

Um Ambiente para Síntese de Imagens de Documentos Históricos

Trabalho de Conclusão de Curso

Engenharia da Computação

Milena Rodrigues Pinheiro de Souza
Orientador: Prof. Carlos Alexandre Barros de Mello

Recife, junho de 2006

Um Ambiente para Síntese de Imagens de Documentos Históricos

Trabalho de Conclusão de Curso

Engenharia da Computação

Este Projeto é apresentado como requisito parcial para obtenção do diploma de Bacharel em Engenharia da Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Milena Rodrigues Pinheiro de Souza
Orientador: Prof. Carlos Alexandre Barros de Mello

Recife, junho de 2006



Milena Rodrigues Pinheiro de Souza

Um Ambiente para Síntese de Imagens de Documentos Históricos

Resumo

No decorrer dos últimos anos, vêm crescendo o número de projetos empenhados na preservação e disponibilização de acervos de documentos históricos em portais *Web*. Muitos desses projetos se restringem à transcrição dos documentos do acervo para mídia digital, através da utilização de *scanners* ou câmeras digitais. Outros oferecem, além disto, versões transpostas para texto que podem ser visualizadas através do navegador *Web*, através, por exemplo, da inserção do arquivo texto no código HTML do sítio, ou de *applets* escritos em Java.

Neste trabalho, foi desenvolvido um sistema para geração completa de imagens de documentos históricos, além de um ambiente gráfico que permite integrar todas as fases envolvidas nesse processo. Para isso, foram implementados diversos algoritmos de processamento de imagens como: limiarização, síntese de texturas e coloração de imagens.

Uma vez adquirido o ambiente de geração de documentos históricos aqui desenvolvido, é possível gerar imagens sintéticas similares aos documentos originais a partir de um arquivo binário contendo informações estatísticas relativas à textura do papel e um arquivo texto correspondente ao conteúdo escrito no documento. Dessa maneira, os usuários precisariam fazer *download* de apenas dois arquivos que juntos ocupam centenas de vezes menos espaço de armazenamento que o documento original, facilitando e agilizando, assim, sua transmissão.

É possível ainda, com a ferramenta desenvolvida, gerar esses documentos a partir da imagem do texto, gerada no processo de segmentação. Além de gerar as versões sintéticas, o ambiente é capaz de fazer a segmentação da imagem original, em imagem do texto e textura, e síntese da imagem da textura.

Apesar de termos usado algoritmos específicos de segmentação e síntese de imagens, é possível adicionar facilmente novos algoritmos à aplicação, necessitando-se que esses novos algoritmos estejam escritos em C/C++ ou Java.

Abstract

The number of projects concerned in preserving historical documents and making them available for *Web* users has been growing at a reasonable rate recently. Most of these projects are restricted to transcribing these documents into digital media using a scanner or digital cameras. Other projects are concerned into integrating them into HTML code or putting them on Java *applets*.

In this work a system for complete generation of historical document was developed, beyond a graphical environment that allows integrate all phases involved in this process. To generate this documents we developed some image processing algorithms like threshold, synthesis and image coloration.

With the programs developed here it is possible to generate a synthetic document similar to the original only with a text archive with it's content and a binary archive with statistical informations about the texture. In this way, the users could download only two documents instead of the original document that need much more storage space. Facilitating and speeding, thus, its transmission.

It is still possible, with the tool developed here, to generate these documents from the image of the text, generated in the segmentation process, and no longer from a text archive. Besides generating the synthetic versions, the environment is capable of making the segmentation of the original image and the synthesis of the texture.

Although we had used specific algorithms of segmentation and synthesis of images, is possible to easily add new algorithms the application, it is necessary that these new algorithms have been written in C/C++ or Java.

Sumário

Índice de Figuras	v
Índice de Tabelas	vii
Tabela de Símbolos e Siglas	viii
1 Introdução	10
2 Conceitos Básicos	15
2.1 Introdução	15
2.2 Processamento de imagens	16
2.2.1 Imagem Digital	20
2.2.2 Fundamento de cores	23
2.2.3 Histograma	28
3 Segmentação	31
3.1 Segmentação por detecção de descontinuidade	31
3.1.1 Detecção de pontos	32
3.1.2 Detecção de linhas	33
3.1.3 Detecção de bordas	33
3.2 Segmentação por detecção de continuidade	34
3.2.1 Limiarização	34
3.2.2 Crescimento de regiões	35
3.2.3 Split e merge de regiões	36
3.3 Segmentação de documentos históricos	37
4 Textura	44
4.1 Análise de textura	44
4.1.1 Medidas matemáticas	45
4.1.2 Análise de variância	45
4.1.3 Matriz de co-ocorrência	45
4.1.4 Novos algoritmos de análise de imagens	47
4.2 Novo algoritmo de síntese de textura	47
4.2.1 Análise dos resultados	50
5 Ambiente para geração de imagens de documentos históricos Segmentação	53
5.1 Descrição do ambiente	53
5.2 Resultados gerados	59
5.2.1 Análise dos resultados	63
6 Conclusões e Trabalhos Futuros	65

Bibliografia

Índice de Figuras

Figura 1. Sistema de geração de documentos históricos por composição de imagens.	12
Figura 2. Sistema de geração por composição de imagens a partir de informações da textura e arquivo texto.	13
Figura 3. Imagem digital de 1921 transmitida entre Londres e Nova Iorque.	16
Figura 4. Imagem utilizada para enfatizar técnicas de Processamento de Imagens.	18
Figura 5. Áreas correlacionadas dentro do processamento gráfico.	19
Figura 6 Amostragem de sinais.	20
Figura 7. Diferentes amostragens de uma imagem.	21
Figura 8. Imagem com diferentes níveis de quantização.	22
Figura 9. Imagem e uma aproximação destacando a matriz de pixels.	23
Figura 10. Decomposição da luz branca ao passar por um prisma.	24
Figura 11. Curva de resposta espectral do olho.	25
Figura 12. Cores primárias e secundárias da luz e dos pigmentos.	26
Figura 13. Diagrama de cromaticidade.	27
Figura 14. Exemplo de imagens claras e escuras, e seus respectivos histogramas.	28
Figura 15. Exemplo de imagem com seu respectivo histograma.	29
Figura 16. Imagem com alto e baixo contraste, e seus respectivamente histogramas.	30
Figura 17. Equalização de histograma	30
Figura 18. Exemplo de máscara para detecção de pontos.	32
Figura 19. Exemplo de aplicação da técnica de detecção de pontos.	32
Figura 20. Filtros para detecção de linhas.	33
Figura 21. Exemplo de aplicação da técnica de detecção de linhas.	33
Figura 22. Máscaras correspondentes ao operador de Sobel (a) e Laplaciano (b).	34
Figura 23. Exemplo de aplicação da técnica de detecção de bordas.	34
Figura 24. Exemplo de aplicação de limiarização.	35
Figura 25. Aplicação do algoritmo de crescimento de regiões.	36
Figura 26. Aplicação do algoritmo de <i>Split e Merge</i> .	36
Figura 27. Exemplo de documento histórico.	37
Figura 28. Comparação de algoritmos de segmentação a imagem.	40
Figura 29. Exemplos das classes do acervo: a) $H=0,22$, b) $H=0,27$ e c) $H = 0,32$	41
Figura 30. Novo algoritmo aplicado à imagem da Figura 28.a).	42
Figura 31. Aplicação do algoritmo a imagens coloridas de cartões postais.	43

Figura 32. Imagem e sua matriz de ocorrência com $d = 1$ e $\theta = 0^\circ$.	46
Figura 33. Gaussianas geradas com uso do Desvio Padrão e da Variância.	49
Figura 34. (topo) Textura original e seu histograma, (embaixo) textura sintética e seu histograma.	51
Figura 35. Textura sintética em zoom.	51
Figura 36. Fluxo de caixa dos arquivos de geração de documentos históricos.	54
Figura 37. Interface gráfica.	56
Figura 38. Interface gráfica e seu menu de opções.	57
Figura 39. Interface com menu de seleção de arquivos.	58
Figura 40. Processo de geração da imagem de um documento histórico. Primeira Fase: Criação dos arquivos de dados e do arquivo de texto com o conteúdo da carta.	59
Figura 41. Processo de geração da imagem de um documento histórico. Segunda Fase: Geração da imagem a partir dos arquivos de dados e de texto.	59
Figura 42. (a) Documento original, (b) imagem do texto resultante do processo de segmentação, (c) documento gerado de acordo com o esquema de geração de imagens da primeira fase, (d) textura original, (e) textura sintética.	60
Figura 43. Base de caracteres usada na geração automática da imagem textual.	61
Figura 44. (a) Documento original, (b) imagem gerada a partir do documento de texto, (c) documento gerado de acordo com o esquema de geração de imagens da segunda fase, (d) textura original, (e) textura sintética.	62

Índice de Tabelas

Tabela 1. Comprimento de onda.	24
Tabela 2. Entropia das matizes em amostras de texturas.	48
Tabela 3. Matizes mais freqüentes e porcentagem de aparição de algumas amostras de texturas.	49
Tabela 4. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 200 texturas para o processo de síntese.	52
Tabela 5. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 50 imagens resultantes do esquema da primeira fase.	63
Tabela 6. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 50 imagens resultantes do esquema da segunda fase.	64

Tabela de Símbolos e Siglas

HTML - *HyperText Markup Language* (Linguagem de Marcação de Hipertexto)
BMP - *Bit map* (Mapa de Bit)
OCR - *Optical Character Recognition* (Reconhecimento Óptico de Caracteres)
ASCII - *American Standard Code for Information Interchange* (Código Padrão Americano para o Intercâmbio de Informações)
PROHIST – Projeto Processamento de Imagens de Documentos Históricos
MB – *Mega Bytes* (Milhões de Bytes)
pixel - *picture element* (Elementos da Imagem)
ppi – *pixel per inch* (Pixel por polegada)
dpi – *dots per inch* (Pontos por polegada)
 λ – Comprimento de onda
 μ - Média
RGB - *Reg, Green, Blue* (Vermelho, Verde, Azul)
CIE – *Commission Internationale de l’Eclairage* (Comissão Internacional de Iluminação)
MOS – *Mean Opinion Score* (Taxa Média de Opinião)
EMQ – Erro Médio Quadrático
PSNR – *Peak Signal-to-Nose Ratio* (Razão Sinal-Ruído de Pico)
ANOVA – Análise de Variância
 σ – Desvio padrão
NS – Diferenças Estatisticamente Não-Significativas
S – Diferenças Estatisticamente Significativas
GLCM – *Grey Level Co-Occurrence Matrix* (Matriz de Co-ocorrência em Escala de Cinza)
 θ - *teta*
 \in - pertence a
- número de elementos do conjunto
HVS – *Human Visual System* (Sistema Visual Humano)
dll - *dynamic linkage library* (Biblioteca de ligação dinâmica)

Agradecimentos

Em primeiro lugar, agradeço à minha irmã Mychelle Rodrigues Pinheiro de Souza e aos meus pais, Mirian Rodrigues da Silva Souza e George Pinheiro de Souza, sem os quais todo esse trajeto não seria possível. Agradeço aos meus familiares, em especial àqueles que conviveram comigo durante o percurso da graduação.

Gostaria de agradecer também aos professores e responsáveis pelo curso de Engenharia de Computação e que estão cumprindo tão bem o papel de docentes dentro da UPE.

Aos meus amigos e colegas, pelo companheirismo e troca de conhecimento, idéias e apoio durante todos esses anos. Em especial, agradeço aos smurfinhos, Cleyton Mario de Oliveira Rodrigues, Marcelo Nunes Alves Costa e Nivia Cruz Quental; assim como a nossos amiguinhos, César Augusto Lins de Oliveira, Laura Arahata de Moraes e Petrônio de Luna Braga; com amizade, companheirismo e paciência, conseguimos superar todos os momentos difíceis que passamos nos últimos anos.

Ao meu orientador e amigo, Carlos Alexandre Barros de Mello, pelo incentivo, paciência, dedicação, prestatividade e encorajamento ao longo dos últimos três anos.

Agradeço também ao time de trabalho do CESAR – Centro de Estudos e Sistemas Avançados do Recife, em especial ao meu gerente, Benedito Alberto Macedo, o qual soube demonstrar paciência e discernimento, nos momentos em que estava extremamente atarefada com minhas obrigações da faculdade, como esta monografia. E também a Paloma de Andrade Nunes e Ícaro Azevedo Matias, pela amizade, ajuda e momentos de descontração.

Agradeço a, Jamile de Moura Lima, pelo carinho, apoio, compreensão e imprescindível ajuda ao longo de nossa amizade, e por todos os bons momentos proporcionados por sua companhia. Obrigada por ter tornado todo esse trajeto mais fácil, com sua ternura e afeto incondicional, me dando sempre ânimo para continuar, apesar das pressões e frustrações sofridas ao longo desse caminho.

Agradeço a Emilly Juliana Alves Ferreira, pela amizade e dedicação e a sua mãe e irmão, Eliane de Fátima Alves Ferreira e Thyago Bruno Alves Ferreira, pelos bons momentos compartilhados durante os dias que me abrigaram em sua casa.

A Adélia Carolina de Andrade Barros, pela amizade e atenção devotada durante esses últimos anos, sempre suportando minhas lamúrias e me ouvindo nos momentos difíceis.

Ao CNPq, pelo incentivo provido durante os três anos de iniciação científica.

A Laert, Sandrinha, Vânia e Djalma, por terem tornado extremamente divertido meu curto período de estágio no Tribunal de Justiça de Pernambuco.

Por fim, mas não menos importante, gostaria de agradecer a todos os outros que fizeram e fazem parte de minha vida de uma forma importante, mas que não foram expressamente citados aqui.

Capítulo 1

Introdução

Indiscutivelmente a escrita se consolidou como um dos principais instrumentos para evolução da humanidade. Com seu surgimento por volta de 4000 a.C., a escrita possibilitou a comunicação entre as mais diversas sociedades através dos tempos, e permitiu que a produção intelectual não mais tivesse que começar sempre do zero - passando a ser um processo incremental. Seu meio de armazenamento sofreu uma série de transformações no decorrer do tempo, desde a escrita talhada em pedra e madeira até chegar a sua forma mais popular nos dias atuais, o papel. Apesar de ser, ainda, o meio mais utilizado, o papel enfrenta uma série de problemas, a saber: ocupação rápida de grandes volumes, fragilidade, consumo de recursos ambientais em larga escala, etc.

Em se tratando de documentos históricos a problemática é ainda maior, por consistirem de um patrimônio de valor inestimável, deixado como herança por nossos antepassados. Por meio deles é possível preservar a memória e a cultura das civilizações, passando-as de geração em geração. Devido a sua importância, é primordial que sejam tomadas todas as medidas cabíveis no sentido de preservá-los. Porém, a preservação de documentos não se apresenta como tarefa trivial, principalmente quando se tratam de documentos históricos - os mais antigos podendo datar de até algumas centenas de anos. Ademais, com o passar do tempo o papel sofre uma série de reações químicas e desgastes naturais. Somam-se a isso os riscos relacionados à manipulação e até mesmo desastres, que vão desde ataques de traças até inundações, incêndios, etc.

Ao mesmo tempo em que é imprescindível preservar a integridade física dos documentos que guardam a história de um País, povo ou nação, é preciso e desejável que esses não percam sua essência, que é transmitir informação e conhecimento. Diante disso, simplesmente resguardá-los e inviabilizar seu acesso, apesar de ser uma medida perfeitamente compreensível do ponto de vista de preservação, não se apresenta como uma solução adequada.

Com o surgimento de mecanismos digitalizadores (*scanner*) e câmeras digitais tornou-se possível transcrever informações escritas em papel para informações digitais, acessíveis por computador, possibilitando, assim, seu armazenamento em diversos tipos de mídia: disquetes, CDs, DVDs, etc.

O armazenamento em formato digital se mostrou uma ótima solução, pois, além de ocupar volume reduzido, economiza recursos naturais e não possui a mesma fragilidade apresentada pelo papel, proporcionando, com isso, uma proteção mais eficiente contra o desgaste provocado pelo tempo. Dessa forma, os documentos originais podem ser devidamente guardados, enquanto todos os interessados podem acessar seu conteúdo, em versão digital. Além de facilitar o acesso aos documentos, a versão digital permite que seja feita a cópia e o armazenamento dos mesmos.

À primeira vista essa pode parecer a solução ideal por resolver a maioria das deficiências apresentadas pelo papel, no entanto, a digitalização dos documentos também apresenta alguns problemas. Os arquivos digitalizados são armazenados como imagens, para isso é preciso guardar informações referentes a cada pixel da imagem, que requer grande espaço de armazenamento. Como exemplo, podemos verificar que a digitalização de uma página em papel A4 com 256 tons de cinza e 200 dpi de resolução, gera uma imagem em BMP com aproximadamente 4 MB.

Apenas com o surgimento do Reconhecimento Óptico de Caracteres (OCR - *Optical Character Recognition*) [1] se tornou realmente viável o armazenamento de textos digitalizados. A imagem do documento digitalizado é automaticamente transposta para arquivo de texto, utilizando caracteres em código ASCII (*American Standard Code for Information Interchange*). Sendo o código ASCII, uma codificação numérica utilizada na representação de caracteres em computadores e dispositivos de armazenamento eletrônico de dados, que mapeia 128 caracteres da língua inglesa com códigos de 0 a 127.

As ferramentas OCR são desenvolvidas, em sua grande maioria, utilizando redes neurais artificiais, e atualmente apresentam taxa de acerto superior a 99% para documentos impressos ou datilografados. O mesmo conteúdo em arquivo de texto passa a ocupar centenas de vezes menos espaço - variando de acordo com a quantidade de informação presente na imagem, ademais, possibilita a aplicação de mecanismos de busca.

Visando a preservar e facilitar a divulgação de documentos históricos, o objetivo deste trabalho centra-se em possibilitar a criação de um banco de imagens, com informações a respeito das mesmas, que podem ser usadas para gerar versões sintéticas da imagem original. E, além disso, desenvolver uma ferramenta gráfica que agregue todas as etapas de geração das versões sintéticas dos documentos históricos, conferindo a esse processo um caráter de trabalho fácil e intuitivo.

Anteriormente, esse processo era feito separadamente, sendo preciso digitar linhas de comando no console para cada fase, o que tornava o processo extremamente enfadonho e demorado, atualmente bastam alguns “cliques” no *mouse*.

O acervo a ser trabalhado é composto por, aproximadamente, 6.500 documentos - totalizando mais de 30.000 páginas de documentos do final do século XIX e início do século XX, que fazem parte do Projeto ProHist [2]. Devido ao efeito do tempo e alta alcalinidade do papel desse período, esses documentos apresentam grande desgaste, o que dificulta o uso de diversas técnicas de processamento de imagens.

A geração de imagens sintéticas de documentos históricos (ver Figura 1) consiste, *a priori*, em sua aquisição através de algum dispositivo digitalizador e segmentação do documento original, gerando assim, duas imagens distintas: uma relativa à imagem da tinta (texto) e a outra relativa à imagem do papel (textura) [3].

Após essa fase inicial de segmentação, é possível tratar, separadamente, as duas imagens. A imagem da tinta pode, então, ser transposta para texto utilizando uma ferramenta OCR. Pode-se utilizar a ferramenta OCR antes ou após a segmentação, sendo que, algumas vezes, elas apresentam melhores resultados se aplicadas à imagem do texto e não da imagem original.

A partir da imagem da textura, é possível extrair uma série de informações estatísticas que são necessárias para geração da textura sintética do documento que será gerado. Ao final de todo o processo, a imagem do texto e a imagem da textura gerada são compostas, gerando o documento final, uma versão sintética similar ao documento original.

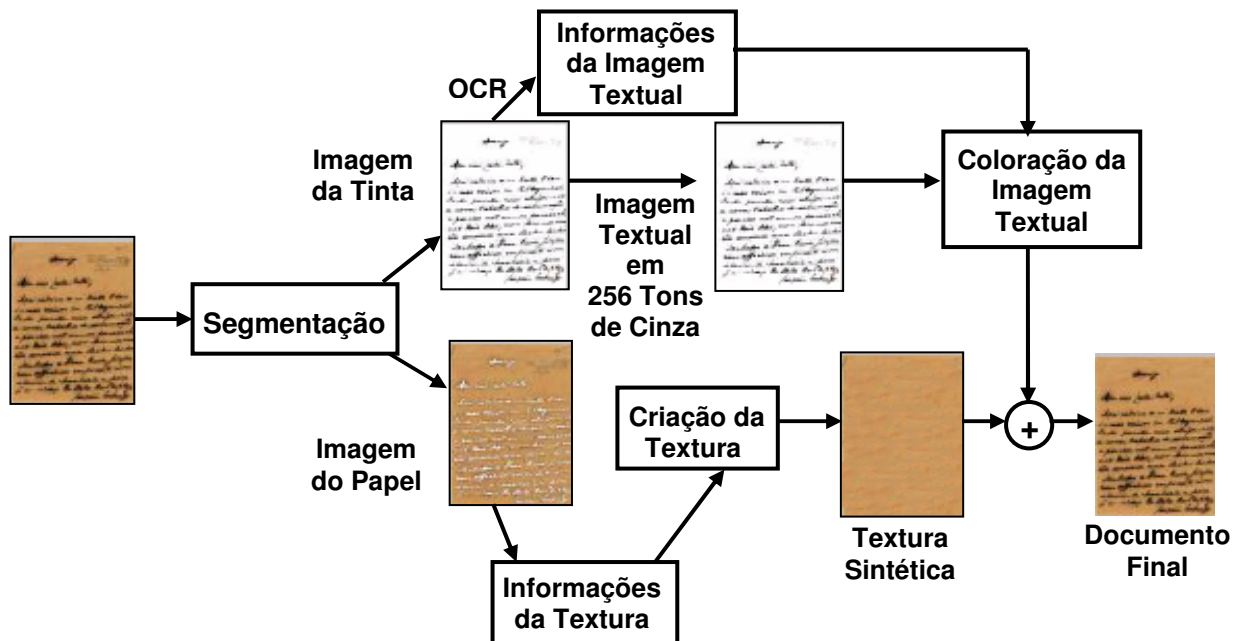


Figura 1. Sistema de geração de documentos históricos por composição de imagens.

Para os interessados apenas no conteúdo presente no documento, o arquivo de texto é suficiente. No entanto, existem inúmeros usuários, que gostariam de possuir uma versão similar, ou até mesmo a versão original, do documento digitalizado, apesar de estas versões requererem um espaço de armazenamento bem maior. Podemos, então, classificar os usuários em quatro tipos/classes:

- Classe 1. Aqueles interessados apenas no conteúdo presente no documento, como mencionado acima; esses só precisariam de um arquivo texto com essas informações.
- Classe 2. Aqueles interessados na imagem textual tal qual se apresenta no texto original e não apenas em arquivo texto com seu conteúdo, mas que não possuem interesse nas informações sobre o fundo do papel (textura). Os usuários inseridos nesta classe poderiam optar entre adquirir a imagem da tinta ou um arquivo texto, a partir do qual a imagem sintética da tinta poderá ser gerada.
- Classe 3. Aqueles que gostariam de ter a impressão de estarem lendo o documento original, precisariam adquirir um arquivo binário, com informações da textura, e uma imagem da tinta ou o arquivo texto correspondente a esta imagem.
- Classe 4. Aqueles que realmente gostariam de acessar a versão original do documento digitalizada.

Quanto maior for a proximidade entre a versão do documento e a versão original, maior será o espaço de armazenamento requerido.

Devido aos documentos das classes 4 e 1 serem gerados diretamente a partir da digitalização e submissão a uma ferramenta OCR, respectivamente, este trabalho se aplica principalmente às classes 2 e 3.

O Capítulo 2 desta monografia faz uma revisão dos principais conceitos básicos necessários para compreensão dos conceitos aqui abordados. Dentre esses conceitos encontram-se os de imagem digital e fundamento de cores.

A segmentação da imagem original em imagem do texto e imagem da textura é a etapa mais importante e delicada de todo o processo, tendo em vista que o resultado da segmentação

tem influência direta na qualidade das imagens que serão geradas. A principal dificuldade dos algoritmos de segmentação está em escolher um ponto de corte no histograma, através do qual é possível classificar a cor do pixel como pertencente ao texto ou à textura. A esse ponto de corte damos o nome de *limiar*. Para efetuar uma segmentação adequada utilizaremos, inicialmente, um algoritmo de segmentação de imagens baseado em entropia [4]. Uma explanação a respeito de algumas técnicas de segmentação, incluindo a técnica baseada em entropia que foi re-implementada, é vista em maiores detalhes no Capítulo 3.

No Capítulo 4, são apresentados alguns procedimentos para análise de texturas, e examinaremos em detalhes a geração da textura sintética [5] da imagem e feita através de informações estatísticas extraídas da textura da imagem do documento original. Essas informações são armazenadas em um arquivo binário e ocupa cerca de 80 bytes.

A partir do arquivo texto e do arquivo binário contendo as informações estatísticas necessárias para geração da textura sintética, é possível a geração completa do documento sintético, como pode ser observado na Figura 2. Esses dois arquivos poderiam ser, *a posteriori*, disponibilizados em um banco de dados na *Web*, para que o usuário não precisasse baixar as imagens originais dos documentos, que ocupam, entre 2MB e 12 MB, cada uma; e sim esses dois arquivos que juntos ocupam poucos *bytes* (80 *bytes* das informações da textura somados com o arquivo texto que ocupa entre 200 a 3000 *bytes*, a depender da quantidade de informação escrita presente no documento).

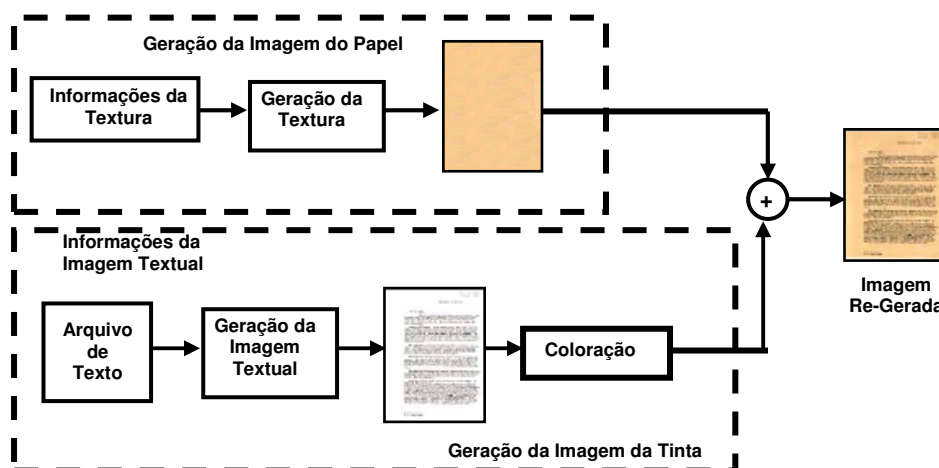


Figura 2. Sistema de geração por composição de imagens a partir de informações da textura e arquivo texto.

Para geração da imagem textual, a partir do arquivo texto, foi utilizado um conjunto de imagens contendo números, letras e caracteres especiais extraídas de imagens do acervo, sendo o arquivo texto percorrido e a cada caractere, a imagem correspondente a ele vai sendo lida e desenhada até que a imagem final seja gerada.

Após a geração da imagem do texto, aplicamos a textura a ela e obtemos, como resultado final, o documento sintético gerado.

No Capítulo 5, é mostrado como foram feitos os algoritmos envolvidos no processo de geração dos documentos, a interface que agrega estes algoritmos e alguns resultados alcançados. Logo após, passamos às conclusões e trabalhos futuros.

Os algoritmos de segmentação, síntese, geração de imagem a partir de arquivos texto e composição das imagens do texto com a textura para geração do documento final foram implementados em C, sendo invocados, utilizando o método *native* por uma interface gráfica implementada em Java.

Outros trabalhos importantes vêm sendo desenvolvidos internacionalmente com o objetivo de disponibilizar versões digitais de documentos históricos, dentre eles merecem destaque:

- Projeto Gutenberg [6]: Desenvolvido pela Universidade de Keio, Tóquio, dentro do Projeto HUMI (*Humanities Media Interface Project*). HUMI, lê-se “fumi”, também é uma palavra japonesa que significa documentos, literatura e história. Nesse projeto a versão digital da Bíblia impressa originalmente por Gutenberg, composta por dois volumes, com mais de 1.300 páginas foram digitalizadas em apenas 4 dias. Desde então, outros livros raros vêm sendo adquiridos.
- Livro de Emblemas [7]: Projeto desenvolvido pela Universidade de Glasgow, para criação de um acervo digital de emblemas franceses.
- Projeto ARTFL (*American and French Research on the Treasure of the French Language*) [8]: Desenvolvido pela Universidade de Chicago, EUA. Possui um banco de dados de enciclopédias diversas coleções de textos e enciclopédias francesas do século XVI, totalizando mais de 18.000 páginas.
- *Koninklijke Bibliotheek* [9]: Uma iniciativa da Biblioteca Nacional dos Países Baixos, apresenta uma seleção de livros, jornais, revistas e diversas publicações eletrônicas referentes aos Países Baixos. O portal do projeto apresenta diversos *links* interessantes para outros bancos de dados de documentos.
- *The Digital Scriptorium* [10]: Desenvolvido por diversas bibliotecas filiadas às universidades de Berkeley e Columbia nos Estados Unidos da América, o projeto apresenta uma base de dados digital de manuscritos medievais e da renascença. O acervo possui mais de 15.000 imagens coloridas apresentadas em três tamanhos para visualização em tela (pequeno, médio e grande).
- Cervantes Digital Library [11]: Portal dedicado às obras de Miguel de Cervantes, contendo sua bibliografia, fotos e pinturas, e uma coleção de arquivos digitais de suas obras, incluindo, evidentemente, Dom Quixote, a qual aparece tanto como imagens digitalizadas quanto como texto em inglês e espanhol. O projeto é fruto da colaboração de diversas universidades e do Centre de Estudios Cervantinos na Espanha.
- Projeto DEBORA (*Digital AccEss to BOoks of the RenAissance*) [12]: Projeto Europeu que visa o desenvolvimento de um conjunto de ferramentas para acesso remoto a livros do século XVI digitalizados, sem a necessidade de se deslocar até as bibliotecas onde os livros originais estão localizados.

O trabalho apresentado nessa monografia gerou o seguinte artigo:

MELLO, C. A. B. ; CAVALCANTI, C. S. ; **RODRIGUES, M. P. S.** . Generating True Color Paper Texture of Historical Documents. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005, Filadélfia, EUA.

Capítulo 2

Conceitos Básicos

Sendo o foco principal deste trabalho a manipulação de imagens digitais, faz-se necessário apresentar alguns conceitos relacionados ao tratamento de imagens.

2.1 Introdução

Guardada a notável fama dos computadores no rápido processamento de tarefas, na maioria das vezes irrealizáveis por seus operadores, existe grande dificuldade na implementação de algoritmos que realizem tarefas fáceis e naturais para nós - como a de reconhecimento de voz, sabores e odores conhecidos, entre outras coisas. Contudo, muitas das atividades que facilmente desempenhamos exigem um grande dispêndio de processamento mental, como é o caso da visão, que ocupa, aproximadamente, 55% de nossa atividade cerebral.

Nossa visão nos proporciona um imenso fascínio, face às imagens que processamos continuamente. Com suas formas e cores diversas, essas imagens adquirem semântica e são, muitas vezes, associadas a alguma emoção ou sentimento. No entanto, a que se destacar o caráter efêmero dessas imagens, captadas por nossos olhos e armazenadas, algumas vezes, em nossas mentes e que, com o passar do tempo, vão se apagando, passando a se constituir em uma vaga lembrança, culminando no esquecimento. Outrossim, essas imagens não podem ser compartilhadas, em sua essência primeira, com ninguém.

Diante da natureza “tênue” dessas imagens, é compreensível que busquemos capturá-las, na tentativa de imortalizar certos momentos e figuras queridas e importantes.

Apesar das inúmeras tentativas, desde a pré-história, de representar a realidade através das pinturas rupestres, apenas em 1825 foi produzida a primeira fotografia reconhecida, que possibilitou mais que uma simples representação da realidade, proporcionou um registro material dela. A partir desse momento, iniciou-se uma busca constante pelo aperfeiçoamento dos processos e métodos fotográficos, até chegarmos ao filme colorido, o foco automático, a exposição automática e, mais recentemente, às câmeras digitais.

Concomitantemente ao rápido avanço das técnicas fotográficas, estavam sendo desenvolvidos os primeiros aparelhos de televisão - que também sofreram diversas transformações, indo da exibição em preto em branco à colorida, passando de tela convencional à tela plana e, nos últimos anos, a revolução tecnológica ameaça o clássico aparelho de televisão, que utiliza tubos de raios catódicos, fazendo surgir os televisores de plasma.

Vivemos e sobrevivemos da busca incessante por meios que nos possibilitem representar fidedignamente nossa realidade. Por isso, imbuídos pelo avanço tecnológico desta nova Era, nada mais natural que tentar transpor nossa realidade aos computadores, permitindo a detenção interfaces mais amigáveis, representação de imagens e filmes, além do armazenamento, transmissão e manipulação dessas imagens. Esse movimento originou áreas como a computação gráfica, processamento de imagens e visão computacional.

2.2 Processamento de imagens

O processamento de imagens foca a melhoria da qualidade das imagens, seja tornando-as mais próximas do real, de acordo com a visualização e interpretação humana, ou tratando imagens que servirão como entrada para outros sistemas, a exemplo do pré-processamento de imagens para sistemas de visão computacional.

Uma das primeiras aplicações conhecidas, em que foram utilizadas técnicas de processamento de imagens ocorreu na década de 20 e objetivava aperfeiçoar a transmissão de figuras de jornal - essas seriam transmitidas entre Londres e Nova Iorque. Uma das imagens transmitidas dessa forma pode ser vista na Figura 3. A aplicação consistia em codificar as imagens, transmiti-las através de um cabo submarino e reconstruí-las quando chegassem ao seu destino, a mudança da forma com que as imagens eram codificadas resultou em uma redução no tempo de transmissão, que antes era de mais de uma semana, para menos de três horas [13].



Figura 3. Imagem digital de 1921 transmitida entre Londres e Nova Iorque.
(Figura retirada do livro “Digital Image Processing”, R. GONZALEZ e, P. WINTZ, ed. Addison-Wesley Publishing Company, Inc., 2ª edição, 1987)

Seguiram-se vários estudos, não apenas no sentido de aperfeiçoar as técnicas de transmissão, como, também, para otimizar a qualidade das imagens transmitidas.

O desenvolvimento dos computadores e o aperfeiçoamento das imagens proporcionaram um aumento na resolução e nos níveis de quantização [14]. Por esse motivo, as pesquisas na área de compressão de imagens tornaram-se cada vez mais críticas, visto que o número de informações contidas em uma imagem de qualidade exigia um maior espaço para o armazenamento da mesma.

Existe até hoje um *trade-off* entre qualidade e tamanho de armazenamento. Os usuários de grande parte das aplicações gráficas esperam uma ótima qualidade das imagens, todavia, em geral, quanto melhor a imagem, maior o espaço em disco por ela ocupado.

A partir dessa dicotomia entre qualidade e espaço de armazenamento em disco, surgiram duas sub-áreas de pesquisa dentro do processamento de imagens: quantização e compressão de imagens.

Essas duas áreas possuem o objetivo comum de reduzir o espaço de armazenamento das imagens, porém, fazem uso de técnicas distintas para isso. Enquanto a quantização de imagem (examinaremos o conceito de quantização com maiores detalhes na seção 2.2.1) procura reduzir o número de cores utilizados na representação de uma imagem, a compressão de imagens procura criar algoritmos de codificação que permitam sintetizar a forma como essa imagem é representada internamente para o mecanismo de exibição. As técnicas de compressão podem, ou não, ocasionar a perda de informações.

A área de processamento de imagens tem um caráter multidisciplinar e vem sendo aplicada na intenção de facilitar a visualização e melhorar a interpretação das imagens tratadas.

São inegáveis os benefícios trazidos pelo uso do processamento de imagens. É quando da utilização desses recursos que conseguimos lograr êxito na reconstrução de fósseis, utilizando figuras de achados arqueológicos, também em ultra-sonografias - onde já é possível observar o feto com impressionante riqueza de detalhes [14]. Além de tudo isso, ainda podemos citar a otimização de imagens captadas por satélites e uma série de outras aplicações que são realizadas nas áreas da astronomia, biologia, física nuclear, indústria e órgãos de defesa [14].

Na Figura 4 podem ser vistas algumas imagens e o resultado do processamento delas. Primeiro temos uma imagem da superfície de Marte enviada à Terra Figura 4.a, cujos ruídos de transmissão foram reduzidos Figura 4.b; em seguida, a imagem de um raio-X Figura 4.c teve seu contraste e os contornos melhorados Figura 4.d; e, por fim, uma terceira imagem que se encontrava desbotada e sobreposta Figura 4.e foi submetida à aplicação de um filtro Figura 4.f [14].

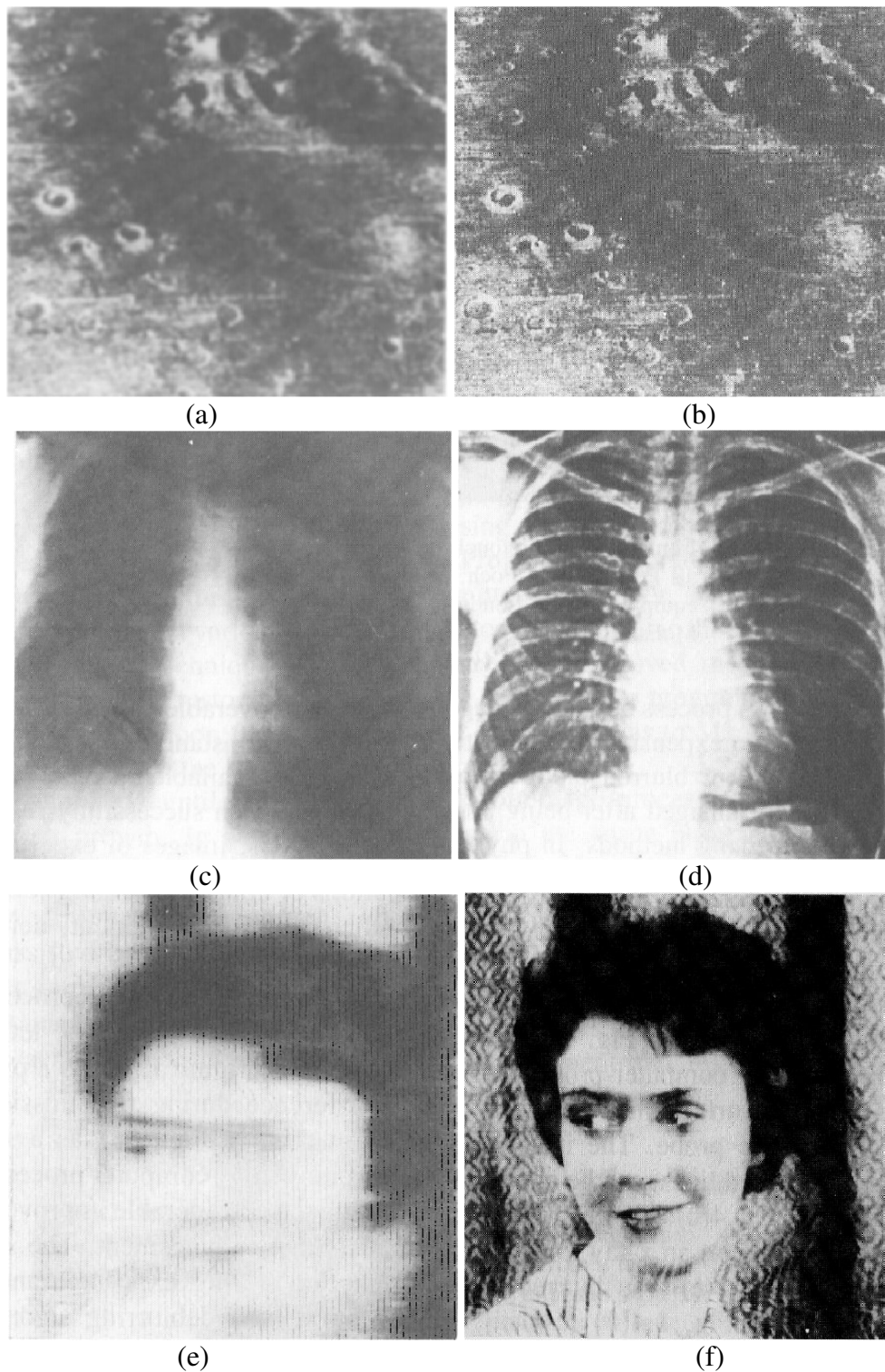


Figura 4. Imagem utilizada para enfatizar técnicas de Processamento de Imagens. (a) Imagem da superfície de Marte enviada à Terra, (b) imagem (a) com redução dos ruídos de transmissão, (c) imagem de um raio-X, (d) imagem do raio-X (c) com contraste e contornos melhorados, (e) imagem desbotada e (f) imagem (e) submetida à aplicação de um filtro. (Figura retirada do livro “Digital Image Processing”, R. GONZALEZ e, P. WINTZ, ed. Addison-Wesley Publishing Company, Inc., 2ª edição, 1987).

Três grandes áreas necessitam da manipulação de imagens: computação gráfica, processamento de imagens e visão computacional. Essas áreas são, por muitas vezes, tão próximas que podem ser confundidas. Para diferenciá-las entre si, é comum observarmos a natureza da entrada e saída utilizadas. De acordo com esse critério de avaliação, é possível construir o fluxograma da Figura 5.

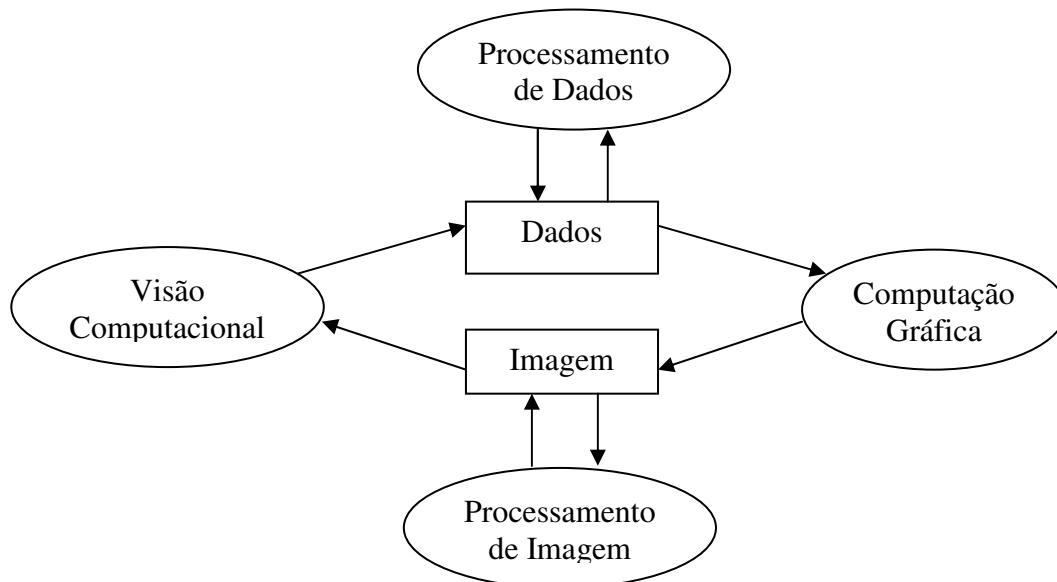


Figura 5. Áreas correlacionadas dentro do processamento gráfico.

As quatro áreas representadas na Figura 5 podem ser definidas, informalmente, da seguinte forma [14]:

- **Processamento de dados:** aplicação onde tanto as entradas quanto as saídas consistem em dados. Grande parte das aplicações tradicionais utilizam apenas processamento de dados.
- **Computação gráfica:** a partir de dados geram-se imagens. Com a sofisticação das ferramentas de desenvolvimento de aplicações de computação gráfica, a geração de imagem é construída de forma quase automática, onde o conhecimento da ferramenta é, muitas vezes, mais primordial que o conhecimento dos dados que estão sendo gerados para formar as imagens. Nos últimos anos, a área de computação gráfica vem despertando grande interesse, o que impulsionou pesquisas na área. Pesquisas estas que se fizeram presentes, com veemência, na indústria de jogos e na indústria cinematográfica.
- **Processamento de imagens:** aplicação em que as entradas e saídas se constituem de imagens. Consiste, basicamente, na manipulação de imagens, objetivando melhorá-las do ponto de vista visual, tanto para visualização humana quanto para a interpretação automática.
- **Visão computacional:** gera dados a partir de imagens. Essa área possui uma forte correlação com a área de inteligência artificial, e é especialmente utilizada com o objetivo de dar semântica às imagens captadas por dispositivos de aquisição de imagens que podem estar servindo, por exemplo, como olhos de um robô.

Como mencionado anteriormente, a classificação de um projeto, entre essas áreas, pode não se mostrar de forma clara, uma vez que, em grande parte dos casos, estas áreas cooperam entre si, maximizando suas potencialidades.

2.2.1 Imagem Digital

A imagem captada por nossos olhos é produto da reação a estímulos luminosos que permitem a projeção, na nossa retina, do universo tridimensional à nossa frente. Nesse sentido, podemos entender uma imagem como uma projeção em uma superfície bidimensional (nesse caso a retina) de intensidades de estímulos luminosos. A imagem captada por nossos olhos recebe impulsos luminosos de cada ponto do espaço visível, assim, cada ponto será associado a uma cor. Como esses pontos são infinitos, a imagem percebida consiste em um sinal contínuo.

Uma imagem digital é uma representação discreta do número de pontos contidos no universo visível e do número de cores que se associam a cada ponto. Uma representação discreta significa tomar um subconjunto finito (discreto) do universo infinito que define a imagem.

De acordo com a nomenclatura utilizada na teoria de sinais, denominamos amostragem toda discretização feita no domínio, e de quantização o processo de discretização do contra-domínio [14]. Em imagens a amostragem ocorre no espaço (2D) e não no tempo como é mais comum em amostragem dos sinais. Na Figura 6.a temos um sinal contínuo, na Figura 6.b foram demarcados os instantes deste sinal que se deseja e na Figura 6.c é mostrado o sinal amostrado nos instantes escolhidos.

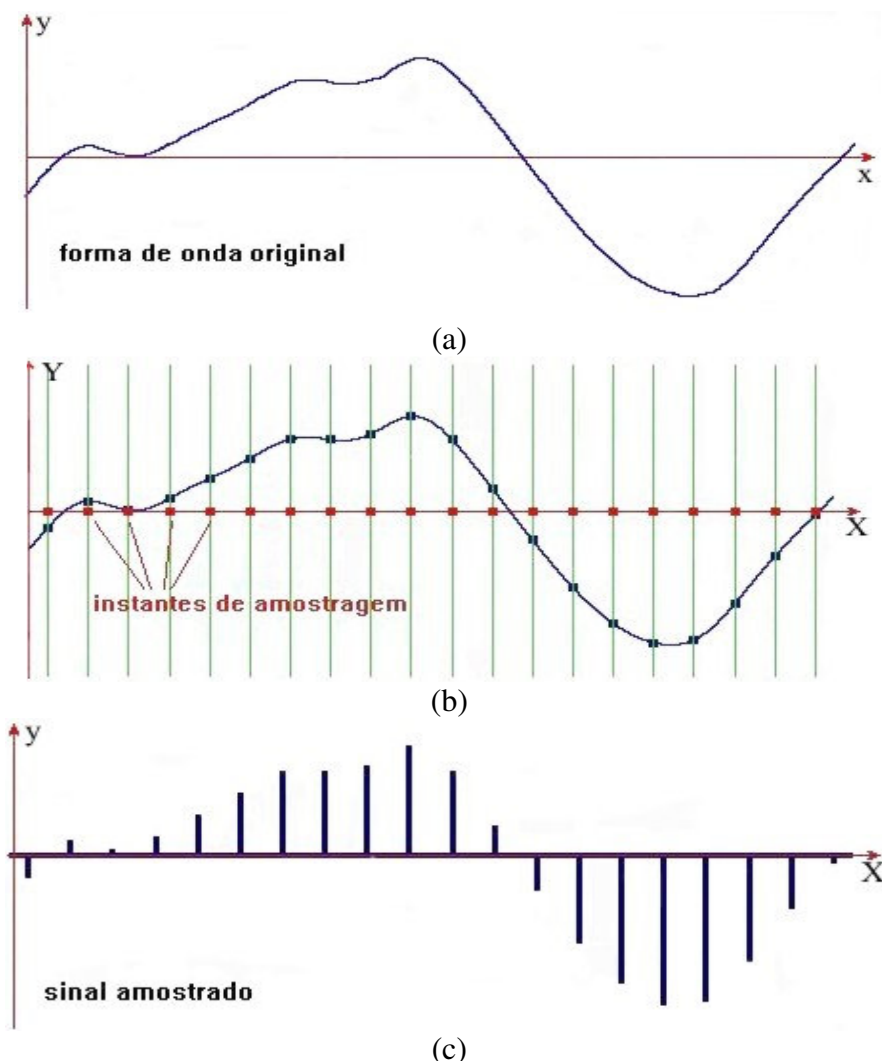


Figura 6. Amostragem de sinais. (a) Sinal contínuo, (b) sinal contínuo com demarcação dos instantes de amostragem e (c) sinal amostrado.

Cada ponto da amostragem utilizado para representar a imagem constitui-se como a menor unidade dela e recebe o nome de elemento da imagem, ou, como é mais chamado, *pixel*.

Dessa forma, denominamos a amostragem de uma imagem como sendo a quantização do número de pontos existentes na imagem; o número de pixel presentes na imagem dependerá desse processo de amostragem. Na Figura 7, temos, como exemplo, uma imagem com diferentes amostragens. A imagem mostrada na figura 7.a foi amostrada com 420 x 348 pixels; a imagem 7.b foi amostra com 305 x 261 pixels, já é possível observar sutis diferenças entre as imagens 7.a e 7.b; essas diferenças se acentuam nas imagens 7.c, com 210 x 174 pixels, e 7.d com 105 x 87 pixels. Todas as imagens estão sendo apresentadas de forma ampliada para melhor visualização.



Figura 7. Diferentes amostragens de uma imagem. (a) Imagem com amostrada com 420 x 348 pixels, (b) imagem com amostrada com 305 x 261 pixels, (c) imagem com amostrada com 210 x 174 pixels e (d) imagem com amostrada com 105 x 87 pixels.

Já a quantização de imagens nada mais é que a discretização do espaço de cor da mesma. Na Figura 8, as imagens (a), (b), (c) e (d) apresentam diferentes níveis de quantização, tendo sido discretizadas com, respectivamente, 16 milhões, 256, 16 e 2 cores (geralmente preto e branco).

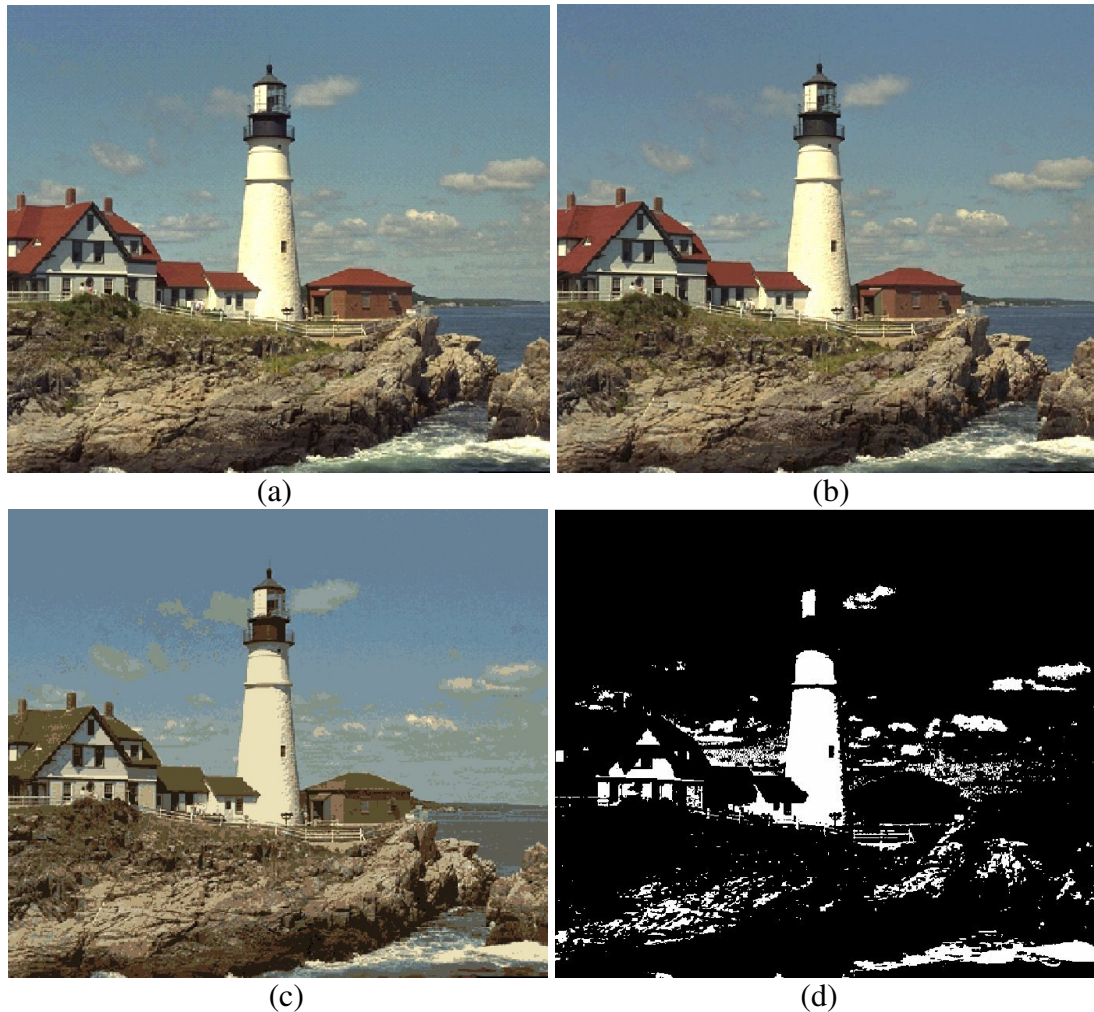


Figura 8. Imagem com diferentes níveis de quantização. (a) Imagem com 16 milhões de cores, (b) imagem com 256 de cores, (c) imagem com 16 de cores e (d) imagem com 2 cores.

A forma mais comum de representar uma imagem consiste em transformá-la em matriz de pixels, onde cada pixel possui uma coordenada e um valor de cor associado a ele. Podemos tomar como exemplo uma pequena imagem retangular, mostrada na Figura 9. A imagem é vista como uma malha de pixels, os números de linhas na vertical e horizontal definem, respectivamente, a resolução vertical e horizontal. Seguindo o exemplo, se tivermos 39 pixels na horizontal e 23 pixels na vertical, seu produto, denominado resolução espacial ou geométrica da imagem, será 897.

Para cada um dos 897 pixels temos um valor que representa a intensidade luminosa associada ao pixel; esse valor varia sempre entre a menor intensidade luminosa (preto) e a maior intensidade luminosa (branco).

A resolução espacial nos diz a frequência espacial com que a imagem foi amostrada, dessa forma, podemos concluir que quanto maior a resolução, maior a riqueza dos detalhes percebidos na imagem. No entanto, devido ao tamanho físico do pixel nos diferentes dispositivos onde a imagem será visualizada variar bastante, a qualidade da imagem estará vinculada, além da distância do observador em relação a mesma, ao número de pixels em uma determinada área - o que é denominado densidade de resolução da imagem e, em geral, é representada pelo número de pixels por polegadas, ppi (*pixel per inch*) ou dpi (*dots per inch*).

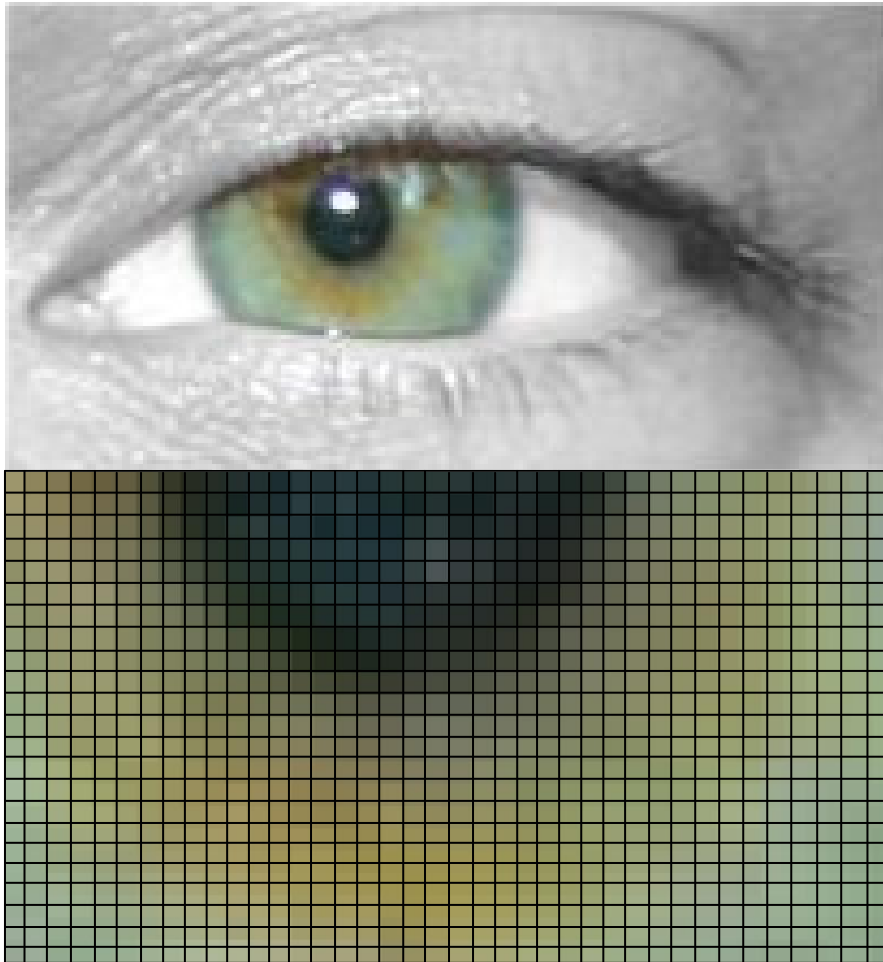


Figura 9. Imagem e uma aproximação destacando a matriz de pixels. No topo a imagem de um olho e abaixo uma aproximação da imagem do olho, destacando a matriz de pixels.

O espaço de cor de uma imagem, ou gamute, está relacionado com o número de cores que podem ser associadas a um pixel. Em uma imagem monocromática temos um gamute de um bit, onde podemos representar apenas duas cores, geralmente preto e branco. Como o sistema utilizado pelo computador é o binário, o número de cores possíveis em uma imagem digital se encontra entre 0 e uma potência de dois. O número máximo de cores de uma imagem costuma variar entre 2 e aproximadamente 16 milhões, sendo assim, seu gamute varia entre 1 e 24 bits.

2.2.2 Fundamento de cores

A cor é um fenômeno psicofísico decorrente da interação entre a luz e o sistema visual do observador [14].

A luz apresenta um comportamento dual, ou seja, tanto de partícula quanto de onda. Um raio luminoso é formado por fótons, que se movimentam em forma de onda em uma determinada frequência. Esses fótons chegam até a retina, gerando impulsos elétricos que, ao chegarem ao cérebro, são traduzidos em cores. Sendo uma onda eletromagnética, a luz se propaga no vácuo com velocidade (c) de $3 \cdot 10^8$ m/s (velocidade da luz). Juntamente com a velocidade de propagação, a frequência (f) com que a onda se propaga determina o comprimento de onda (λ) desta radiação, como pode ser visto na Eq. 1.

$$\lambda = \frac{c}{f} \quad (\text{Eq. 1})$$

O olho humano é capaz de distinguir cores que possuem comprimento de onda entre 380 μm e 780 μm , intervalo denominado de faixa visível do espectro de luz.

Em 1666, Isaac Newton descobriu que, quando um feixe de luz branca atravessa um prisma, a luz emergente não é branca, e sim, um espectro de cores que vão do violeta ao vermelho, podendo ser dividido em seis grandes regiões: violeta, azul, verde, amarelo, laranja e vermelho (ver Figura 10). A separação das cores no prisma ocorre porque o vidro apresenta um índice de refração diferente para cada λ . De acordo com a 2ª lei de Snell, cada λ possuirá um ângulo de refração diferente.

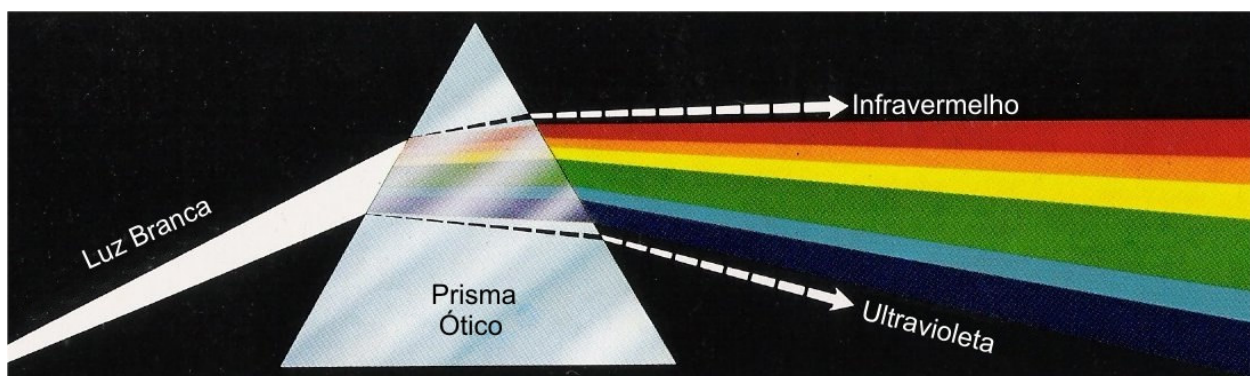


Figura 10. Decomposição da luz branca ao passar por um prisma.
(Figura retirada do livro “Digital Image Processing”, R. GONZALEZ e, P. WINTZ, ed. Addison-Wesley Publishing Company, Inc., 2ª edição, 1987)

Os comprimentos de onda dessas seis regiões variam nas seguintes faixas:

Tabela 1. Comprimento de onda

Cor	Faixa do Espectro
Violeta	380 – 440 μm
Azul	440 – 490 μm
Verde	490 – 565 μm
Amarelo	565 – 590 μm
Laranja	590 – 630 μm
Vermelho	630 – 780 μm

Newton acreditava que o olho possuía um número infindável de células fotossensíveis. Porém, em meados do século XIX, o físico T. Young, a partir de uma série de experimentos, desenvolveu o modelo tricromático para o olho humano. Young afirmava que possuímos apenas três tipos de células fotossensíveis, que captariam estímulos de baixa, média e alta frequência. Sendo assim, segundo a teoria de Young, em vez de um número infinito de sensores, possuímos apenas três tipos [14].

A superfície da retina humana é formada por dois tipos de receptores de raios luminosos, os cones e os bastonetes [13]. Os bastonetes fornecem a sensação de intensidade luminosa (tons de cinza, que variam da menor intensidade (preto) a maior (branco)). Por sua vez, os cones se localizam na parte central da superfície da retina e são responsáveis pela percepção de cor.

Existem três tipos de cones, aqueles sensíveis aos comprimentos de onda curto, médio e longo, a curva de resposta espectral dos cones pode ser visualizada na Figura 11.

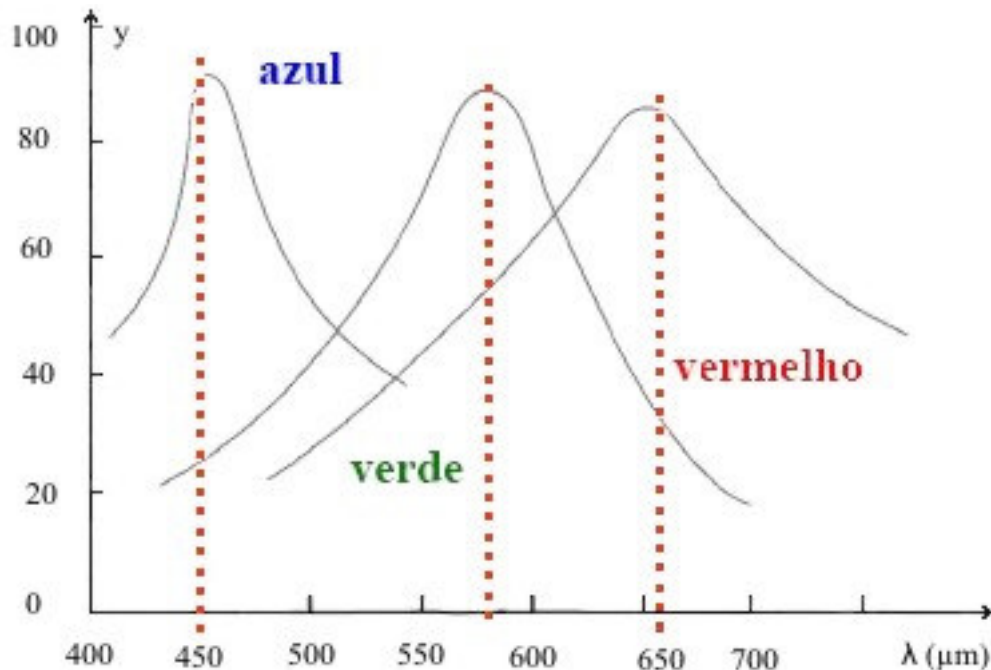


Figura 11. Curva de resposta espectral do olho.

Os pontos máximos das curvas de resposta espectral mostradas na Figura 11 correspondem, respectivamente, as faixas de vermelho, verde e azul, que representam as frequências baixa, média e alta.

Posteriormente, Helmholtz passou a apoiar a teoria de Young e essa passou a ser conhecida como teoria de Young-Helmholtz.

Dessa forma, as cores são vistas a partir de uma combinação de três cores primárias: vermelho (*Red - R*), verde (*Green - R*) e azul (*Blue - B*). Um sistema de representação tridimensional do espaço de cor foi, então, adotado como padrão em 1931 pela comissão Internacional de iluminação (*Commission Internationale de l'Eclairage - CIE*).

Essas três cores, tomadas como primárias, podem representar nos monitores de vídeo até 16.777.216 cores, ou 16 milhões de cores, utilizando 24 bits, ou 1 byte para o armazenamento de cada componente de cor. Não é interessante aumentar o número de cores além de dezesseis milhões, visto que o olho humano não tem sensibilidade suficiente para distinguir matizes tão próximas.

A combinação entre as cores primárias, duas a duas, produz as cores ditas secundárias, que são: magenta (vermelho e azul), ciano (verde e azul) e amarelo (vermelho e verde). Finalmente, a mistura entre as três cores primárias, na intensidade correta, resulta na cor branca. No topo da Figura 12 são mostrados três círculos representando as cores primárias da luz e as interseções destes ilustrando a mistura entre elas.

Existem diferenças entre as cores primárias da luz e as cores primárias de pigmentos e corantes. Nos pigmentos e corantes as cores primárias são: magenta, ciano e amarelo. Enquanto como cores secundárias têm-se, o vermelho, verde e azul. A combinação entre as três cores primárias dos pigmentos, na intensidade correta, produz a cor preta. Na base da Figura 12 pode-se ver as cores primárias de pigmentos e corantes, bem como o resultado da mistura entre eles.



Figura 12. Cores primárias e secundárias da luz e dos pigmentos.
 (Figura retirada do livro “Digital Image Processing”, R. GONZALEZ e , P. WINTZ, ed.
 Addison-Wesley Publishing Company, Inc., 2ª edição, 1987)

Existem três características bastante utilizadas para fazer a distinção entre cores, são elas: matiz, brilho ou luminância e saturação.

A teoria de Hering afirma a existência de três canais duplos de cores no cérebro, onde um contém as componentes de preto e branco, e os outros dois canais, as cores vermelho-verde e azul-amarelo.

Se combinarmos a teoria de Young-Helmholtz à teoria de Hering, teremos que, os três canais conteriam os sinais combinados na forma R-G, B-Y e R+G. Onde R+G corresponderia à cor amarela (Y), que está diretamente associada com a luminância (L), visto que a componente azul é praticamente desprezível para seu cálculo:

$$L = 0,176R + 0,81G + 0,011B \quad (\text{Eq. 2})$$

As componentes R-G e R-Y fornecem a informação de tonalidade (matiz e saturação) da cor, também conhecida como crominância.

O brilho, ou luminância, se refere à intensidade de luz presente na cor.

A matiz fornece a informação sobre o comprimento de onda dominante, representando, assim, a cor propriamente dita - azul, verde, amarelo, etc.

A saturação está associada ao grau de pureza da cor, ou intensidade de luz branca adicionada a ela. Quanto maior a quantidade de luz branca, menor a saturação. Sendo assim, uma cor pura teria o máximo de saturação e nenhuma mistura com a cor branca.

O plano do crominância ou plano de Maxwell (Eq. 3) é uma representação tridimensional do espaço de cor. A partir das três coordenadas (x,y,z) é possível obter qualquer tonalidade de cor.

$$X + Y + Z = 1 \quad (\text{Eq. 3})$$

Na Figura 13, pode-se visualizar uma projeção ortogonal do diagrama de cromaticidade nos eixos x e y, retirando a coordenada y, que representa o brilho ou a luminosidade da definição da cor. As cores nas bordas do diagrama são ditas completamente saturadas. À medida que nos aproximamos do centro do diagrama, mais luz branca é adicionada e menor é a saturação.

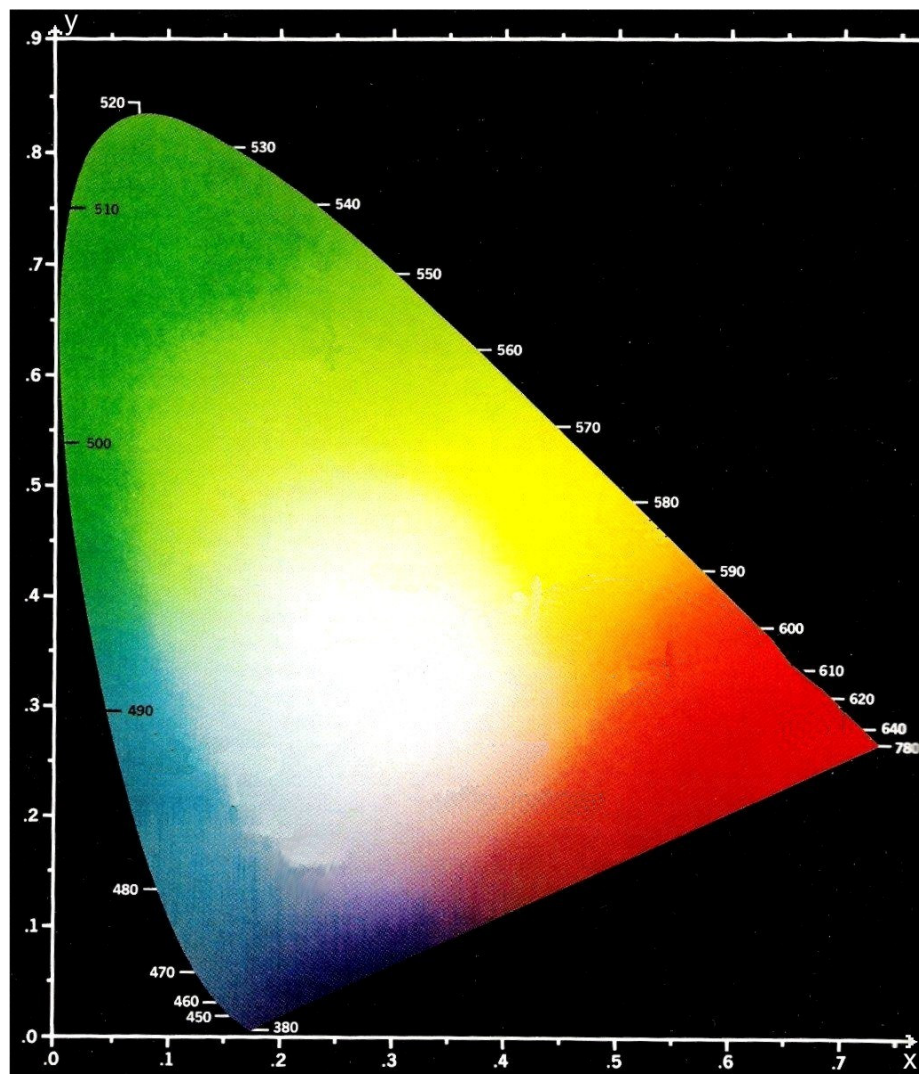


Figura 13. Diagrama de cromaticidade.

(Figura retirada do livro “Digital Image Processing”, R. GONZALEZ e , P. WINTZ, ed. Addison-Wesley Publishing Company, Inc., 2ª edição, 1987)

2.2.3 Histograma

O histograma é um vetor de frequência de cores. No histograma é calculada a soma de vezes que uma cor aparece na imagem. Sendo assim, o histograma de uma imagem é uma matriz que relaciona uma cor ao número de vezes que ela está presente em uma imagem.

No caso do histograma de uma imagem colorida, essa informação é apresentada para cada tom das componentes RGB da imagem. Ou seja, para cada componente R, G e B tem-se o número de vezes que este componente se apresenta na imagem.

É possível representar graficamente o histograma de uma imagem, apresentando sua distribuição de cores. Com o gráfico da distribuição de cores da imagem é possível identificar mais facilmente as faixas de cores que estão mais presentes. A maioria das ferramentas gráficas de qualidade exibe o gráfico do histograma, e, além das curvas R, G e B, oferecem, também, distribuições como: saturação, matiz e escala de cinza.

Quando olhamos um histograma da esquerda para a direita, estamos indo dos valores de menor intensidade (cores escuras próximas ao preto) para o valor de maior intensidade (cores claras próximas ao branco).

Na Figura 14, está apresentado um exemplo de uma figura escura Figura 14.a e seu histograma Figura 14.b. Nota-se, como esperado, que as cores do histograma concentram-se mais à esquerda do histograma. Já na Figura 14.c temos uma figura clara, e diante disso, percebemos que em seu histograma na Figura 14.d as cores se concentram principalmente à direita do mesmo.

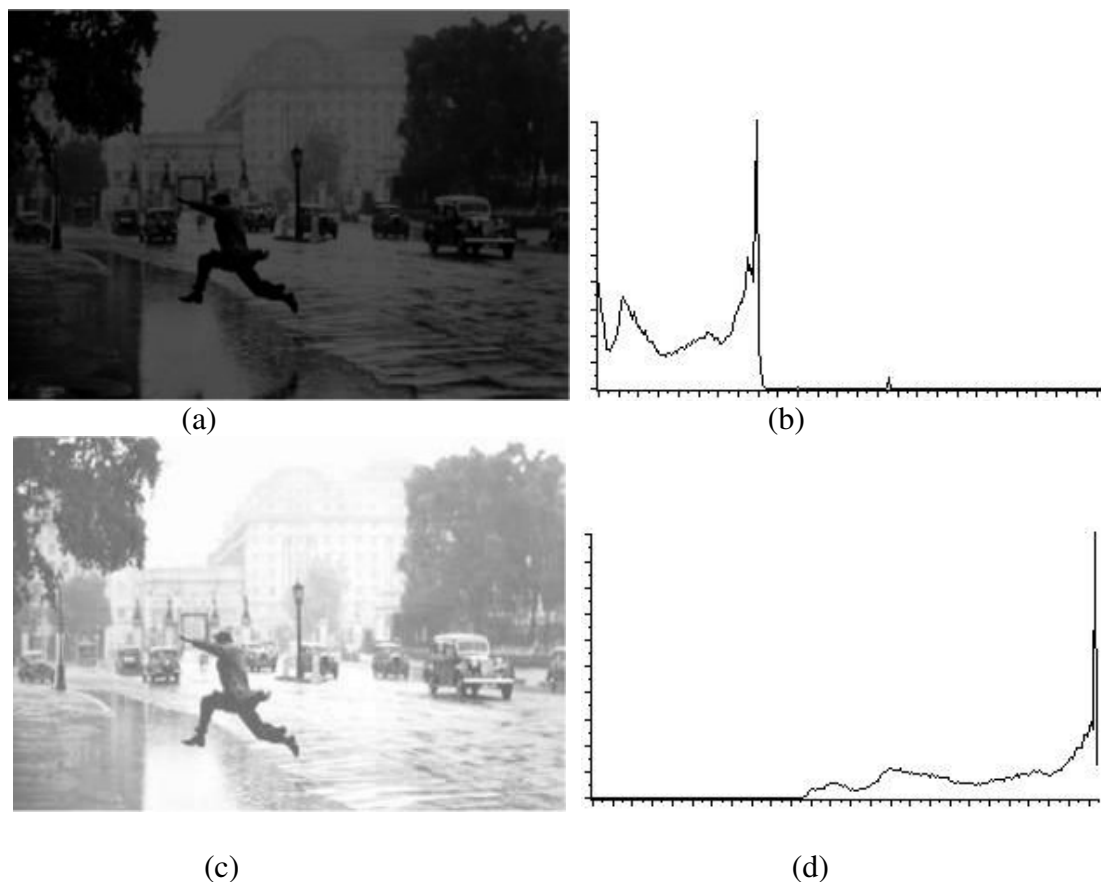


Figura 14. Exemplo de imagens claras e escuras, e seus respectivos histogramas. (a) Imagem escura e (b) seu histograma e (c) imagem clara e (d) seu histograma.

A seguir, na Figura 15, é apresentada uma imagem colorida e seu histograma RGB. É possível notar uma elevação das curvas nas extremidades do histograma, o que indica que a imagem é formada principalmente de áreas claras (próximas ao branco) e escuras (próximas ao preto), o que pode ser comprovado visualmente.

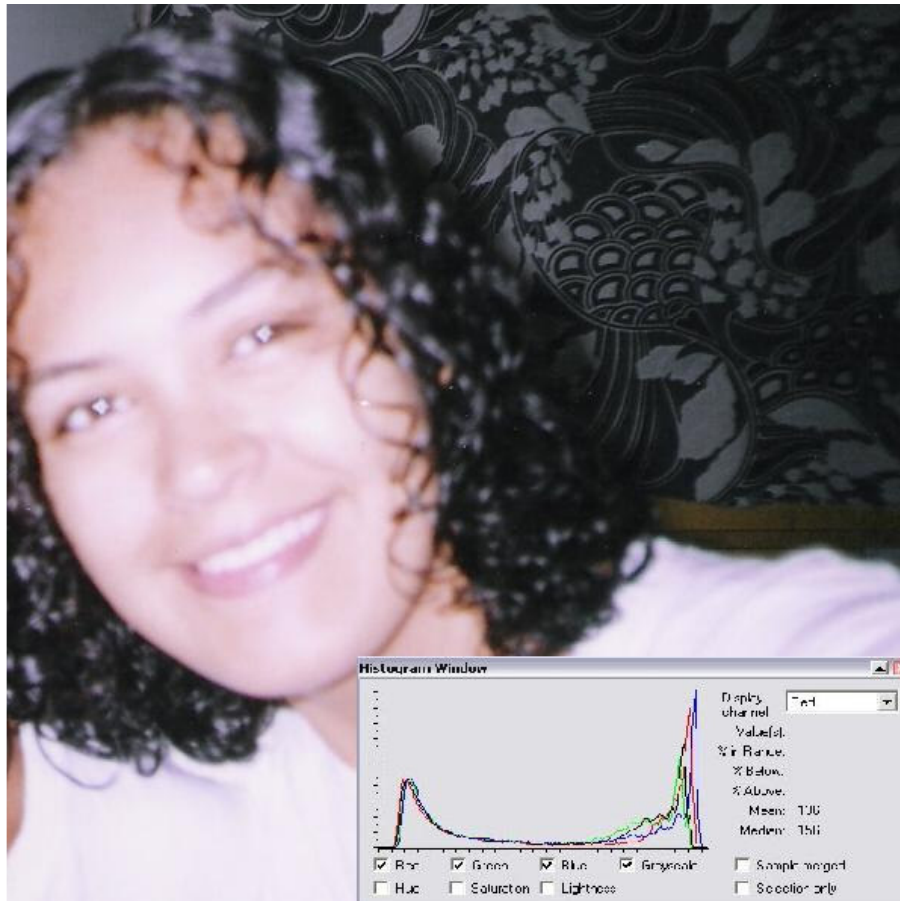


Figura 15. Exemplo de imagem com seu respectivo histograma.

O contraste de uma imagem é manipulado com o objetivo de melhorar a qualidade da visual da mesma, apesar dos resultados serem subjetivos, notamos que numa imagem com alto contraste as informações são muito mais perceptíveis que numa imagem com baixo contraste, veja o exemplo mostrado na Figura 16.

É possível avaliar o contraste de uma imagem a partir da observação de seu histograma. Em imagens com alto contraste, Figura 16.a, as cores do histograma se encontram espalhadas ao longo de todo o intervalo da escala, como mostrado na Figura 16.b; por outro, lado é possível notar que em imagens com baixo contraste, Figura 16.c, apresentam uma concentração das cores em uma determinada faixa do histograma, vide Figura 16.d [15].

Devido a essa relação entre o contraste de uma imagem e sua distribuição de cores ao longo do histograma de frequência, um realce automático do contraste pode ser feito facilmente através da equalização do histograma, a exemplo da Figura 17. Esse processo visa homogeneizar um dado histograma [13].

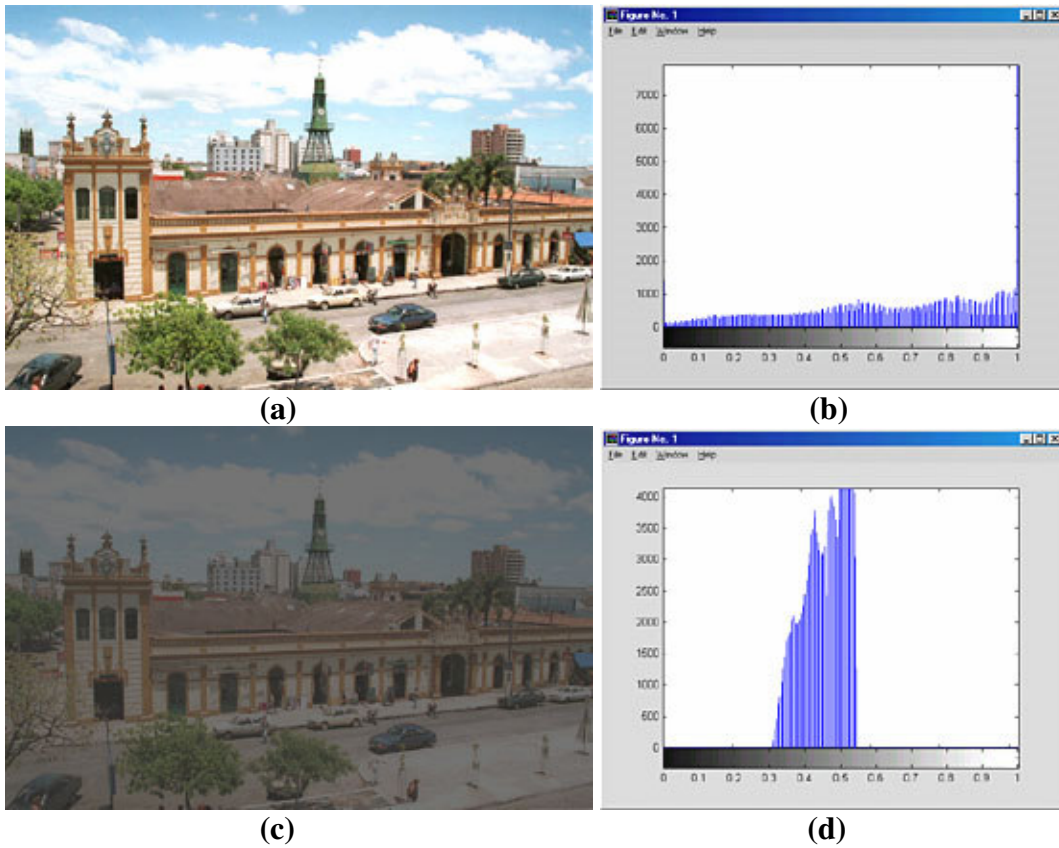


Figura 16. Imagem com alto e baixo contraste, e seus respectivos histogramas. (a) Imagem com alto contraste e (b) seu histograma, (c) Imagem com baixo contraste e (d) seu histograma.

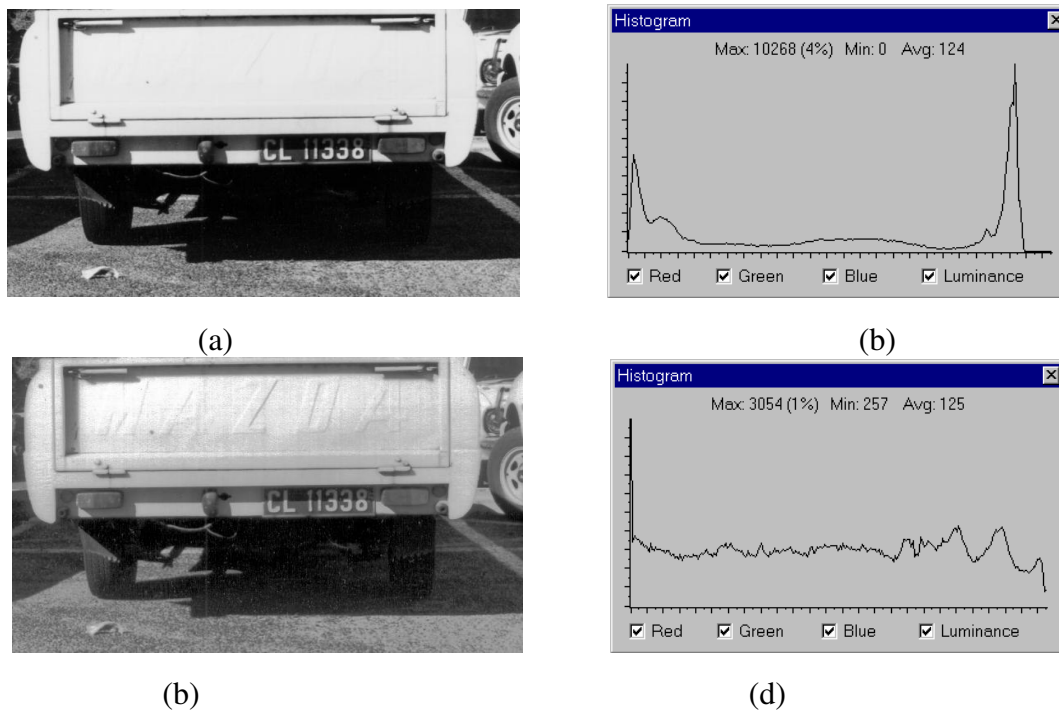


Figura 17. Equalização de histograma. (a) Imagem original e (b) seu histograma, (c) Imagem equalizada e (d) seu histograma.

Capítulo 3

Segmentação

Segmentar uma imagem consiste em reconhecer e subdividir as distintas regiões que a constituem como, por exemplo, separar diferentes objetos numa cena [16]. Na fase de segmentação não existe ainda nenhuma preocupação a respeito da identificação desses objetos.

A segmentação, na maioria dos casos, é apenas um pré-processamento das imagens que servirão de entrada para alguma das diversas aplicações de processamento de imagens e visão computacional, seja no reconhecimento, representação ou interpretação semântica dessas imagens.

O fato de ser apenas um dos passos iniciais para análise das imagens que serão tratadas não diminui a importância da segmentação. Essa é uma das fases mais difíceis e determinantes, podendo ser responsável pelo sucesso ou fracasso de uma determinada aplicação.

Decorrente principalmente da escassez dos recursos computacionais, até pouco tempo, os algoritmos de segmentação se restringiam a imagens em tons de cinza. Porém, nas últimas décadas, com os avanços dos computadores, as pesquisas na área de segmentação passaram a se concentrar em algoritmos para segmentação de imagens coloridas. Mesmo com o crescente aumento da velocidade de processamento dos computadores, o processamento de imagens coloridas requer muito mais tempo que o necessário para imagens em tons de cinza [16].

Existe na literatura uma série de técnicas de segmentação vastamente conhecidas; a maioria delas se baseia na busca por descontinuidades ou similaridades na imagem [13]. A seguir, será feita uma breve explanação a respeito dessas duas abordagens e algumas das técnicas que são utilizadas em cada uma. Além dos algoritmos que utilizam essas duas abordagens, existem ainda outros, como os que se utilizam de algum mecanismo de inteligência artificial como redes neurais artificiais, por exemplo [16].

O melhor algoritmo a ser usado dependerá fundamentalmente do escopo da aplicação. Não existe um algoritmo de segmentação que seja ótimo para todas as aplicações; a maioria deles é adequada a algum tipo específico de caso.

3.1 Segmentação por detecção de descontinuidade

As técnicas baseadas em descontinuidade percorrem a imagem em busca de mudanças abruptas da intensidade dos pixels. São exemplos de técnicas que utilizam descontinuidade a detecção de pontos, linhas e bordas na imagem.

3.1.1 Detecção de pontos

A detecção pontos [16] consiste em reconhecer pontos isolados numa imagem com fundo homogêneo. Na maioria das vezes a detecção de pontos é feita através da aplicação de máscaras similares à mostrada na Figura 18.

Máscaras são filtros que agem sobre a imagem por meio de convolução discreta, esses filtros são representados por uma matriz, geralmente quadrada. Cada pixel e seus oito vizinhos (vizinhança 8-conectada) são multiplicados por uma constante, e a soma destes valores representa a resposta deste pixel.

-1	-1	-1
-1	8	-1
-1	-1	-1

Figura 18. Exemplo de máscara para detecção de pontos.

Para esse tipo de detecção, utiliza-se um filtro passa-alta, e ao valor central da máscara atribui-se o modulo da soma dos outros valores, sendo esses outros valores iguais e negativos. Dessa forma, quando a máscara é aplicada a um pixel do fundo da imagem, o valor encontrado é igual ou bem próximo a zero, enquanto que quando aplicada a um ponto isolado, cuja intensidade é maior que a do fundo da imagem, o valor resultante é bem maior que zero. Uma das grandes dificuldades da detecção de pontos ocorre quando existe a presença de ruídos na imagem. A Figura 19 mostra um exemplo da aplicação da máscara descrita na Figura 18.

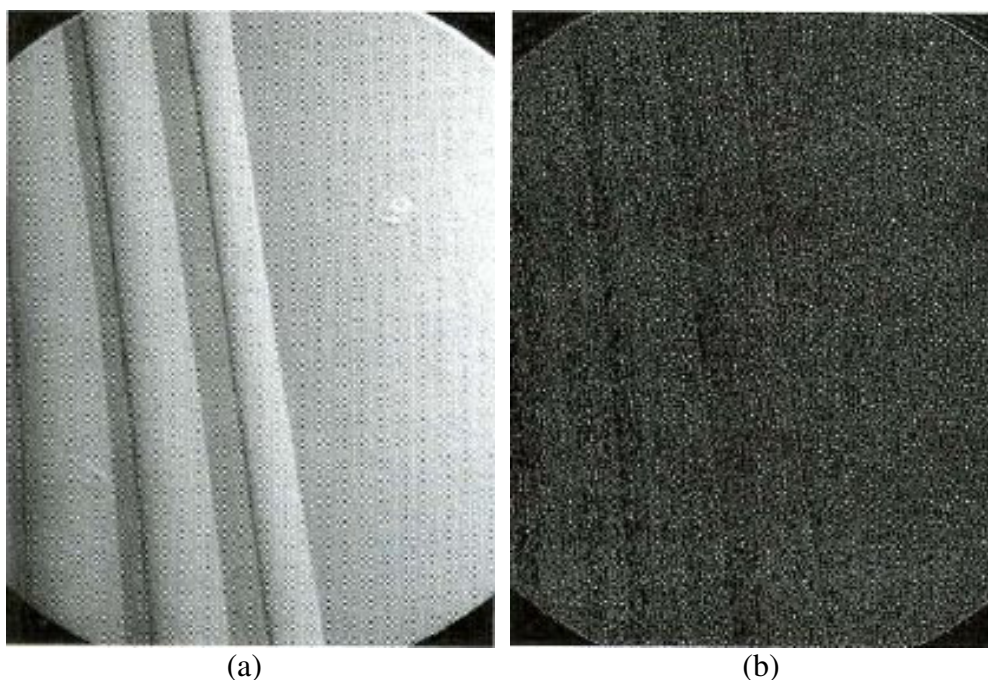


Figura 19. Exemplo de aplicação da técnica de detecção de pontos. (a) Imagem antes da aplicação do filtro e (b) imagem após a aplicação do filtro.

Os dois filtros mais comuns na detecção de bordas são os operadores de Sobel e Laplaciano [13,14], que podem ser vistos na Figura 22. O operador de Sobel encontra os gradientes tanto na vertical, utilizando a máscara na Figura 22.a, quanto na horizontal, com uma máscara cuja matriz é a transposta desta. Já o operador de derivada de segunda ordem (Laplaciano) não faz distinção em relação a direção da borda.

-1	0	1
-2	0	2
-1	0	1

(a)

0	1	0
1	-4	1
0	1	0

(b)

Figura 22. Máscaras correspondentes ao operador de Sobel (a) e Laplaciano (b).

Na Figura 23, a seguir, podemos observar o resultado da aplicação dos operadores de Sobel (a) e Laplaciano (b) a uma imagem (a).

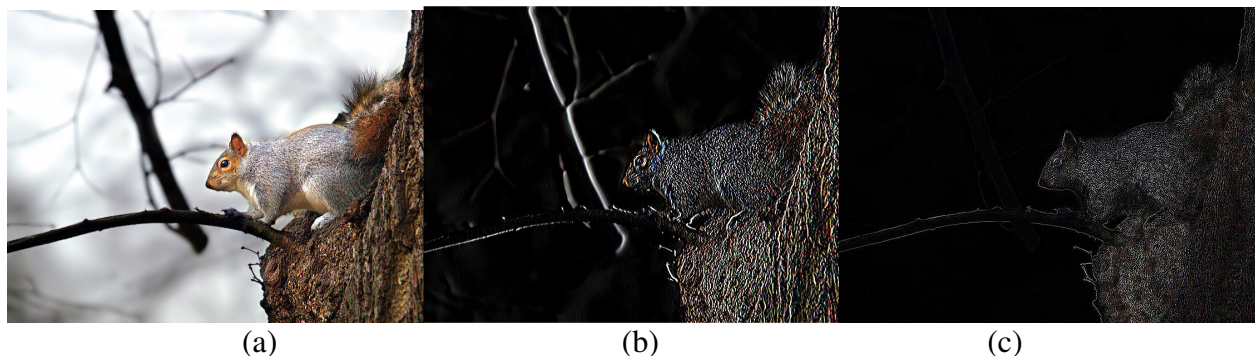


Figura 23. Exemplo de aplicação da técnica de detecção de bordas. (a) Imagem original, (b) aplicação do operador de Sobel e (c) aplicação do operador Laplaciano

3.2 Segmentação por detecção de continuidade

Existe um número maior de técnicas que, ao invés de buscar as mudanças entre pixels vizinhos na imagem, procuram justamente o contrário, regiões similares. As principais técnicas baseadas no princípio de similaridade são: limiarização, crescimento de regiões e *split* e *merging* de regiões.

3.2.1 Limiarização

Limiarização [17,19,20,21], também conhecida como binarização, é uma das mais simples e usadas técnicas de segmentação de imagens, consistindo na determinação de um ou mais valores de corte, podendo ser, dessa forma, biníveis ou multiníveis. No modo binível é feita apenas a segmentação entre a imagem de fundo e um objeto; no modo multinível é feita a segmentação entre fundo e vários objetos.

O valor de corte ou limiar representa uma tonalidade intermediária entre as regiões que se deseja separar. A maior dificuldade dos algoritmos de limiarização está justamente na determinação deste valor de corte. Uma vez encontrado este valor limiar, separa-se através de um

processo iterativo, os pixels com tonalidade maior dos pixels com tonalidade menor ao limiar, subdividindo assim, a imagem em duas.

Existem inúmeras implementações de algoritmos de limiarização; a mais simples delas, usadas em ferramentas como o *Photoshop*TM da Adobe Systems [22], consiste em determinar um valor fixo de corte, que dividirá a imagem em duas categorias: os pixel com cor acima (convertidos para branco) ou abaixo do ponto de corte (convertidos para preto).

Outros algoritmos de limiarização se utilizam de informações estatísticas da imagem para determinar o valor de corte, algumas das medidas estatísticas mais utilizadas são: média, moda, mediana, primeira e segunda derivada, e entropia.

Na figura 24, usou-se limiarização para eliminar o fundo do cheque [23].

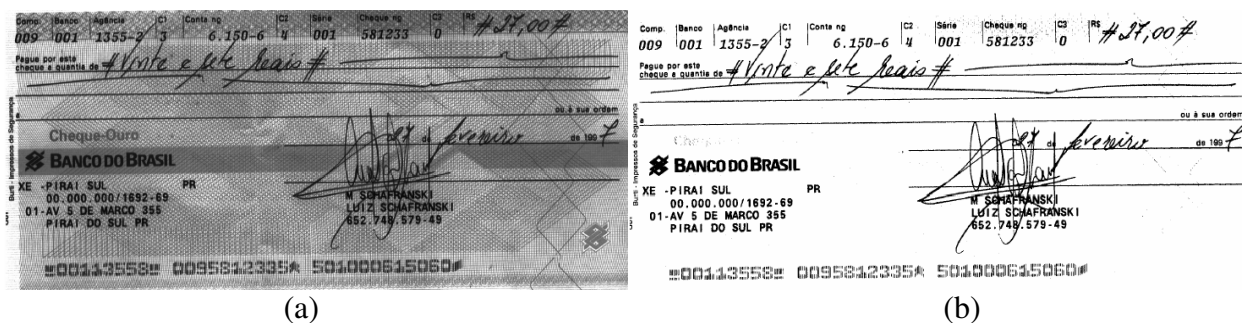


Figura 24. Exemplo de aplicação de limiarização. (a) Imagem original de um cheque e (b) imagem do cheque com fundo eliminado através da aplicação de um algoritmo de limiarização.

Dentre os algoritmos de limiarização destaca-se o algoritmo de Otsu por ser até hoje um dos algoritmos de limiarização mais referenciados [19]. Este algoritmo se baseia na análise do discriminante. Os pixels da imagem são classificados e separados em duas classes, que representam o objeto e fundo da imagem.

O algoritmo de Otsu faz a segmentação usando informações estatísticas como média e variância das duas classes. Através dessas informações é calculado o valor do limiar de corte da imagem. Esse algoritmo apresenta melhor resultado quando o número de pixels das duas classes é próximo [21].

3.2.2 Crescimento de regiões

Os algoritmos de crescimento de regiões [16,17,18] elegem um ou mais pixels como “sementes” e através de um dado critério de homogeneidade os pixels vizinhos a “semente” são classificados como similares ou não. Se forem similares, passam a possuir o mesmo rótulo e a pertencer à mesma região.

À medida que se percorre a matriz de pixels da imagem, as regiões semelhantes vão crescendo e se agrupando em torno da “semente”, obedecendo sempre o critério de semelhança pré-determinada. As diversas implementações dessa técnica variam principalmente em relação aos critérios de escolha das “sementes” e do critério de homogeneidade que são adotados. Na Figura 25, a imagem (a) foi previamente binarizada obtendo-se a imagem (b) que foi usada como semente, na imagem (c) podemos ver o resultado do crescimento ao redor da semente, e em (d) temos o resultado final da segmentação.

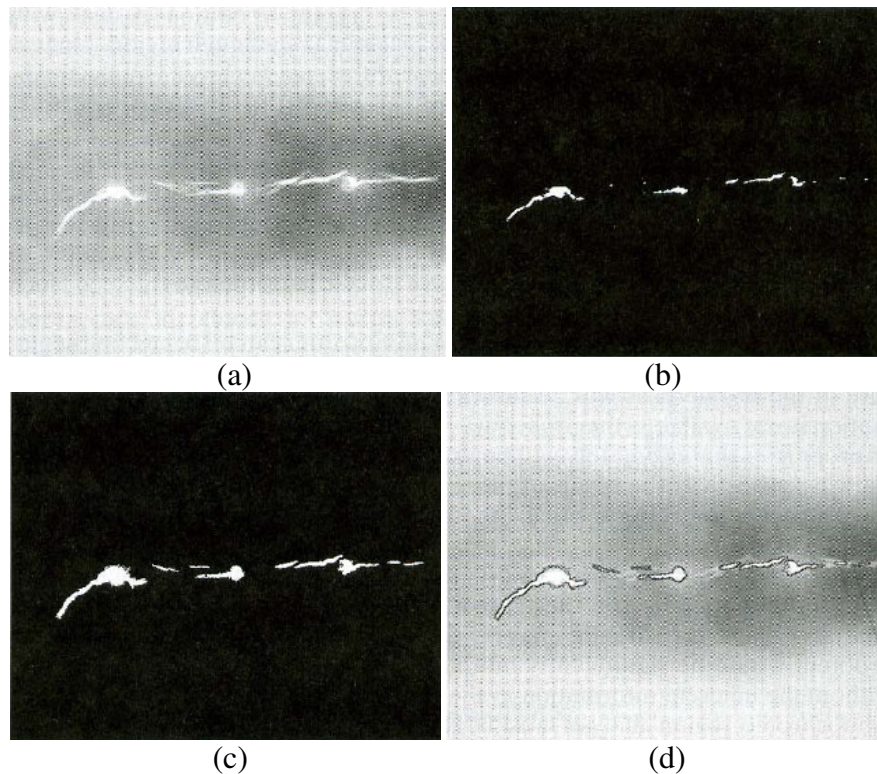


Figura 25. Aplicação do algoritmo de crescimento de regiões. Imagem original (a), semente (b), resultado do crescimento (c) e Imagem segmentada (d).

3.2.3 Split e merge de regiões

Nessa técnica [13,17], a imagem é recursivamente particionada, subdividindo-se arbitrariamente ou de acordo com algum critério pré-estabelecido até que cada parte subdividida da imagem seja homogênea. Após essa fase, conhecida como *split*, tem-se uma série de fragmentos de imagens que necessitam agora ser conectadas da maneira adequada.

Após a fase de *split*, inicia-se a fase de *merge*, responsável por conectar regiões vizinhas garantindo que a homogeneidade seja preservada, classificando dessa forma, as partes vizinhas semelhantes como fazendo parte da mesma região. O processo continua até que o número máximo de segmentos sejam conectados. Na Figura 26.a é aplicada a técnica de *Split e Merge* obtendo a Figura 26.b.

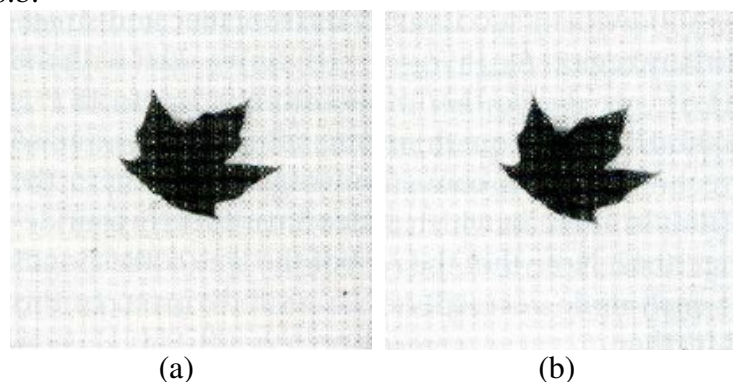


Figura 26. Aplicação do algoritmo de *Split e Merge*. (a) Imagem original e (b) resultado do processo de *Split e Merge* de regiões.

3.3 Segmentação de documentos históricos

A segmentação de documentos históricos se caracteriza pela separação desse documento em duas classes, tinta e textura; ou seja, conteúdo escrito e fundo do papel.

Apesar de só haver duas classes a serem separadas, a segmentação de documentos encontra uma série de dificuldades cujo grau varia de acordo com o acervo utilizado. Os documentos aqui manipulados datam do final do século XIX e início do século XX, período este em que o papel fabricado apresentava uma alta alcalinidade, o que fez os documentos adquirirem um aspecto amarelado mais rapidamente. Na Figura 27 é possível ver um desses documentos.

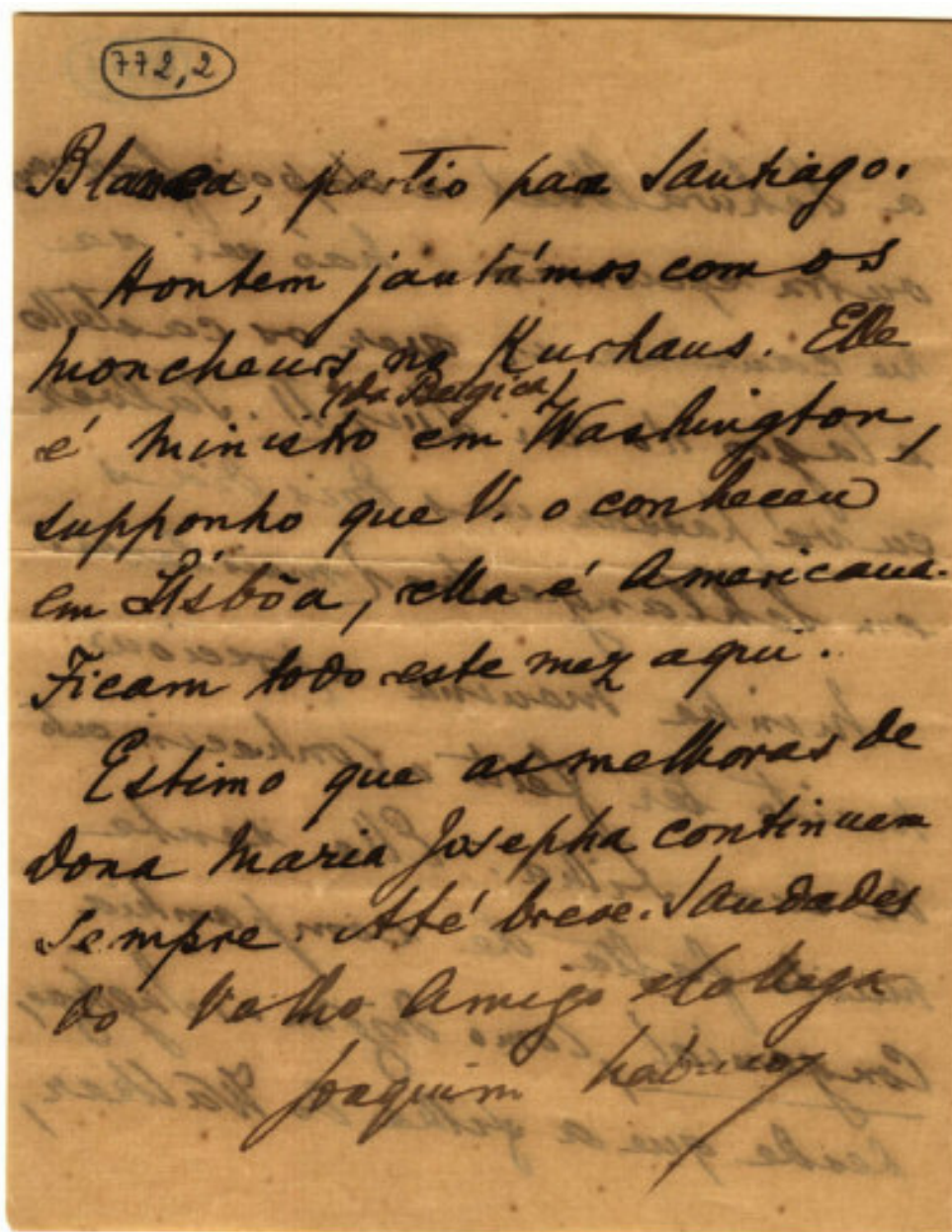


Figura 27. Exemplo de documento histórico.

Outros fatores como esmaecimento da tinta ou sua transposição do verso para frente do papel, sujeira, fungos, etc também contribuem de forma negativa no processo de segmentação.

Devido ao fato de se tratar da segmentação de apenas duas classes bem singulares, tinta e papel, a técnica utilizada na segmentação de documentos é, na maioria dos casos, feita por limiarização. Um estudo comparativo entre técnicas de limiarização foi feito em [24], onde nenhuma delas conseguiu gerar um resultado satisfatório para o conjunto de imagens testado.

Dentre os algoritmos comparados estão o algoritmo de Otsu, que como mencionado anteriormente é um dos algoritmos de segmentação ainda mais utilizado, e três algoritmos baseados na entropia da imagem, são eles: Algoritmo de Pun, algoritmo de Kapur-Sahoo-Wong e o algoritmo de Johannsen e Bille.

A entropia foi definida por Claude E. Shannon em 1948 como sendo uma medida de informação. A teoria da informação desenvolvida por Shannon, procura quantificar a quantidade de informação (I) envolvida no acontecimento de um determinado evento (E). Tendo $p(E)$ como a probabilidade do evento E ocorrer, a quantidade de informação recebida quando esse evento ocorre é dada pela Eq. 4.

$$I = \log\left(\frac{1}{p(E)}\right) \quad (\text{Eq. 4})$$

Pode-se observar que um evento com probabilidade 1 de ocorrer terá um valor nulo para a informação associada a este evento, visto que o evento certamente acontecerá, não existe nenhuma surpresa ou informação associada a ele. Sendo assim, quanto menor for a probabilidade de ocorrer um determinado evento, maior será a surpresa caso ele ocorra e conseqüentemente, mais informação será obtida com isto.

Suponha uma imagem (X) com n cores, utiliza-se o índice (i) que varia de 1 a n , para indicar uma determinada cor da imagem, e a probabilidade dessa cor na imagem é dada por $p(X_i)$. Sendo assim, a entropia, medida em bits/cores, associada a imagem é dada pela Eq. 5.

$$H = -\sum_{i=1}^n p(x_i) \log[p(x_i)] \quad (\text{Eq. 5})$$

O algoritmo de Pun, primeiramente são calculadas as entropias dos pixel pertencentes a tinta (H_b) e do fundo (H_w), dadas por Eq. 6 e Eq. 7, respectivamente.

$$H_b = -\sum_{i=0}^t p(x_i) \log[p(x_i)] \quad (\text{Eq. 6})$$

$$H_w = -\sum_{i=t+1}^{255} p(x_i) \log[p(x_i)] \quad (\text{Eq. 7})$$

As entropias H_b e H_w , são delimitadas pelo um valor de corte (t), que está entre 0 e 255 – considerando uma imagem com 256 tons de cinza. O valor de t deve ser tal que maximize a equação da entropia, dada como a soma das entropias H_b e H_w , como pode ser verificado na Eq. 8.

$$H = H_b + H_w \quad (\text{Eq. 8})$$

O algoritmo de Kapur-Sahoo-Wong, assim como no de Pun, busca um valor de t que maximize a Eq. 8. No entanto, agora as probabilidades são definidas por uma distribuição de probabilidade para a tinta e para o fundo da imagem, utilizando a probabilidade e a probabilidade acumulada de t , como pode ser visto na Eq. 9 e na Eq. 10.

$$H_b = -\sum_0^t \frac{p(x_i)}{P_t} \log\left(\frac{p(x_i)}{P_t}\right) \quad (\text{Eq. 9})$$

$$H_w = -\sum_{t+1}^{255} \frac{p(x_i)}{1-P_t} \log\left(\frac{p(x_i)}{1-P_t}\right) \quad (\text{Eq. 10})$$

O algoritmo de Johannsen e Bille é uma outra variação de algoritmos de segmentação baseados em entropia, desta vez, busca-se um limiar t que minimize a Eq. 11, onde S_b e S_w são dados pela Eq. 12 e Eq. 13, e $E(x)$ correspondente a Eq. 14.

$$S(t) = S_b(t) + S_w(t) \quad (\text{Eq. 11})$$

$$S_b(t) = \log\left(\sum_{i=0}^t p_i\right) + \left(\frac{1}{\sum_{i=0}^t p_i}\right) \left[E(p_t) + E\left(\sum_{i=0}^{t-1} p_i\right) \right] \quad (\text{Eq. 12})$$

$$S_w(t) = \log\left(\sum_{i=t}^{255} p_i\right) + \left(\frac{1}{\sum_{i=t}^{255} p_i}\right) \left[E(p_t) + E\left(\sum_{i=t+1}^{255} p_i\right) \right] \quad (\text{Eq. 13})$$

$$E(x) = -x * \log(x) \quad (\text{Eq. 14})$$

O fracasso da aplicação desses algoritmos nesse conjunto de imagens se deve principalmente à interferência frente-verso existente em muitas das imagens. Na Figura 28, podemos visualizar alguns documentos e os resultados da aplicação dessas técnicas.

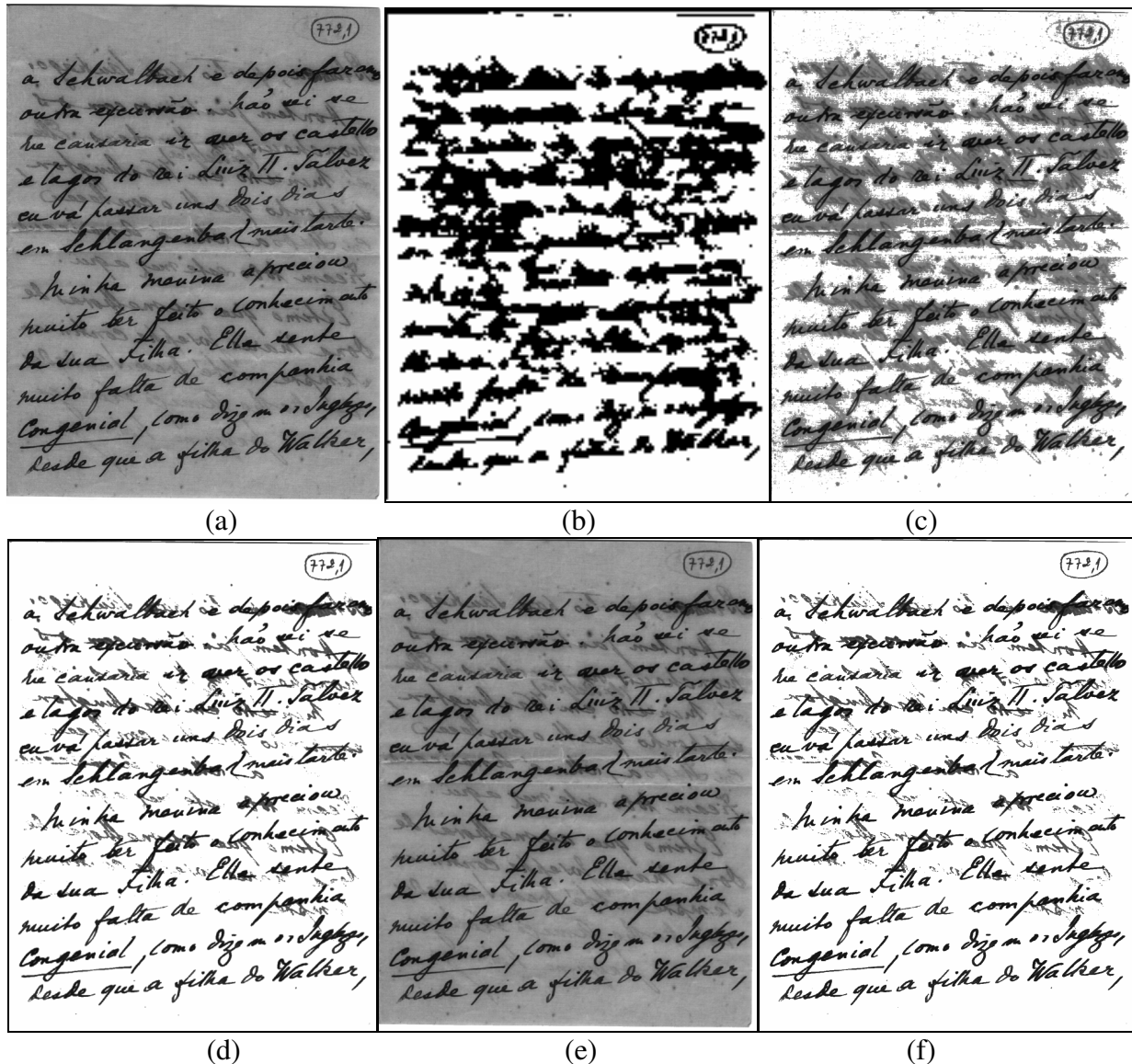


Figura 28. Comparação de algoritmos de segmentação a imagem. (a) Software Photoshop com valor de corte de 127 (b), Pun (c), Kapur-Sahoo-Wong (d), Johannsen e Bille (e) e Otsu (f).

Diante dos resultados apresentados, foi desenvolvido um novo algoritmo de segmentação baseado em entropia [25].

Visto que a maior parte dos documentos históricos são formados por pixels relativos ao papel, a cor mais freqüente na imagem corresponde a um elemento da textura do papel, e foi tomada como separador entre as entropias do papel e da tinta. O cálculo da entropia no novo algoritmo de segmentação usa uma base igual ao número de pixels da imagem. A entropia de 500 imagens do acervo foi calculada e pôde-se observar que seus valores variavam entre 0,2 e 0,4. Diante das imagens e suas respectivas entropias, foi possível classificá-las em três grandes categorias (Figura 29):

1. Documentos claros ou com pouco texto, cuja entropia é menor que 0,25;
2. Documentos mais comuns, cuja entropia varia entre 0,25 e 0,3;
3. Documentos escuros ou com muita interferência frente-verso, cuja entropia é maior que 0,3.

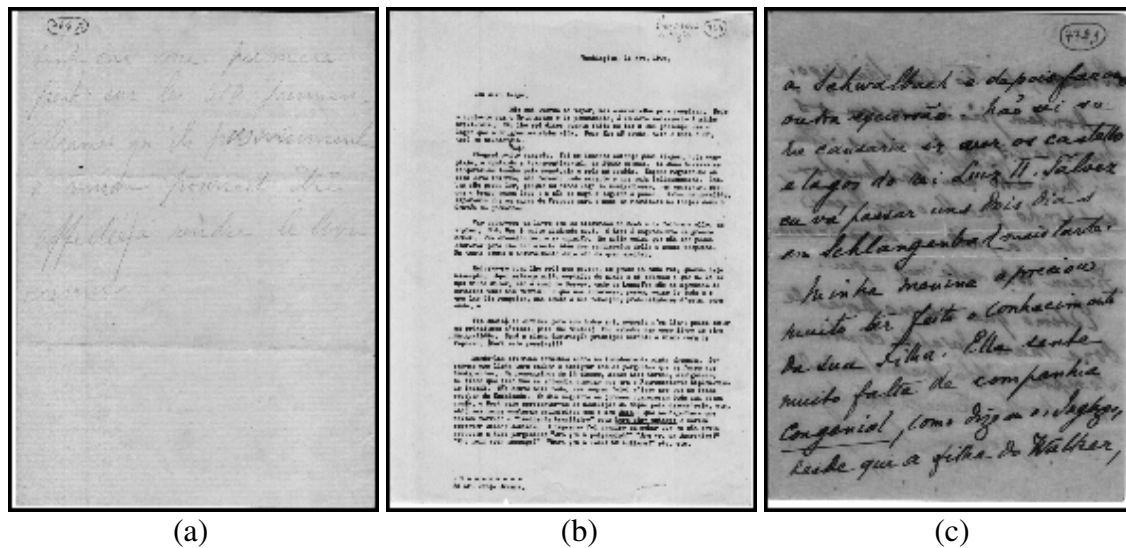


Figura 29. Exemplos das classes do acervo: a) $H=0,22$, b) $H=0,27$ e c) $H = 0,32$.

Com o objetivo de dosar o peso da entropia do papel (H_w) e do texto (H_b) durante a segmentação, duas constantes multiplicativas foram usadas, m_w e m_b , variando seus pesos de acordo com a faixa de entropia do documento. Os valores dessas constantes foram calculados empiricamente tendo como parâmetros de avaliação a análise qualitativa das imagens resultantes e o grau de acerto pelas ferramentas OCR a que foram submetidas. Após uma série de experimentos, essas variáveis foram definidas da seguinte forma:

1. Para os documentos pertencentes à primeira categoria, com $H \leq 0,25$, os valores definidos para os fatores m_w e m_b foram, respectivamente, 2 e 3 com o objetivo de aumentar o valor de corte;
2. Para os documentos pertencentes a segunda categoria, $0,25 < H < 0,3$, o valor do fator m_b foi 2,6, com o objetivo de reduzir as interseções, e a entropia da textura não foi modificada, atribuindo-se assim 1 ao m_w ;
3. Para os documentos pertencentes a terceira categoria, $H \geq 0,3$, não houve alteração nos valores das entropias, permanecendo $m_w = m_b = 1$.

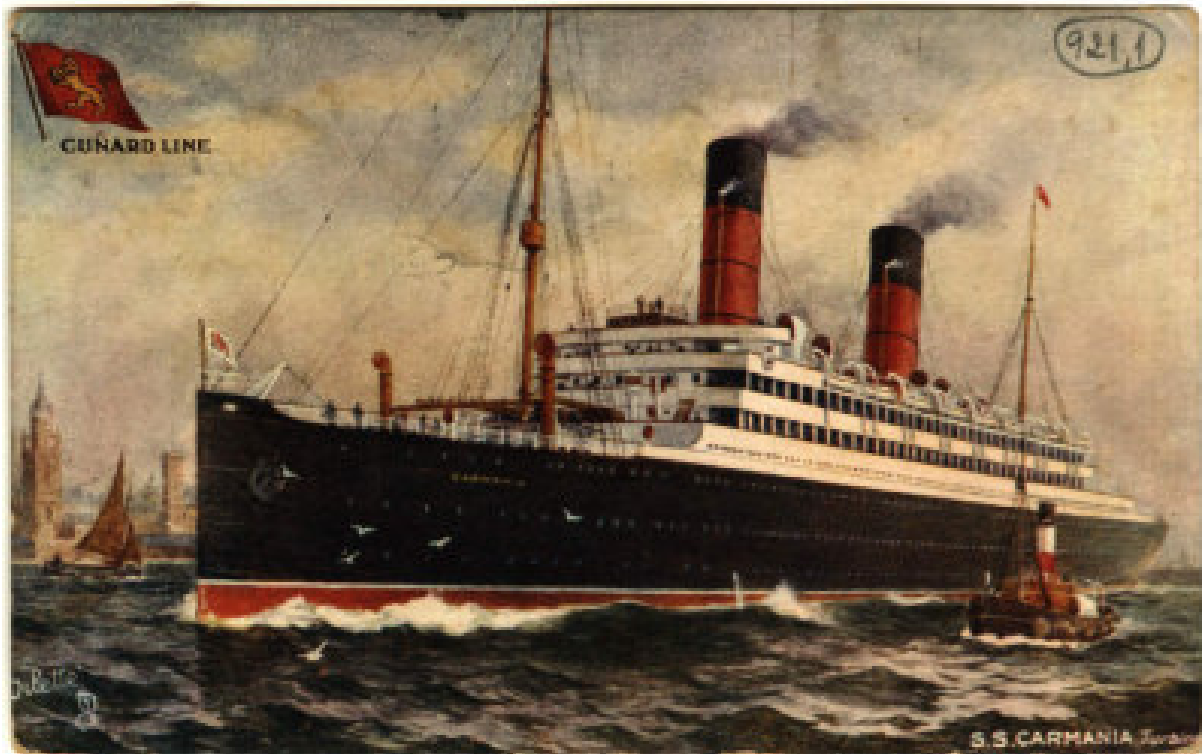
O novo algoritmo proposto em [25] apresentou bons resultados, não apenas quando aplicado a documentos históricos (tanto em tons de cinza como coloridos), mas também com cartões postais, o que demonstra o potencial e versatilidade do algoritmo. A seguir (Figura 30) podemos verificar alguns o resultado da aplicação deste novo algoritmo a imagem da Figura 28. a). E na Figura 31, é mostrada a aplicação desse algoritmo a uma imagem de um cartão postal colorido.

(772,1)

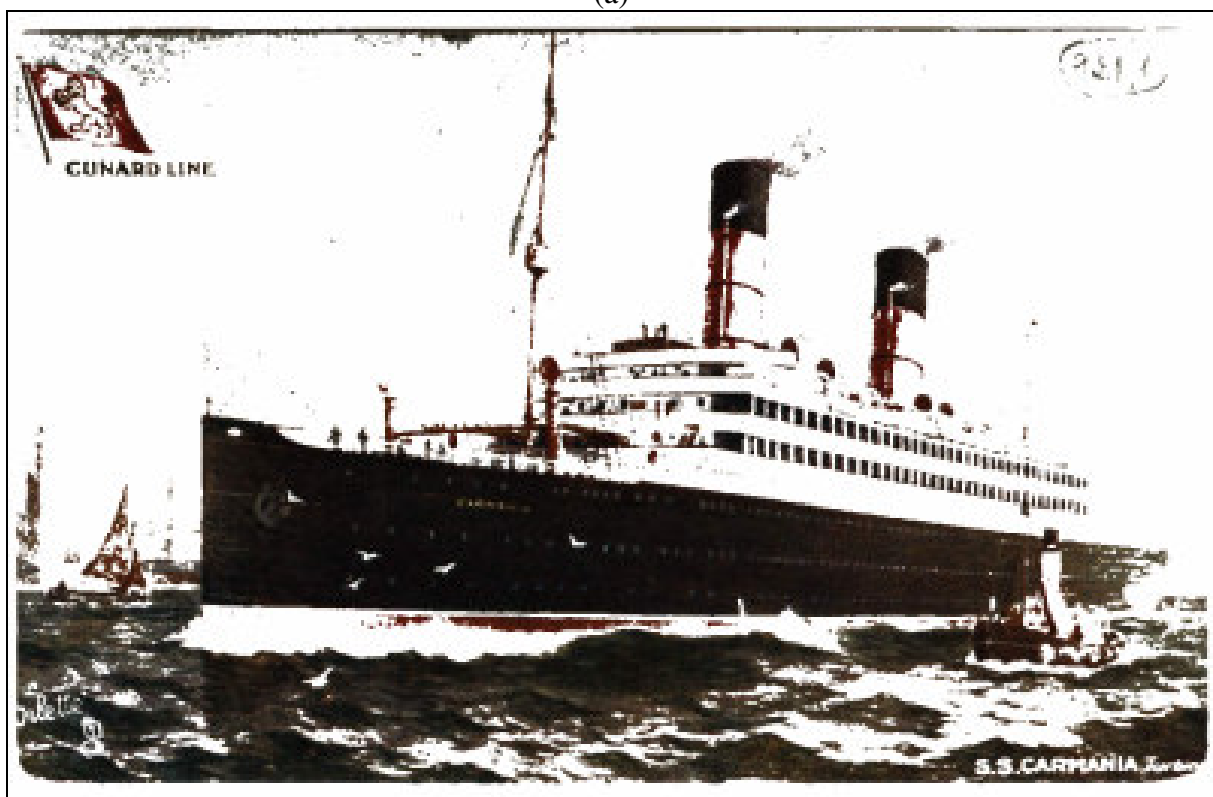
a. Schwaltach e depois fazem
outra experiência: há-se se
de causaria ir quer os castello
e lago do rei Luiz II. Salvo
cu va' passar uns dois dias
em Schlangenthal mais tarde.

Minha mãe me apreciou
muito ter feito o conhecimento
da sua filha. Ela sente
muito falta de companhia
Congenial, como dizem os Ingleses,
desde que a filha do Walker,

Figura 30. Novo algoritmo aplicado à imagem da Figura 28.a).



(a)



(b)

Figura 31. Aplicação do algoritmo a imagens coloridas de cartões postais. Imagem original e b) Imagem segmentada.

Capítulo 4

Textura

Apesar da inexistência de uma definição formal acerca do que vem a ser textura, esta se caracteriza, principalmente, pela repetição de um padrão ou mais em uma região, e tem como principal aplicação conferir um caráter mais real e natural a modelos artificiais [1].

Quando lidamos com documentos históricos, o papel envelhecido devido à ação do tempo, apresenta uma tonalidade mais escura, geralmente amarelada. Objetivando produzir uma representação sintética o mais fiel possível desses documentos, não seria natural representá-los com um fundo branco. Aplicar texturas que simulem características do papel original é, dessa forma, fundamentalmente importante para conceder caráter mais verossímil ao documento.

4.1 Análise de textura

A análise da qualidade de imagens exerce um papel primordial no processamento de imagens, tendo como principal objetivo oferecer subsídios para comparação e avaliação dos resultados gerados.

Durante muito tempo avaliou-se a qualidade das imagens por meio da média dos resultados obtidos a partir de pesquisa de opinião (Mean Opinion Score - MOS). Nessa pesquisa, cada um escolhe um número de 1 à 5 para determinar a qualidade da imagem, seguindo a correspondência: (1) ruim; (2) pobre; (3) justo; (4) bom; (5) excelente. Tira-se então a média das notas individuais, que pode variar de 1 (no pior caso) à 5 (no melhor caso).

Talvez essa fosse a forma mais fácil e adequada de assegurar a qualidade dessas imagens. Grande parte dos algoritmos de processamento de imagens objetiva empreender algum tipo de melhoramento às imagens originais, e como essas imagens, na maioria das vezes, são geradas para serem vistas por outras pessoas, nada mais natural que simplesmente olhar para as duas imagens e analisá-las visualmente [26].

Apesar de ser uma forma simples e natural de avaliação, o MOS pode ser bastante inconveniente quando um número grande de imagens precisa ser comparado (tornando o processo muito lento), além de esse processo estar sujeito a medidas subjetivas de qualidade. Para suprir as deficiências do MOS, outras formas de analisar texturas foram desenvolvidas. Veremos algumas delas a seguir.

4.1.1 Medidas matemáticas

Devido à necessidade de formas rápidas e automáticas de se obter métricas qualitativas, o Erro Médio Quadrático (EMQ) e a Razão Sinal-Ruído de Pico (*Peak Signal-to-Noise Ratio* - PSNR) foram, e ainda são, as medidas estatísticas mais utilizadas [26]. O EMQ e o PSNR mensuram, cada um a sua maneira, a média das diferenças entre duas imagens, quanto menor o EMQ, e conseqüentemente quanto maior o PSNR, menos diferença existem entre as duas imagens.

Apesar disto, o EMQ e o PSNR são muito criticados, uma vez que seus valores em si não têm significado algum, só fazendo sentido quando se compara o valor encontrado para imagens produzidas por diferentes algoritmos.

Aplicando-se o EMQ e o PSNR, Eq. 15 e Eq. 16, à imagem original e cada uma das imagens geradas, a imagem mais fiel à original será aquela que obtiver o menor EMQ, e conseqüentemente maior PSNR, já que estas grandezas são inversamente proporcionais.

$$EMQ = \sum_{i=0, j=0}^{i=M, j=N} \frac{(I_1(i, j) - I_2(i, j))^2}{M * N} \quad (\text{Eq. 15})$$

$$PSNR = 10 * \log_{10} \frac{255^2}{EMQ} \quad (\text{Eq. 16})$$

Onde $I_1(i,j)$ é a cor do pixel na posição (i,j) da primeira imagem e $I_2(i,j)$ é a cor da segunda imagem na mesma posição (i,j) e M e N são as dimensões da imagem, linhas e colunas, respectivamente.

4.1.2 Análise de variância

O estatístico R.A. Fisher [27] desenvolveu um procedimento denominado Análise de Variância (ANOVA), com o qual pretendia controlar a rejeição de uma hipótese nula. A hipótese nula nada mais é do que uma hipótese inicial a respeito de um experimento, e que deverá ser ou não rejeitada como uma possibilidade viável. Apesar de ser possível que a hipótese nula envolva qualquer parâmetro, ela geralmente sugere que em média os dois grupos testados são iguais, ou seja, a diferença das médias dos dois grupos é nula ($\mu_1 - \mu_2 = 0$).

Com a ANOVA, mesmo que a hipótese nula seja rejeitada, ela ainda pode ser validada, de acordo com um teste de significância [28]. Nesse teste, é calculada a probabilidade de um resultado diferente da hipótese nula ocorrer. Se essa probabilidade for muito baixa, a hipótese nula é rejeitada. A determinação do que seria uma probabilidade muito baixa é uma escolha arbitrária, contudo, convencionalmente, utiliza-se os valores 0,05 ou 0,01.

A ANOVA pode ser aplicada a uma série de características das imagens [29,30,31], todavia, o método que vem se apresentando como o mais adequado é a matriz de co-ocorrência [32,33,34].

4.1.3 Matriz de co-ocorrência

A matriz de co-ocorrência, por ter sido aplicada durante muito tempo apenas em imagens em tons de cinza, é largamente conhecida como Matriz de Co-ocorrência de tons de cinza (*Grey Level Co-Occurrence Matrix* - GLCM). Para imagens coloridas toma-se três matrizes de co-ocorrência, uma para cada tom no Sistema RGB.

A GLCM contém informações sobre a posição dos pixels que possuem tons de cinza semelhantes, guardando a frequência com que pixels vizinhos e similares são separados por uma certa distância (d) em múltiplas direções (θ) - onde θ pode ser 0° , 45° , 90° ou 135° , e d é um número inteiro pequeno, geralmente 1 ou 2. A matriz de co-ocorrência é uma matriz quadrada, de dimensões iguais ao número de tons possíveis na imagem. Por exemplo, para uma imagem (I) com 256 possíveis tons de cinza, sua matriz de co-ocorrência correspondente terá dimensão de 256×256 .

Tomando um tom por i e outro por j , podemos definir as matrizes de co-ocorrência da seguinte forma [35]:

$$P(i, j, d, 0^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid k - m = 0, |l - n| = d, I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 45^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = -d) \vee (k - m = -d, l - n = d), I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 90^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid |k - m| = d, l - n = 0, I(k, l) = i, I(m, n) = j\}$$

$$P(i, j, d, 135^\circ) = \#\{(k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = d) \vee (k - m = -d, l - n = -d), I(k, l) = i, I(m, n) = j\}$$

Onde, L_x é o espaço horizontal, L_y é o espaço vertical, $L_y \times L_x$ é a resolução espacial, (k, l) é a coordenada do primeiro tom tomado, (m, n) é a coordenada do segundo tom tomado e $\#$ representa o número de elemento no conjunto.

Na Figura 32, temos a matriz de uma imagem e sua matriz de ocorrência para $d = 1$ e $\theta = 0^\circ$. As setas da mesma cor fazem referência à mesma posição na matriz de co-ocorrência. Por exemplo, na posição $(0, 1)$ na matriz de co-ocorrência (ver Figura 32.b) tem a quantidade de pixels da cor zero, separados de uma distância igual a 1 na horizontal, aos pixels da cor 1. As relações que satisfazem esta condição estão mostradas com setas na cor laranja. Como existem duas relações entre as cores 0 e 1, nestas condições, na matriz de co-ocorrência, o valor da posição $(0, 1)$ é igual a 2. É possível observar que essa matriz é simétrica, e que relações entre pixels da mesma cor, são contadas como duas, visto existe uma relação tanto da direita para esquerda, como da esquerda para direita.

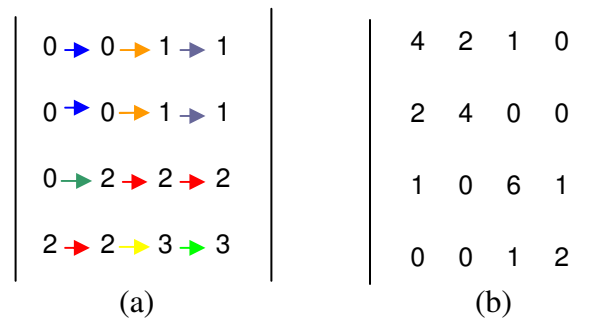


Figura 32. Imagem (a) e (b) sua matriz de ocorrência com $d = 1$ e $\theta = 0^\circ$.

Um dos grandes benefícios na utilização das matrizes de co-ocorrência diz respeito às inúmeras informações que podem ser extraídas delas. Em [35] foram definidas 14 características para texturas de imagens, dentre elas: segundo momento angular, contraste, correlação, variância,

momento diferencial inverso, soma e diferença de variância e diferença de entropia. Essas medidas são largamente utilizadas na análise da qualidade das texturas geradas.

Uma das desvantagens das matrizes de co-ocorrência é o grande espaço que elas ocupam. Uma alternativa seria utilizar a soma ou diferença de histogramas [1], onde se somam ou subtraem-se os histogramas dos pixels a uma certa distância (d) um do outro.

4.1.4 Novos algoritmos de análise de imagens

Em busca de medidas percentuais de qualidade, novos algoritmos de análise de imagens vêm sendo desenvolvidos.

O algoritmo desenvolvido em [36] propõe um novo índice universal para avaliar a qualidade de imagens. Baseado em perda de correlação, distorção de luminância e distorção de contraste, apresentou bons resultados, se mostrando bem mais representativo, por exemplo, que o EMQ. Uma das grandes vantagens deste algoritmo está na apresentação dos resultados entre -1 e 1, onde 1 é o melhor valor e ocorre apenas quando as duas imagens são iguais. Dessa forma, é possível ter uma noção percentual da qualidade das imagens, o que não ocorre quando se utiliza EMQ ou PSNR.

Alguns outros algoritmos vêm incorporando características do sistema visual humano (*Human Visual System* - HVS) [37 - 45]. Uma vez que a principal função do olho humano é extrair informações estruturais do campo de visão observado, uma medida de distorção estrutural poderia ser uma boa aproximação para avaliação de distorções em imagens.

4.2 Novo algoritmo de síntese de textura

Como mencionado anteriormente, as texturas dos documentos trabalhados são, em sua grande maioria, amareladas, assim, o tom azul tem pouca influência sobre as cores presentes na imagem. Baseamo-nos em um algoritmo para geração automática de texturas coloridas de papel, a partir dos modelos dos histogramas RGB da textura original [3], para desenvolver um novo algoritmo. O algoritmo leva em consideração essas características, inerente ao próprio papel do acervo trabalhado. Nesse sentido, optamos por fazer um sistema que gera automaticamente os histogramas dos tons vermelho e verde, e, no lugar do histograma de azul, utilizar os histogramas da luminância - cujo valor corresponde, também, ao tom de cinza (Gr) associado a cada pixel [5].

A textura original do documento é obtida durante a segmentação - fase em que a imagem original é separada em duas, uma correspondente à tinta e outra correspondente à textura do papel. Na imagem correspondente à textura, após a limiarização, onde anteriormente havia a tinta do papel, existe agora a cor branca. De acordo com as necessidades da aplicação, pode-se fazer uso de uma série de algoritmos de preenchimento [3] a fim de preencher/completar tais regiões de branco.

Para geração da textura sintética são extraídas informações estatísticas de uma amostra da textura da imagem do documento original. Dentre essas informações estão: média, desvio-padrão, histograma, entropia e matiz mais comum na imagem. Este conjunto de informações é armazenado em um arquivo binário, que ocupa aproximadamente 80 bytes. Vale salientar que os pixels na cor branca, relativos as lacunas deixadas pelo texto extraído durante o processo de segmentação, não são levados em consideração.

De posse dos dados estatísticos coletados anteriormente, são criados os histogramas RGGr da imagem sintética. Os histogramas das imagens analisadas se mostraram similares a funções gaussianas; essa aproximação pode ser avaliada pela entropia da imagem, uma vez que quanto menor a entropia mais a função se aproxima de uma distribuição normal [46].

Embora seja gerado o histograma de cinza ao invés do histograma de azul, é possível calcular o valor de azul que deverá formar o pixel na imagem final através da equação da luminância (EQ. 2).

Para fins de análise, foram calculadas as entropias dos tons RGGr e da matiz (h_hue). A Tabela 2, tirada de [47], apresenta os valores de h_hue calculados para algumas imagens do acervo, na primeira coluna encontra-se o nome da imagem, e na segunda o valor da entropia encontrado.

Foram utilizadas a Eq. 6, e Eq. 7 a Eq. 8. Ressaltando mais uma vez que no cálculo da entropia, o produto das dimensões da imagem foi usado como base logarítmica.

Tabela 2. Entropia das matizes em amostras de texturas.

Imagem	Entropia (h_hue)
testebak2	0,169175
Teste4bak2	0,160451
Teste5bak2	0,180805
Teste6bak2	0,164334
Teste7bak2	0,113249

Com o objetivo de ajustar o histograma de forma a assemelhar-se a uma função gaussiana (assim como ocorre nas imagens das texturas originais) o espaço de cor foi tomado no intervalo entre a Eq. 17 e a Eq. 18, e o fator (F) é uma variável calculada iterativamente. Enquanto a soma das amplitudes do histograma for maior ou igual ao número de pixels da imagem ou se aproximar de até 20% deste valor, a variável fator é incrementada, e para evitar que o valor de fator cresça indefinidamente, o mesmo foi limitado como no máximo igual a três vezes o valor da variância.

$$\mu - F * \sigma^2 \quad (\text{Eq. 17})$$

$$\mu + F * \sigma^2 \quad (\text{Eq. 18})$$

A variância foi utilizada em detrimento do desvio padrão, na reconstrução do histograma da imagem da textura, por ter gerado gaussianas mais próximas as originais. Quando utilizado o desvio padrão as funções ficaram ‘achatadas’, e o uso da variância provocou um “esticamento” do ponto central das mesmas.

A Figura 33 mostra uma comparação entre os histogramas reconstruídos utilizando desvio padrão e a variância. O gráfico se encontra normalizado, no eixo x tem-se as cores da imagem, e no eixo y o número de vezes que as cores aparecem na imagem.

Além de determinar o espaço de cor da nova imagem, é preciso limitá-lo, visto a evidência de que combinações RGB, no espaço de cor definido anteriormente, poderão produzir cores não presentes na imagem original.

Observamos então que, os documentos apresentam um valor de matiz que se sobressai em relação aos demais pixels (max_hue). O max_hue representa, em média, cerca de 40% das cores presentes na imagem, ou seja, uma determina cor ocupa em média 40% da imagem da textura.

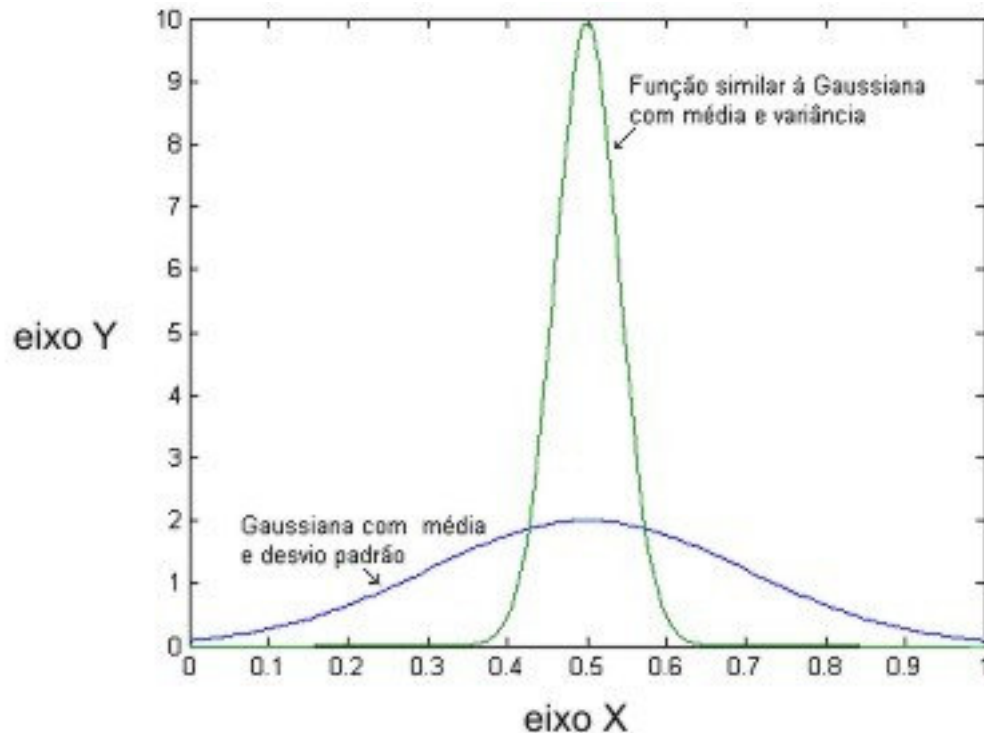


Figura 33. Gaussianas geradas com uso do Desvio Padrão e da Variância.

Na Tabela 3, pode-se observar os valores e porcentagens de aparecimento do *max_hue* encontrados em algumas imagens do acervo.

Tabela 3. Matizes mais freqüentes e porcentagem de aparição de algumas amostras de texturas.

Imagem	max_hue	Freqüência
testebak2	30	37,75%
Teste4bak2	33	26,51%
Teste5bak2	48	19,76%
Teste6bak2	31	41,62%
Teste7bak2	35	42,34%

Calculamos, então, o valor da matiz mais freqüente em cada documento e utilizamos esse valor como fator limitante para as possíveis tuplas (R,G,B) que poderão ser usadas na textura do papel. Para aumentar, ainda que de maneira diminuta, o número de matizes possíveis, utilizou-se as informações de *h_hue* das imagens para determinar um grau de tolerância (*delta*). Determinou-se empiricamente o valor de delta da seguinte forma:

- Se $h_hue > 0,17$, então $delta = 10$
- Senão, $delta = 1$.

Com isso, a matiz dos pixels que estarão presentes na nova textura deverá estar entre $max_hue - delta$ e $max_hue + delta$, e a este intervalo denominamos espaço da matiz.

O parâmetro *h_hue* é também usado para definir quantas cores do espaço RGB terão a matiz definida no espaço de matiz. Uma análise das Tabelas 2 e 3 indicam uma relação entre a entropia e o número de vezes que a entropia mais freqüente aparece na imagem. A relação entre a

entropia h_hue e o número de cores no espaço da matiz (n_cores) foi calculada empiricamente definida de acordo com a Eq. 19 [4]:

$$n_cores = 152,64 * h_hue + 16,0089 \quad (\text{EQ. 19})$$

Uma nova imagem é então criada, onde o valor de azul é recalculado e cada pixel tem sua cor determinada pela sua componente RGB, definida por buscas pseudo-aleatórias em uma tabela de cores, e pelo valor da matiz mais freqüente no documento. Se a tupla RGB tomada estiver dentro do espaço da matiz, o pixel é gravado e o histograma dos tons tomados são decrementados, do contrário, outra combinação é tentada. Caso o número de cores no espaço da matiz seja alcançado, o resto da imagem é preenchido sem restrições em relação ao valor da matiz.

4.2.1 Análise dos resultados

Apesar de, visualmente, as imagens sintéticas serem diferentes das texturas originais, suas características são bastante similares. A Figura 34 a seguir apresenta uma amostra de textura e sua versão sintética, assim como seus histogramas, e na Figura 35 é mostrada uma ampliação da imagem sintética.

As texturas sintéticas foram analisadas qualitativamente por inspeção visual, e quantitativamente, medindo a Razão Sinal-Ruído de Pico (PSNR) e Análise de Variância (ANOVA). Em ambos os casos um conjunto de 200 texturas foi gerado e comparado alcançando resultados satisfatórios.

Aplicou-se a ANOVA à medidas estatísticas da imagem, como: média, desvio padrão, assimetria e curtose (achatamento) dos histogramas das imagens. Os resultados são apresentados na Tabela 4. Na coluna de conclusões, S significa diferenças estatisticamente significativas e NS significa diferenças estatísticas não-significativas.

Tendo como hipótese nula que cada medida estatística da imagem original é igual a esta medida da imagem sintética, é calculada a probabilidade da hipótese nula ocorrer. Caso essa probabilidade seja menor que 0,01, a hipótese nula é rejeitada e a diferença entre as imagens é significativa, caso contrário as diferenças são não-significantes.

A ocorrência dos vários resultados significantes no tom azul é justificada, uma vez que, seu histograma não foi gerado automaticamente, sendo extraído dos demais histogramas de acordo com a (Eq. 2). Como o contraste, ou variância de uma imagem, representa a variação local de níveis de cinza, é imediata a verificação de que as diferenças significativas entre o contraste das imagens originais e o contraste das imagens sintéticas se dão pelo fato das imagens sintéticas terem sido obtidas tomando como fator limitante o valor da matiz mais presente na imagem original - o que acarreta uma diminuição na variação das cores na imagem gerada.

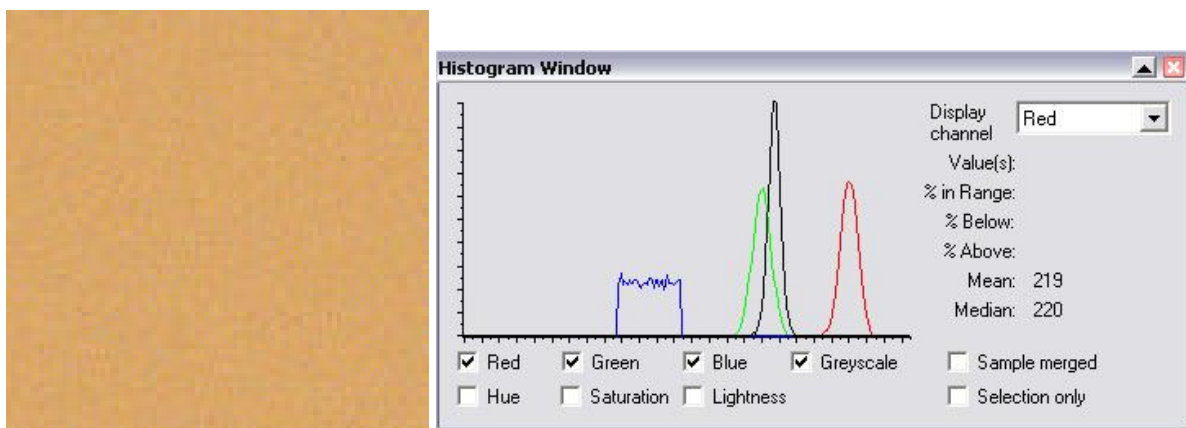
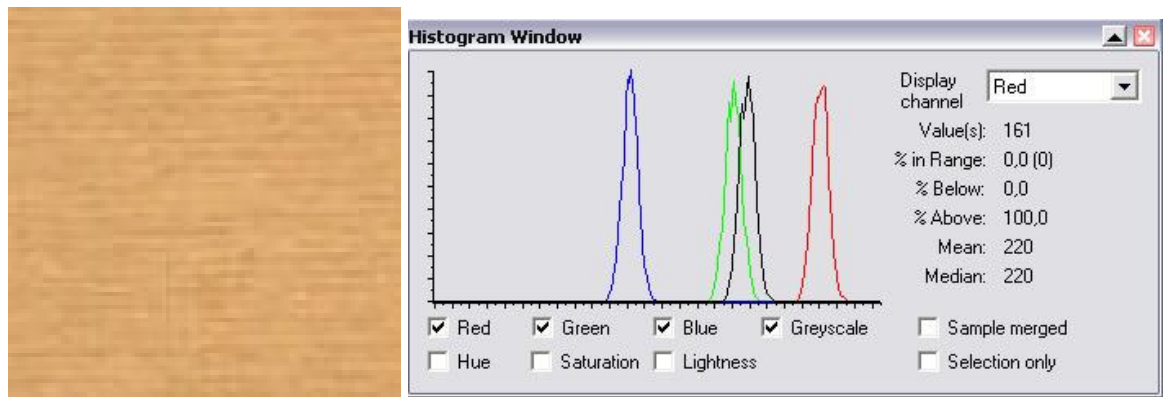


Figura 34. (topo) Textura original e seu histograma, (embaixo) textura sintética e seu histograma.



Figura 35. Textura sintética em zoom.

Tabela 4. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 200 texturas para o processo de síntese.

Medida	Vermelho			Verde			Azul		
	F(x)	p	Conclusão	F(x)	p	Conclusão	F(x)	p	Conclusão
Entropia	7,61	0,007	NS	1,54	0,22	NS	20,51	2,30E-05	S
Segundo Momento Angular	7,86	0,006	NS	0,90	0,35	NS	1,28	0,26	NS
Momento Inverso Diferencial	8,00	0,006	NS	0,99	0,32	NS	1,95	0,167	NS
Diferença das Entropias	0,02	0,885	NS	18,86	4,53E-05	S	93,63	1,10E-14	S
Soma das Entropias	18,30	5,72E-05	NS	14,24	3,28E-04	S	13,04	5,60E-04	S
Média X	0,003	0,953	S	0,02	0,88	NS	4,60	0,03	NS
Média Y	0,003	0,959	NS	0,03	0,87	NS	4,62	0,03	NS
Variância X	8,95	0,004	NS	2,17	0,14	NS	30,12	5,75E-07	S
Variância Y	8,87	0,004	NS	2,13	0,15	NS	30,25	5,48E-07	S
Energia	5,21	2,50E-02	NS	0,45	0,51	NS	0,028	0,866	NS
Contraste	21,81	1,35E-05	S	14,62	2,76E-04	S	5,93	0,02	NS
Assimetria	2,54	0,12	NS	10,75	0,002	NS	72,91	1,51E-12	S
Curtose	3,94	0,05	NS	7,44	0,008	NS	30,62	4,79E-07	S

Capítulo 5

Ambiente para geração de imagens de documentos históricos

Interfaces gráficas têm como principal objetivo tornar as aplicações mais fáceis e práticas de usar. Durante os últimos anos crescem as preocupações em desenvolver interfaces cada vez mais amigáveis, e grandes empresas de *software* vêm investindo muitos recursos em melhorias e inovações nessa área. Atualmente, já temos até mesmo um conjunto de dispositivos de iteração com os usuários que permitem, inclusive, o desenvolvimento de interfaces em 3 dimensões (3D).

Com o mesmo objetivo, de tornar a aplicação mais fácil e intuitiva, foi desenvolvida uma interface gráfica que, apesar de simples, cumpre com este objetivo. A partir dessa interface é possível executar cada fase envolvida na geração dos documentos históricos.

5.1 Descrição do ambiente

Como mencionado anteriormente, a geração de imagens sintéticas de documentos históricos consiste, *a priori*, em sua aquisição através de algum dispositivo digitalizador e segmentação do documento original, gerando assim, duas imagens, uma relativa à imagem do texto e a outra relativa à imagem do papel (textura) [3].

Após essa fase inicial de segmentação, é possível tratar, separadamente, as duas imagens. A imagem do texto pode, então, ser transposta para texto utilizando uma ferramenta OCR. E, a partir da imagem da textura, é possível extrair informações necessárias para geração da textura sintética do documento que será gerado. Ao final de todo o processo, a imagem do texto e a imagem da textura são compostas, gerando o documento final, uma versão sintética similar ao documento original. Todo esse processo pode ser visualizado na Figura 1.

O ambiente desenvolvido consiste em 5 programas, escritos na linguagem de programação C, e uma interface gráfica escrita em Java, que chama os programas escritos em C através do método *native* de Java. A relação entre os algoritmos de processamento de imagem, bem como os arquivos de entrada e saída de cada um são mostrados na Figura 36.

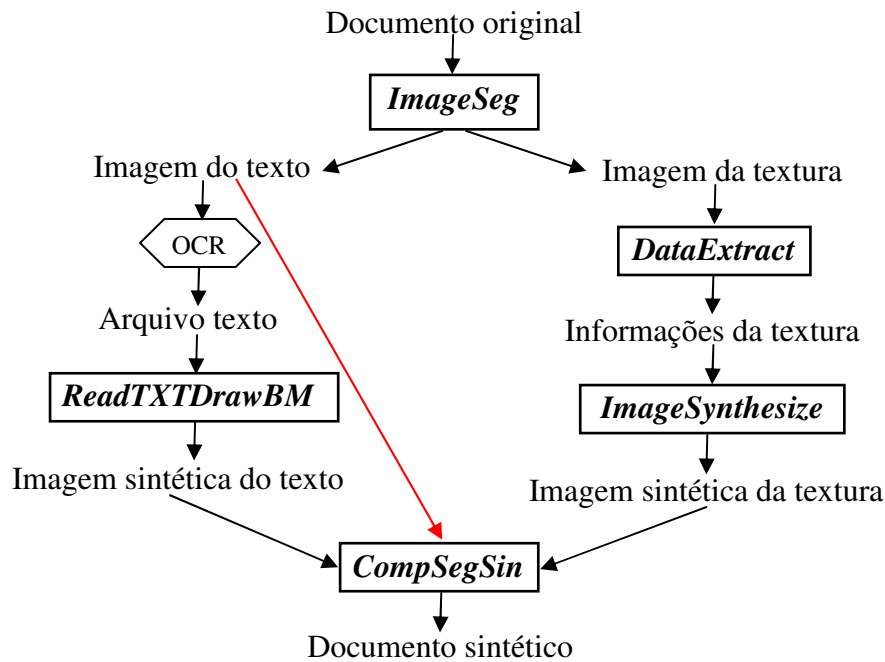


Figura 36. Fluxo de caixa dos arquivos de geração de documentos históricos.

Os 5 programas em C, destacados na Figura 36, são responsáveis por toda manipulação e processamento das imagens:

- *ImageSeg*: Responsável pela segmentação da imagem do documento original em duas outras, imagem do texto e imagem da textura, implementado de acordo a seção 3.3.
- *DataExtractor*: Responsável por ler um fragmento da textura do documento original e extrair dele um conjunto de informações estatísticas que serão posteriormente utilizadas para geração da textura sintética da textura. Essas informações são armazenadas em um arquivo binário.
- *ImageSynthesize*: Recebe o arquivo *.dat* gerado por *DataExtractor* e, a partir dele, gera a textura sintética do documento, de acordo com a seção 4.2.
- *ReadTXTDrawBMP*: Responsável por ler um arquivo de texto (txt) como entrada e gerar a imagem de um documento correspondente a esse arquivo.
- *CompSegSin*: Responsável por realizar a composição entre as imagens texto, seja ela gerada sinteticamente a partir de um texto ou resultante da etapa de segmentação, e da textura do papel.

Para invocar em Java os programas escritos em C é preciso seguir os 5 procedimentos listados abaixo, e em seguida pode-se executar a aplicação:

- 1) Criar uma classe em Java declarando o método nativo que será chamado posteriormente. Podemos ver, como exemplo, essa declaração sendo feita no trecho de código a seguir, onde na classe *Synthesize* é declarado e carregado o método nativo *ImageSynthesize*, que recebe uma *String* como argumento.

```

public class Synthesize {
    public native String ImageSynthesize(String stg);
    static {

```

```

        System.loadLibrary("ImageSynthesize");
    }
    Synthesize(){
        ...
    }
}

```

- 2) Compilar o código Java escrito anteriormente.
- 3) Criar o cabeçalho da classe Java que será usado como assinatura do programa em C. Com a classe já compilada, executamos o comando *javah* e o nome da classe, de acordo com o nosso exemplo, *javah CompSegSin*. Um arquivo *.h* é gerado contendo o cabeçalho, a seguir vemos o trecho mostrando o cabeçalho gerado.

```

JNIEXPORT jstring JNICALL Java_Synthesize_ImageSynthesize
(JNIEnv *, jobject, jstring);

```

- 4) Escrever o código em C utilizando o cabeçalho gerado e incluindo *jni.h* e o arquivo *.h* correspondente ao cabeçalho gerado. Abaixo temos a assinatura do código em C com, e sem o cabeçalho.

Sem o cabeçalho:

```

void main(int argc, char **argv){
    ...
}

```

Com o cabeçalho e os *includes*:

```

#include <jni.h>
#include "DocFrame.h"

JNIEXPORT jstring JNICALL Java_DocFrame_ImageSegmentation
(JNIEnv *env, jobject, jstring target)
{
    ...
}

```

- 5) Gerar a *dll* (*dynamic linkage library*) correspondente ao código em C. Para isso é preciso estar usando um compilador C que gere *dlls* e ter escolhido DLL como tipo de aplicação, nesse trabalho utilizamos o *Microsoft Visual Studio 2005*.

A interface gráfica foi criada utilizando os pacotes *java.awt* (*Abstract Windowing Toolkit*) e *javax.swing*. As classes do pacote *java.awt* e seus subpacotes, tais como *java.awt.event*, *java.awt.image* e *java.awt.Graphics*, agrupam as funcionalidades gráficas que estão presentes desde as primeiras versões de Java, que operam tendo por base as funcionalidades do gerenciador de janelas e bibliotecas nativas do sistema onde a aplicação é executada. Já o framework *Swing*, associado ao pacote *javax.swing* e seus subpacotes, é uma extensão definida a partir de AWT que define componentes gráficos que utilizam exclusivamente Java (*lightweight components*), com funcionalidades e aparência independentes do sistema onde a aplicação é executada.

A classe *DocFrame* que herda de *JFrame*, cria um *JFrame*, que é um panel onde será apresentado o menu de opções e o papel de fundo da aplicação. O menu de opções é criado

adicionando uma barra de menu (*JMenuBar*) ao frame principal, e adicionando itens a essa barra de menu (*JMenuItem*).

Na Figura 37, a seguir, podemos visualizar a interface gráfica criada.

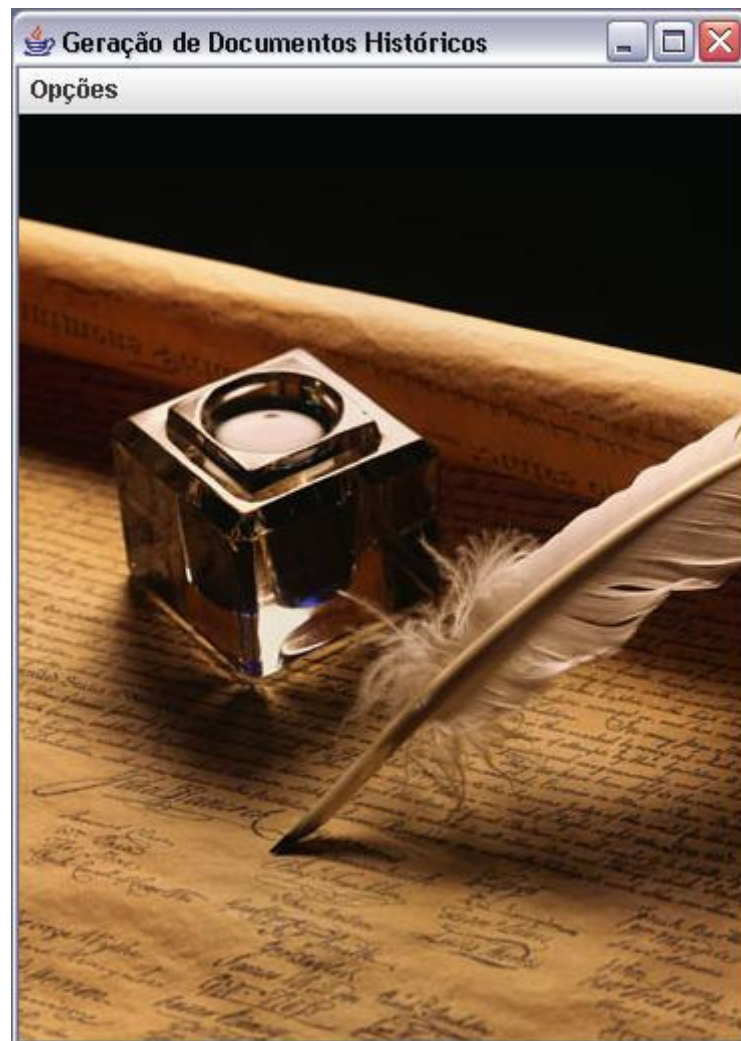


Figura 37. Interface gráfica.

Para cada funcionalidade da aplicação foi criado um item de menu, quando um desses itens é selecionado a classe correspondente à opção escolhida é chamada. Os itens de menu disponíveis são os seguintes:

- Segmentar: Chama a classe *Segmentation*, que por sua vez declara e carrega a dll *ImageSeg*, responsável pela segmentação dos documentos.
- Extrair informações da textura: Chama a classe *DataExt*, que declara e carrega a dll *DataExtractor*, responsável pela extração das informações estatísticas de um fragmento da textura do documento original.
- Sintetizar: Chama a classe *Syntheticize*, que declara e carrega a dll *ImageSyntheticize*, responsável pela síntese da textura do documento original a

partir de um arquivo de dados contendo as informações estatísticas necessárias para geração.

- Gerar imagem a partir do txt: Chama a classe *ReadTXTDrawBMP*, que declara e carrega a dll *ReadTxtDrawBMRGB*, responsável por gerar a imagem de um documento a partir de um arquivo texto com seu conteúdo.
- Geração completa sem txt: Chama a classe *CompSegSin*, que declara e carrega a dll *CompSegSis*, responsável por compor a imagem do texto gerada na fase de segmentação com a textura sintética.
- Geração completa com txt: Chama a classe *CompTxtSin*, que faz uma chamada a classe *ReadTxtDrawBMRGB* (para gerar a imagem correspondente ao arquivo txt escolhido) e a *CompSegSin* (para compor a imagem resultante com a textura sintética).
- Sair: Fecha a aplicação.

A seguir, na Figura 38, podemos ver a interface gráfica com seu menu de opções.

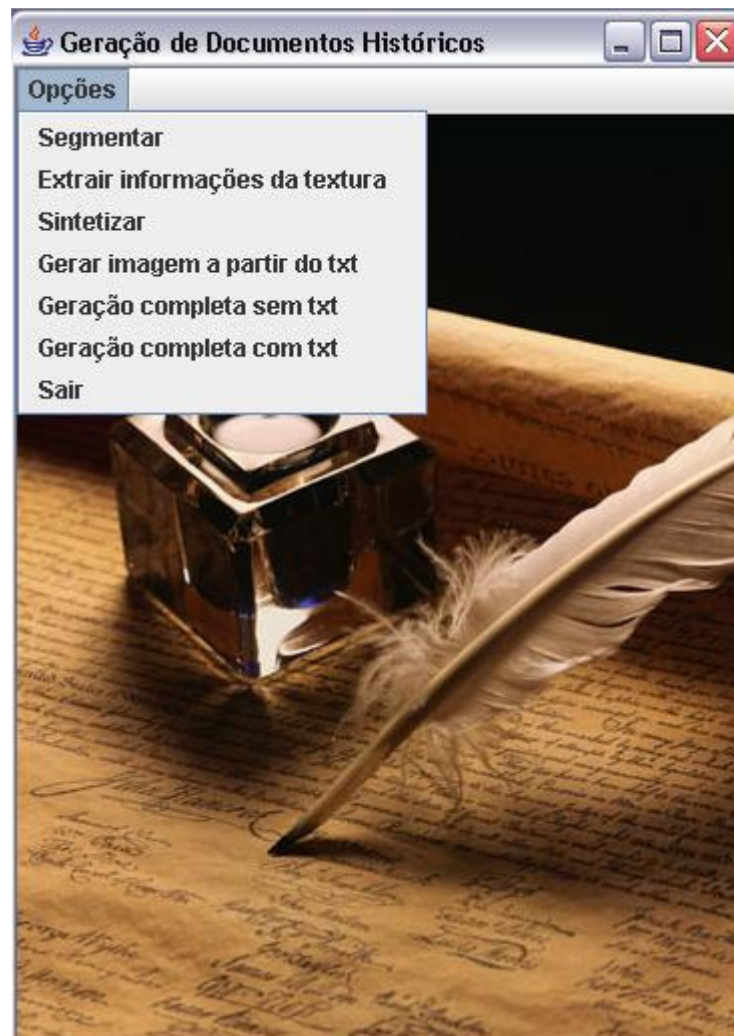


Figura 38. Interface gráfica e seu menu de opções.

Como para cada opção do menu é preciso escolher um arquivo de entrada, seja ele uma imagem, um arquivo de texto ou um arquivo .dat, para cada item de menu é apresentada uma

janela para escolha do arquivo de entrada (*JFileChooser*). Para geração completa sem txt, são necessários dois arquivos, o primeiro correspondendo a imagem do documento e o segundo a textura sintética. Da mesma forma, para geração completa com txt, é preciso escolher três arquivos, o arquivo texto cuja imagem será gerada, a imagem resultante dessa geração e por fim a textura sintética.

Na Figura 39 podemos visualizar a apresentação de uma dessas janelas de escolha de arquivo, a imagem em questão foi produzida com a escolha da opção “Gerar imagem a partir do txt”, onde se espera a escolha de um arquivo texto.

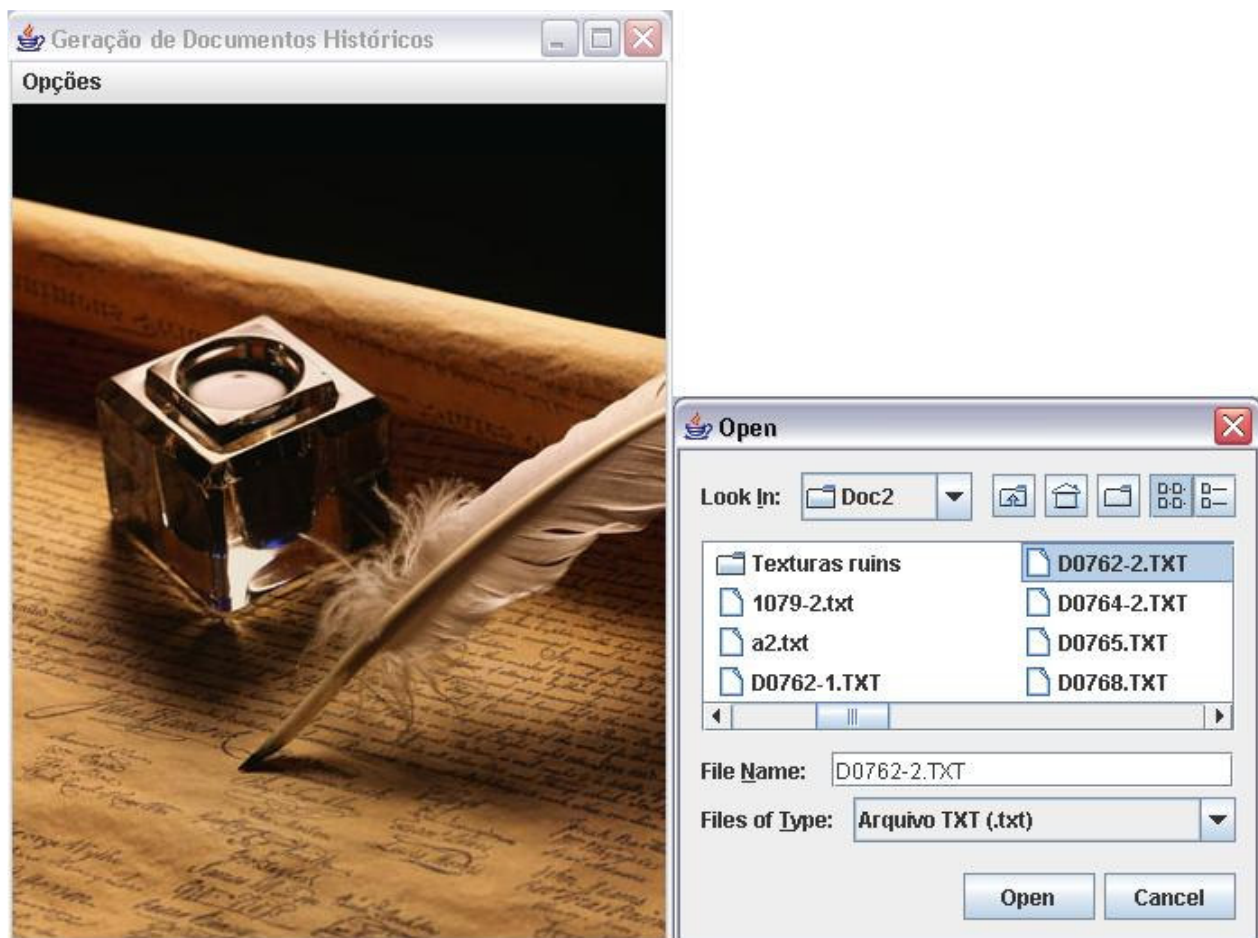


Figura 39. Interface com menu de seleção de arquivos.

Como o ambiente gráfico é feito unicamente para facilitar a interação entre o usuário final e a aplicação, e consiste em várias chamadas a métodos nativos escritos em C, é extremamente fácil adicionar novos algoritmos de segmentação, síntese e processamento de imagens. Para isso, basta que esses algoritmos estejam implementados em C/C++ ou em Java. Caso escritos em C/C++ é preciso seguir os 5 passos para que o esse possa ser invocado como método nativo pela interface Java.

5.2 Resultados gerados

Como mencionado anteriormente, a geração completa das imagens pode se dar de duas fases, a primeira a partir da imagem do texto, obtido no processo de segmentação (Figura 40), e a segunda a partir do arquivo de texto correspondente ao seu conteúdo (Figura 41).

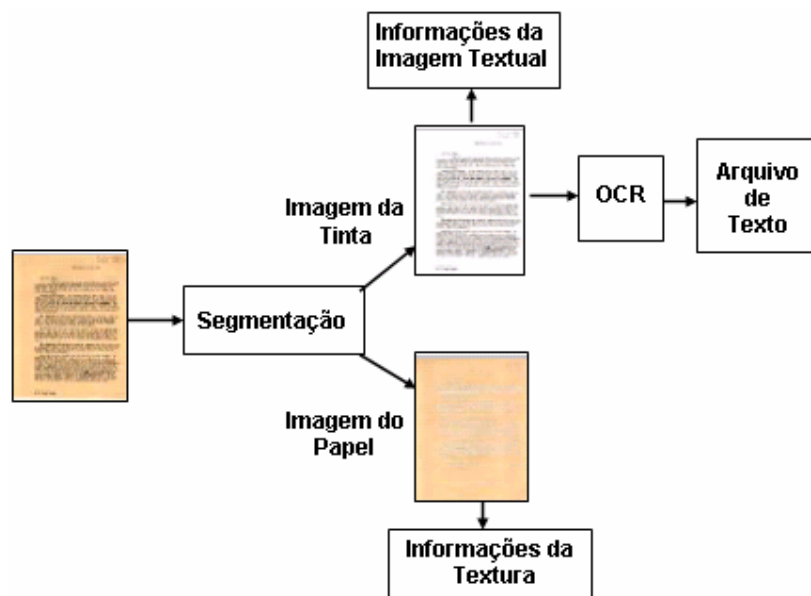


Figura 40. Processo de geração da imagem de um documento histórico. Primeira Fase: Criação dos arquivos de dados e do arquivo de texto com o conteúdo da carta.

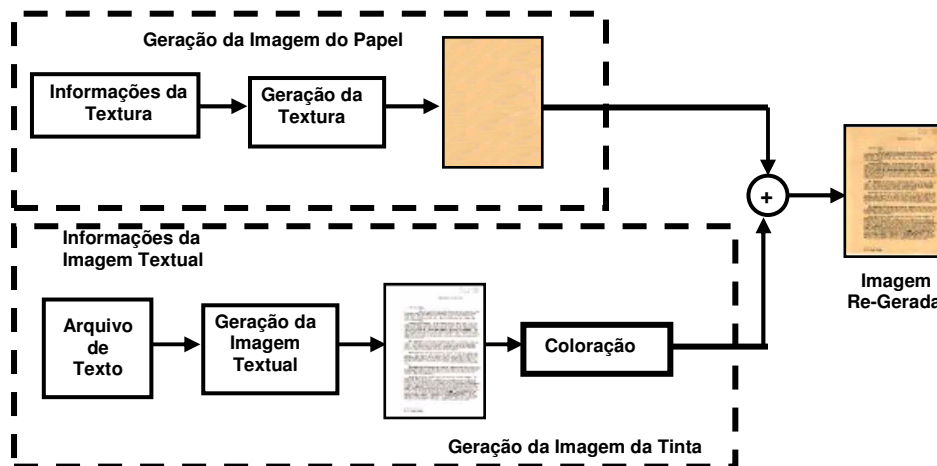
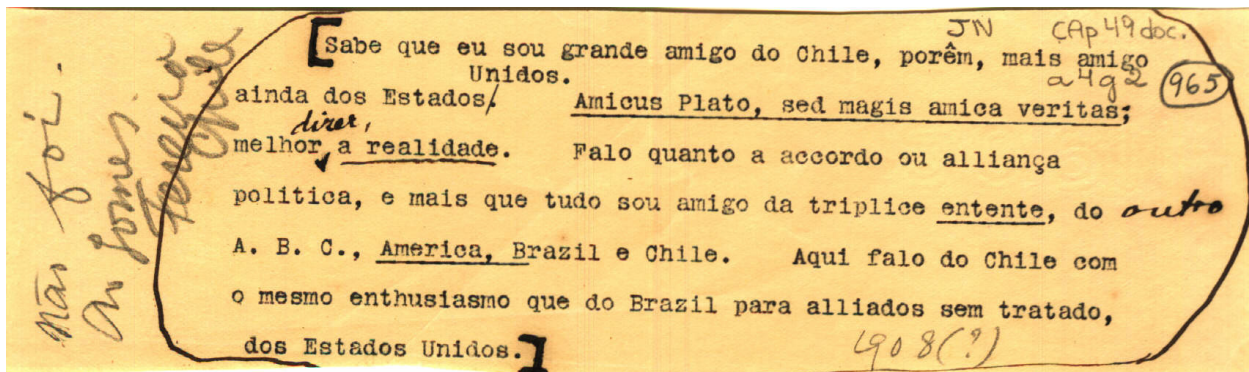
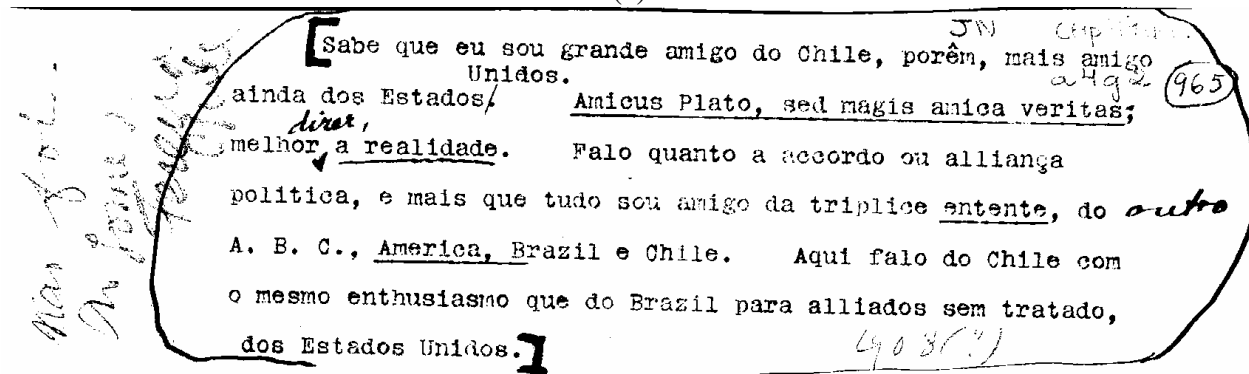


Figura 41. Processo de geração da imagem de um documento histórico. Segunda Fase: Geração da imagem a partir dos arquivos de dados e de texto.

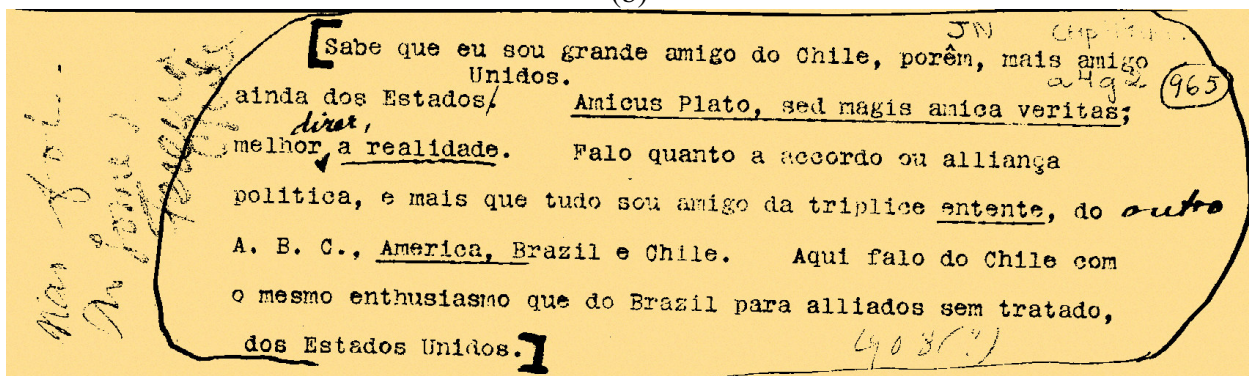
Na Figura 42, podemos ver o documento original (a) e o resultado de sua geração completa de acordo com a o esquema da primeira fase (c). Como podemos observar os documentos são visualmente diferentes, apesar de bem parecidos. A diferença maior se encontra em relação à textura. Além disso, durante a segmentação optou-se por passar os pixel relativos à tinta do papel para a cor preta, dessa forma, a imagem da tinta é mais escura que na imagem original.



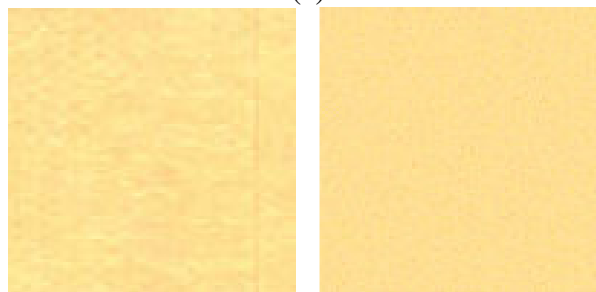
(a)



(b)



(c)



(d)

(e)

Figura 42. (a) Documento original, (b) imagem do texto resultante do processo de segmentação, (c) documento gerado de acordo com o esquema de geração de imagens da primeira fase, (d) textura original, (e) textura sintética.

Para geração dos documentos da segunda fase foi utilizado um conjunto de imagens de letras, números e caracteres especiais, extraídos das imagens segmentadas. Esses são armazenados em 256 tons de cinza e possuem dimensão de 22 x 31 pixels. Na Figura 43, são mostrados os caracteres usados como base para criação das imagens a partir de um documento de texto.

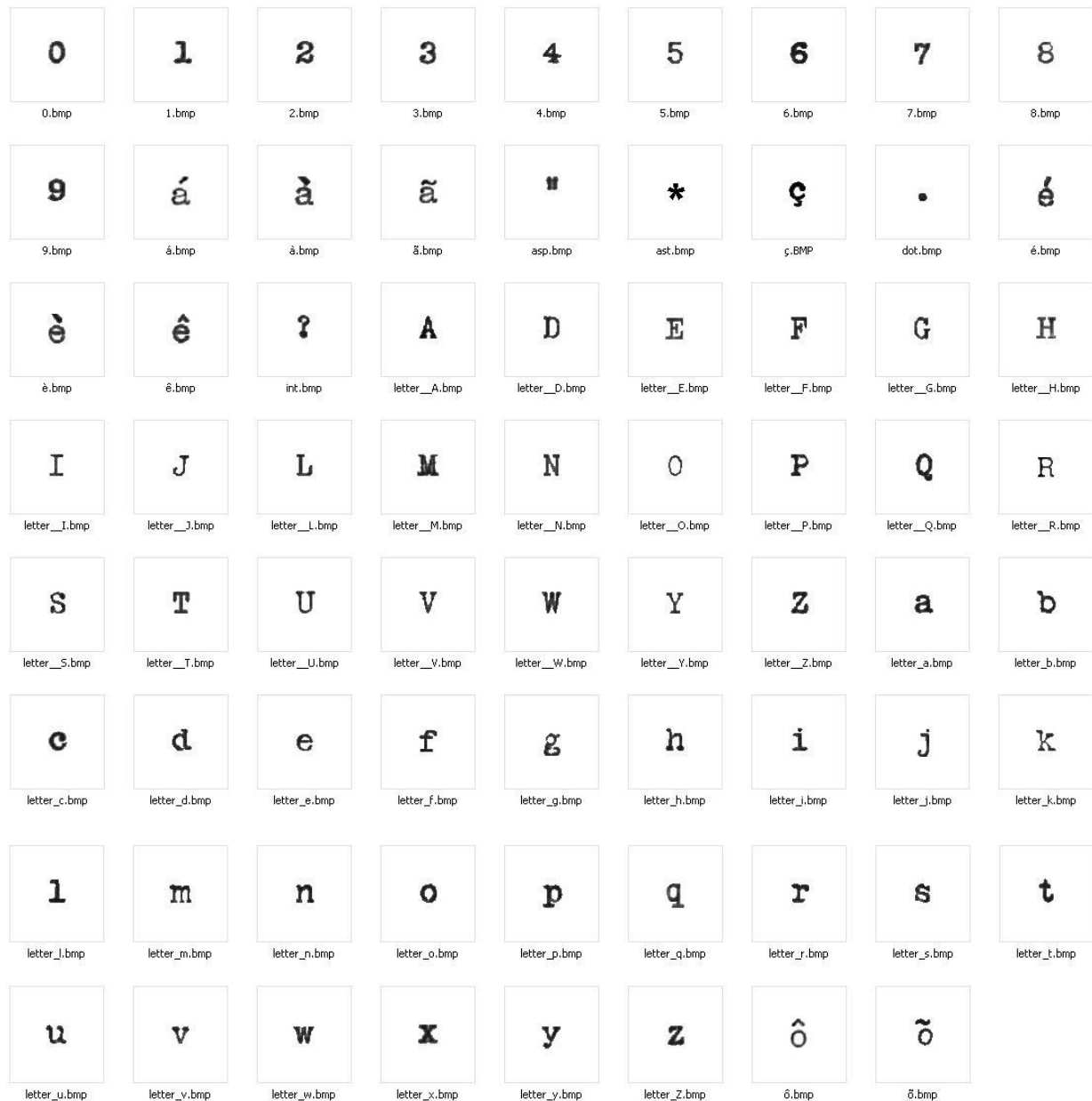
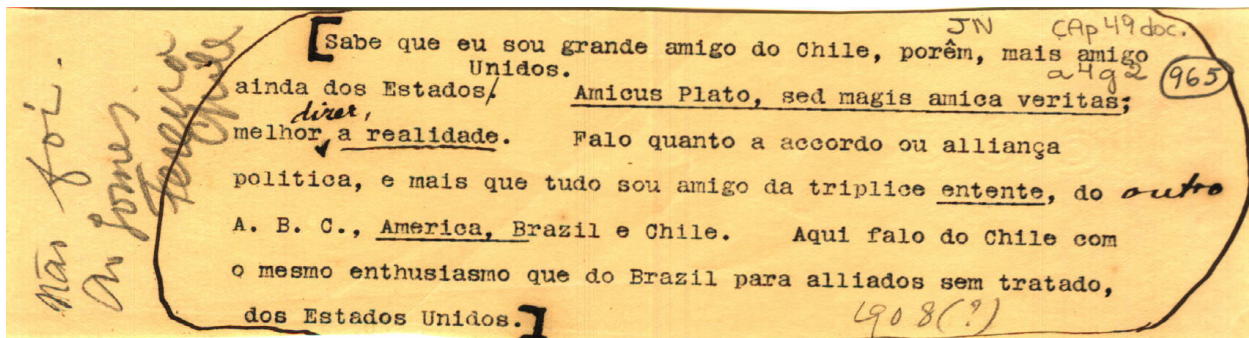
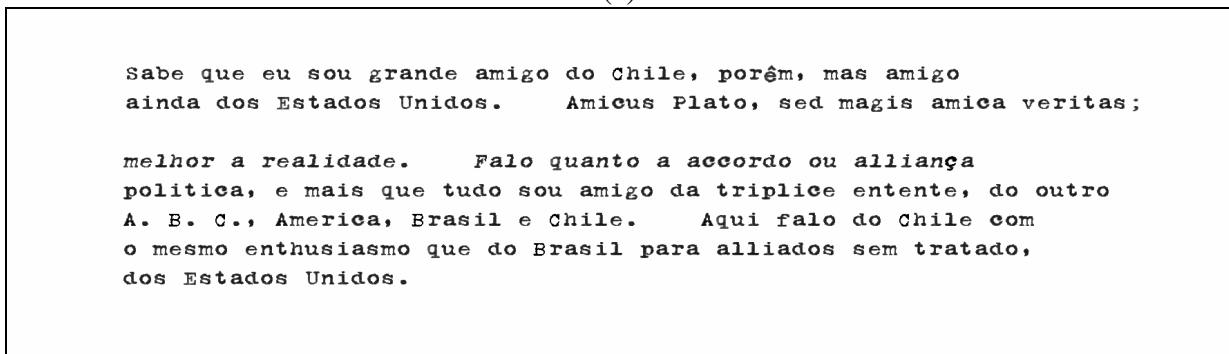


Figura 43. Base de caracteres usada na geração automática da imagem textual.

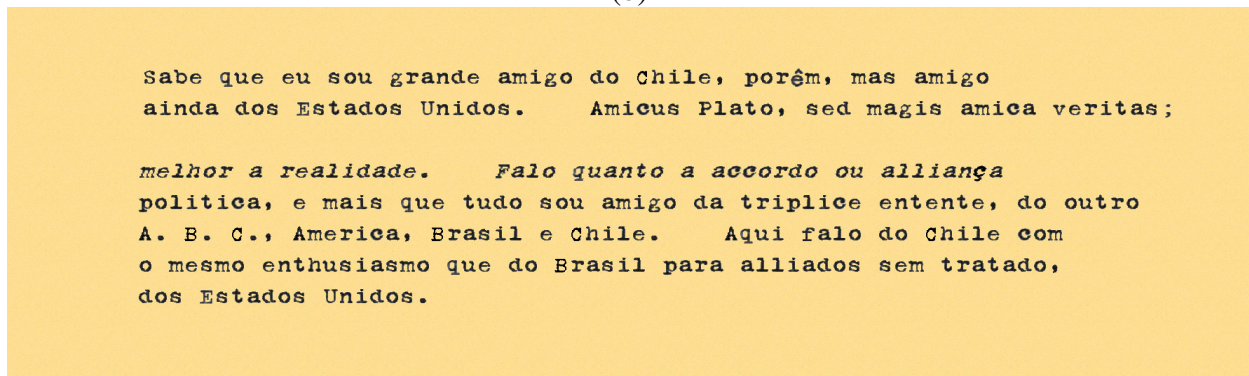
Na Figura 44, podemos ver o documento original e o resultado de sua geração completa de acordo com a o esquema da segunda fase. Os resultados gerados por este segundo esquema, como era esperado, apresentou resultados visualmente menos semelhantes à imagem original. Diferente do esquema da primeira fase, aqui as informações manuscritas nos documentos são perdidas.



(a)



(b)



(c)

Figura 44. (a) Documento original, (b) imagem gerada a partir do documento de texto, (c) documento gerado de acordo com o esquema de geração de imagens da segunda fase.

Para melhorar a estética dos documentos houve uma preocupação em determinar sua formatação (em pixels), tendo sido definida da seguinte forma:

- Margem esquerda = 150 pixels;
- Margem superior = 100 pixels;
- Espaço em branco = 22 pixels;
- Espaçamento entre linhas = 20 pixels;
- Tabulação = 4 espaços em branco.

Foi estabelecido que os caracteres não reconhecidos seriam representados pelo caractere coringa ‘*’ (asterisco).

Apesar de melhorar a estética dos documentos gerados pelo esquema da segunda fase, a pré-definição de margens e espaçamentos fixos contribuem ainda mais com as diferenças entre os

documentos originais e sintéticos, visto que essas medidas podem diferir muito de um documento para outro.

5.2.1 Análise dos resultados

As imagens sintéticas, resultantes tanto do esquema de geração da primeira quanto da segunda fase, foram analisadas qualitativamente por inspeção visual, e quantitativamente, medindo a Razão Sinal-Ruído de Pico (PSNR) e Análise de Variância (ANOVA). Em ambos os casos um conjunto de 50 texturas foi gerado e comparado alcançando resultados satisfatórios.

A ANOVA às medidas estatísticas da imagem [35], e os resultados para os documentos sintéticos gerados de acordo com o esquema de geração de imagens da primeira e segunda fase são apresentados na Tabela 5 e Tabela 6, respectivamente.

Pode-se observar que para a maioria das medidas estatísticas não houve diferenças significantes, o que mostra que para o conjunto de imagens originais e sintéticas são estatisticamente semelhantes.

A ocorrência de um maior número resultados significantes para o tom azul é justificada uma vez que a maior parte da imagem é formada por pixel relativos a textura do documento e, como visto na seção 4.2, seu histograma não foi gerado automaticamente para criação da textura sintética, sendo extraído dos demais histogramas de acordo com a (EQ. 2).

Outras diferenças estatisticamente significantes ocorrem por perdas de algumas informações no processo de segmentação e por ter se gerado a textura do documento sintético final replicando-se a textura sintética gerada com apenas um fragmento da textura original.

Tabela 5. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 50 imagens resultantes do esquema da primeira fase.

Medida	Vermelho			Verde			Azul		
	F(x)	p	Conclusão	F(x)	p	Conclusão	F(x)	p	Conclusão
Entropia	56,57	7,30E-12	S	92,75	3,74E-13	S	77,61	6,84E-12	S
Segundo Momento Angular	13,50	0,0006	S	17,33	0,0001	S	25,46	5,86E-06	S
Momento Inverso Diferencial	16,50	0,0002	S	16,58	0,0002	S	19,64	4,84E-05	S
Correlação	0,16	0,6936	NS	0,39	0,5346	NS	1,58	2,15E-01	NS
Diferença das Entropias	0,00	1	NS	0,00	1	NS	0	1	NS
Soma das Entropias	0,00	1	NS	0,00	1	NS	0	1	NS
Média X	0,020	0,8783	NS	0,03	0,8635	NS	0,06	0,8065	NS
Média Y	0,050	0,8181	NS	0,06	0,8003	NS	0,02	0,7212	NS
Variância X	0,04	0,8494	NS	0,06	0,8111	NS	0,08	0,7839	NS
Variância Y	0,05	0,8217	NS	0,08	0,7806	NS	0,1	7,50E-01	NS
Energia	0,06	8,10E-01	NS	0,12	0,7342	NS	0,41	0,5229	NS
Contraste	12,76	8,00E-04	S	5,39	0,0242	NS	1	0,3219	NS
Assimetria	2,87	0,0963	NS	13,10	0,0007	S	51,81	2,39E-09	S
Curtose	0,24	0,6248	NS	1,14	0,29	NS	17,76	5,00E-04	S

Para as imagens geradas de acordo com o esquema da segunda fase, o número de diferenças estatísticas significante são bem maiores. Nessa fase, como mencionado anteriormente, são consideradas apenas as informações datilogradas, sendo perdida qualquer outra informação escrita. Além disso, a estrutura do documento sintético e do original é diferente, para os documentos sintéticos foi adotada uma formatação, referente a margens e espaçamentos, que não necessariamente, será igual àquela dos documentos originais.

Diante disto, apesar da análise dos resultados com ANOVA terem apresentados uma grande quantidade de resultados estatisticamente significantes, é possível verificar, por inspeção visual, que estas diferenças decorrem principalmente da mudança estrutural do texto original para sua versão sintética.

Tabela 6. Resultado da ANOVA para diversas propriedades aplicada nos histogramas de vermelho, azul e verde de um conjunto de 50 imagens resultantes do esquema da segunda fase.

Medida	Vermelho			Verde			Azul		
	F(x)	p	Conclusão	F(x)	p	Conclusão	F(x)	p	Conclusão
Entropia	8,90	5,60E-03	S	29,11	7,63E-06	S	15,53	0,0004	S
Segundo Momento Angular	3,78	0,0614	NS	17,43	0,0002	S	15,94	0,0004	S
Momento Inverso Diferencial	0,22	0,6429	NS	9,07	0,0052	S	9,44	0,0045	S
Correlação	1,10	0,3033	NS	1,13	0,2931	NS	1,21	0,2809	NS
Diferença das Entropias	0,00	1	NS	0,00	1	NS	0	1	NS
Soma das Entropias	0,00	1	NS	0,00	1	NS	0	1	NS
Média X	27,70	1,10E-05	S	26,63	1,48E-05	S	25,86	1,84E-05	S
Média Y	9,53	0,0043	S	9,09	0,0052	S	8,99	0,0054	S
Variância X	26,98	1,34E-05	S	19,57	0,0001	S	17,26	0,0002	S
Variância Y	19,53	0,0001	S	14,37	0,0007	S	12,82	2,60E-03	S
Energia	14,47	7,00E-04	S	10,76	0,0026	S	10,82	0,0026	S
Contraste	4,90	3,46E-02	NS	11,66	0,0026	S	0	1	NS
Assimetria	0,03	0,8543	NS	3,74	0,0627	NS	17,47	0,0002	S
Curtose	1,31	0,2615	NS	0,00	0,9654	NS	18,03	0,0002	S

Capítulo 6

Conclusões e Trabalhos Futuros

Ao longo deste trabalho foram estudados e desenvolvidos diversos algoritmos de processamento de imagens que envolveram a segmentação, síntese e geração automática de imagens de documentos históricos. Foi criada uma interface gráfica para integrar todos os algoritmos criados, tornando o uso das funcionalidades desenvolvidas mais simples, bastando alguns cliques no *mouse* para geração completa da versão sintética do documento desejado.

Apesar do aumento do interesse mundial em preservar seus documentos históricos, pouca pesquisa é realizada em relação ao processamento das imagens desses documentos. A maioria dos trabalhos relacionados à disponibilização de versões digitais de documentos históricos se restringem a transcrição desses documentos para mídia digital, utilizando *scanners* ou câmeras digitais. Dessa forma, destacamos aqui a importância e potencial desse trabalho, que apresenta possibilidades de aplicação em diversas áreas de processamento de imagens.

Os algoritmos desenvolvidos a despeito de aplicados a um determinado acervo [2] podem ser aplicados a outros acervos que possuam características semelhantes. Além disso, a interface gráfica criada permite adicionar facilmente novos algoritmos e funcionalidades aos sistemas, aumentando ainda mais a gama de acervos com os quais poderemos trabalhar.

Uma vez que objetivamos a preservação e divulgação de quaisquer documentos históricos, a ferramenta aqui desenvolvida poderá ser disponibilizada e utilizada por instituições que tenham interesse em divulgar seus acervos.

Portanto, nosso trabalho possui uma forte contribuição social, permitindo que acervos, algumas vezes até mesmo desconhecidos ou inacessíveis à sociedade, sejam facilmente disponibilizados, sem que isso cause nenhum tipo de dano aos mesmos.

Além da contribuição social, nosso trabalho poderá proporcionar, através do incentivo a utilização da mídia digital, a diminuição dos danos ecológicos e ambientais causados pelo uso exacerbado e crescente desmatamento para fabricação de papel.

Dando prosseguimento aos trabalhos aqui desenvolvidos, é preciso desenvolver e investigar a existência de algoritmos eficazes de geração automática de texturas coloridas, que se aproximem o máximo possível das texturas originais. É preciso também, procurar meios de otimização de código, objetivando a diminuição do tempo computacional necessário para geração dos documentos. Pode-se ainda, melhorar a geração automática do texto, acrescentando informações nos arquivos a respeito das margens e espaçamentos utilizados em cada documento em particular.

Por fim, é preciso aprimorar a interface gráfica objetivando torná-la mais amigável, e permitir, por exemplo, a abertura e visualização dos documentos gerados na própria interface.

Bibliografia

- [1] PARKER, J. and WILEY, J. *Algorithms for Image Processing and Computer Vision*, John Wiley & Sons, Inc., 1995.
- [2] Disponível na Internet via WWW. URL: recpad.dsc.upe.br/site_hist. Visitado em 12/05/2006.
- [3] MELLO, C e LINS, R. *Generating Paper Texture of Images of Historical Documents*. IEEE Int. Conf. on Accoustic, Speech and Signal Processing, Turquia, 2000.
- [4] MELLO, C. *Synthesis of Images of Historical Documents for Web Visualization*. IEEE International Multi-Media Modeling Conference, Brisbane, Austrália, 2004.
- [5] CAVALCANTI, C., MELLO, C. e SOUZA, M. *Generating True Color Paper Textures of Historical Documents*. IEEE Int. Conf. on Accoustic, Speech and Signal Processing, Filadélfia, 2005.
- [6] Projeto Gutenberg. Disponível na Internet via WWW. URL: <http://www.humi.keio.ac.jp>. Visitado em 12/05/2006.
- [7] Livro de Emblemas – Universidade de Glasgow. Disponível na Internet via WWW. URL: <http://www.emblems.arts.gla.ac.uk>. Visitado em 12/05/2006.
- [8] Projeto ARTFL – American and French Research on the Treasure of the French Language – Universidade de Chicago. EUA. Disponível na Internet via WWW. URL: <http://humanities.uchicago.edu/orgs/ARTFL/>. Visitado em 12/05/2006.
- [9] Koninklijke Bibliotheek. Disponível na Internet via WWW. URL: <http://www.kb.nl/index-en.html>. Visitado em 12/05/2006.
- [10] Digital Scriptorium. Disponível na Internet via WWW. URL: <http://sunsite3.berkeley.edu/Scriptorium/>. Visitado em 12/05/2006.
- [11] Cervantes Digital – The Cervanets Project. Disponível na Internet via WWW. URL: <http://www.cSDL.tamu.edu/cervantes/english/index.html>. Visitado em 12/05/2006.
- [12] DEBORA Digital Library. Disponível na Internet via WWW. URL: <http://debora.enssib.fr>. Visitado em 12/05/2006.
- [13] GONZALEZ, R. and WINTZ, P. *Digital Image Processing*, Addison-Wesley Publishing Company, Inc., second edition, 1987.
- [14] GOMES, J. e VELHO, L. *Computação Gráfica: Imagem*, IMPA, 1995.
- [15] Disponível na Internet via WWW. URL: <http://atlas.ucpel.tche.br/~vbastos/realce.htm>. Visitado em 18/04/2006.
- [16] GONZALEZ, R. and WOODS, R. *Digital Image Processing*, Addison-Wesley Publishing Company, Inc., second edition, 2001.
- [17] LUCCHESI, L. and MITRA, S. *Color image segmentation: A state-of-the-art survey*. Proc. of the Indian National Science Academy (INSA-A), 2001.
- [18] FAN, J. et al. *Automatic image segmentation by integrating color-edge extraction and seeded region growing*, IEEE Trans. on Image Processing, 2001.

- [19] SEZGIN, M., SANKUR, B. *Survey over image thresholding techniques and quantitative performance evaluation*. Journal of Electronic Imaging, 2004.
- [20] WU, V. and MANMATHA, R. *Document Image Clean-up and Binarization*. In the Proceedings of the SPIE conf. on Document Recognition V, San José, Califórnia, 1998.
- [21] SANKUR, B., and SEZGIN, M. *Image Thresholding Techniques: A Survey over Categories*. Pattern Recognition, Bogazii University, Electric-Electronic Engineering Department, Bebek, Istanbul, Turquia, 2001.
- [22] Adobe Systems Inc. Disponível na Internet via WWW. URL: <http://www.adobe.com>. Visitado em 15/04/2006.
- [23] SANTOS, J., BORTOLOZZI, F. e SABOURIN, R. *Background Bank Cheques Elimination: A Simple and Efficient Solution*. Centro Federal de Educação Tecnológica do Paraná – CEFET - PR, 1997.
- [24] MELLO, C and LINS, R. *Image Segmentation of Historical Documents*, Visual 2000, Agosto, 2000, México.
- [25] MELLO, C and LINS, R. *A New Segmentation Algorithm for True Colour Images of Historical Documents*, XVIII Simpósio Brasileiro de Telecomunicações, Setembro, 2000, Brasil.
- [26] WANG, Z., BOVIK, A. and LU, L. *Why is Image Quality Assessment so Difícult?*. IEEE Int. Conf. on Accoustic, Speech and Signal Processing, Florida, EUA, 2002.
- [27] MCGRATH, R. *Understanding Statistics – A Research Perspective*. Addison-Wesley Educational Publishers, 1997.
- [28] Disponível na Internet via WWW. URL: http://davidmlane.com/hyperstat/intro_ANOVA.html. Visitado em 18/04/2006.
- [29] SHARMA, M. and SINGH, S. *Evaluation of texture methods for image analysis*. In 7th Australian and New Zealand Intelligent Information Systems Conference, Perth, Western Austrália, 2001.
- [30] REED, T and BUF, J. *A review of recent texture segmentation and feature extraction techniques*, *Computer Vision, Image Processing and Graphics*, vol. 57, no. 3, pp. 359-372, 1993.
- [31] VANGOOL, L., DEWAELE, P, and OOSTERLINCK, A. *Texture analysis*, *Computer Vision, Graphics and Image Processing*, vol. 29, pp. 336-357, 1985.
- [32] WESZKA, J., DYER, C. and ROSENFELD, A. *A comparative study of texture measures for terrain classification*, IEEE Transactions on Systems, man and Cybernetics, vol. 6, pp. 269-285, 1976.
- [33] OHANIAN, P and DUBES, R. *Performance evaluation for four class of texture features*, Pattern Recognition, vol. 25, no. 8, pp. 819-833, 1992.
- [34] CONNERS, R. and HARLOW, C. *A theoretical comparison of texture algorithms*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 3, pp. 204-222, 1980.
- [35] HARALICK, R, SHANMUGAM, K. and DINSTEN, I. *Textural features for image classification*. IEEE Transactions on Systems, Man, and Cybernetics , vol. SMC 3, número 6, Novembro de 1973, p 610.
- [36] WANG, Z. and BOVIK, A. *A universal image quality index*. IEEE Signal Processing Letters, v. 9, n. 3, p. 81-84, Março de 2002.
- [37] PAPPAS, T and SAFRANEK, R. *Perceptual criteria for image quality evaluation, in Handbook of Image and Video Processing* (A. Bovik, ed.), Academic Press, Maio de 2000.

- [38] VQEG, *Final report from the video quality experts group on the validation of objective models of video quality assessment*, Março de 2000. Disponível na Internet via WWW. URL: <http://www.vqeg.org>.
- [39] LAMBRECHT, C. *Special issue on image and video quality metrics*, Signal Processing, vol. 70, Novembro de 1998.
- [40] GIROD, B. *What's wrong with mean-squared error*, in *Digital Images and Human Vision* (A. B. Watson, ed.), pp. 207–220, the MIT press, 1993.
- [41] LUBIN, J. *A visual discrimination mode for image system design and evaluation*, in *Visual Models for Target Detection and Recognition* (E. Peli, ed.), pp. 207–220, Singapore: World Scientific Publishers, 1995.
- [42] DALY, S. *The visible difference predictor: An algorithm for the assessment of image fidelity*, in Proc. SPIE, vol. 1616, pp. 2–15, 1992.
- [43] A. B. Watson, J. Hu, and J. F. III. McGowan, *Digital video quality metric based on human vision*, Journal of Electronic Imaging, vol. 10, no. 1, pp. 20–29, 2001.
- [44] WATSON, A., YANG, G., SOLOMON, J. and VILLASENOR, J. Visibility of wavelet quantization noise, IEEE Trans. Image Processing, vol. 6, pp. 1164–1175, Agosto de 1997.
- [45] MARTENS, J. And MEESTERS, L. *Image dissimilarity*, Sinal Processong, vol. 70, p. 155 – 176, Novembro de 1998.
- [46] KAPUR, J. *Measures of Information and their Applications*, John Wiley and Sons, 1994.
- [47] MELLO, C. Filtragem, *Compressão e Síntese de Imagens de Documentos Históricos*. Tese de doutorado, Universidade Federal de Pernambuco, 2002.