



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Tecnologia

Kauê Tartarotti Nepomuceno Duarte

**Contribuições em Análise de Textura para fins
de Segmentação de Imagens Digitais**

**Contributions to Texture Analysis for Digital
Image Segmentation**

Limeira

2017



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Tecnologia

Kauê Tartarotti Nepomuceno Duarte

Contribuições em Análise de Textura para fins de Segmentação de Imagens Digitais

Dissertação apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Tecnologia, na Área de Sistemas de Informação e Comunicação.

Orientador: Prof. Dr. Marco Antonio Garcia de Carvalho

Coorientador: Prof. Dr. Paulo Sergio Martins Pedro

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Kauê Tartarotti Nepomuceno Duarte, e orientada pelo Prof. Dr. Marco Antonio Garcia de Carvalho

Limeira

2017

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.

ORCID: <http://orcid.org/0000-0002-4074-3672>

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Felipe de Souza Bueno - CRB 8/8577

D85c Duarte, Kauê Tartarotti Nepomuceno, 1993-
Contribuições em análise de textura para fins de segmentação de imagens digitais / Kauê Tartarotti Nepomuceno Duarte. – Limeira, SP : [s.n.], 2017.

Orientador: Marco Antonio Garcia de Carvalho.
Coorientador: Paulo Sergio Martins Pedro.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Características de textura. 2. Transformada watershed. 3. Transformada wavelet. 4. Matriz de coocorrência. 5. Corte normalizado. 6. Teoria dos grafos. 7. Segmentação de imagens. I. Carvalho, Marco Antonio Garcia de, 1970-. II. Pedro, Paulo Sérgio Martins, 1967-. III. Universidade Estadual de Campinas. Faculdade de Tecnologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Contributions to texture analysis for digital image segmentation

Palavras-chave em inglês:

Texture features

Watershed transform

Wavelet transform

Gray-level co-occurrence matrix

Normalized cut

Graph theory

Image segmentation

Área de concentração: Sistemas de Informação e Comunicação

Titulação: Mestre em Tecnologia

Banca examinadora:

Marco Antonio Garcia de Carvalho [Orientador]

Hélio Pedrini

Celmar Guimarães da Silva

Data de defesa: 11-07-2017

Programa de Pós-Graduação: Tecnologia

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de dissertação para o Título de Mestre em Tecnologia na área de concentração de Sistemas de Informação e Comunicação, a que submeteu o aluno Kauê Tartarotti Nepomuceno Duarte, em 11 de julho de 2017 na Faculdade de Tecnologia- FT/ UNICAMP, em Limeira/SP.

Prof. Dr. Marco Antonio Garcia de Carvalho

Presidente da Comissão Julgadora

Prof. Dr. Celmar Guimarães da Silva

Faculdade de Tecnologia – Universidade Estadual de Campinas

Prof. Dr. Hélio Pedrini

Instituto de Computação – Universidade Estadual de Campinas

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do aluno na Universidade.

Dedico essa dissertação em memória de meus avós maternos, em especial a minha vó Margarida, que veio a falecer na metade de meu projeto pela doença de Alzheimer. Gostaria também de dedicar ao meu irmão Kaike, a minha mãe Inês, ao meu pai Landau e aos meus avós paternos Jorge e Sidnea.

Agradecimentos

Agradeço primeiramente a Deus por ter me dado a força interior e me escutado quando diversas dúvidas apareciam em minha cabeça.

Ao meu irmão Kaike, por estar sempre comigo, a minha mãe Inês, por me confortar com suas palavras e me oferecer todo suporte para que eu pudesse continuar, ao meu pai Landau, por sempre me incentivar independente dos obstáculos e mostrar que é acreditando que se consegue. Aos meus avós Jorge e Sidnea, por acreditarem em mim, desde quando eu era apenas um jovem estudante em Embu Guaçu-SP.

Aos meus colegas do grupo ImageLab, por incentivarem-me durante todo o decorrer do trabalho, e ao ex-aluno do grupo Tiago Pinto por oferecer sua ajuda em diversos momentos.

Aos meus professores e orientadores Dr. Marco Antonio Garcia de Carvalho e Dr. Paulo Sergio Martins Pedro, por me auxiliarem durante toda a minha caminhada com conversas técnicas e palavras de incentivo.

Resumo

Segmentação é o processo de dividir a imagem em regiões ou grupos, que permitam a identificação de objetos ou características. Para isso, a área de Segmentação de Imagens possui uma gama de métodos que segmentam os mais diversos tipos de imagens. Dentre estes, existem os que empregam a análise de padrões de textura; análise esta que viabiliza a interpretação de diversas informações. As informações de textura são extraídas, normalmente, pixel a pixel. Extraí-las de regiões pré-definidas ainda é um desafio e necessita, em alguns casos, adaptar o descritor para este fim. Este trabalho apresenta duas contribuições referentes à análise de textura no processo de segmentação de imagens. A primeira consiste na segmentação de estômatos através da combinação da Transformada Wavelet *à trous* e a Transformada Watershed. A segunda consiste na aplicação da Matriz de Coocorrência, juntamente com a Transformada Watershed e o Corte Normalizado, para a segmentação de imagens naturais. As bases de dados utilizadas foram, respectivamente, uma base obtida juntamente com o grupo ScianLab da Universidade do Chile e a BSDS500, da Universidade da Califórnia-Berkeley. Os resultados em ambas as aplicações foram avaliados pela medida-F, amplamente utilizada na literatura, e comparados a situações que incluem diferentes técnicas para abordar a segmentação e o uso de textura. As performances das técnicas propostas foram bastante promissoras, como na primeira aplicação, com a obtenção de 98% de acurácia na identificação e 70% de acurácia na segmentação dos estômatos e, para a segunda, na superação da acurácia de outras técnicas.

Palavras-chaves: Características de Textura; Transformada Watershed; Transformada Wavelet; Matriz de Coocorrência; Corte Normalizado; Teoria dos Grafos; Segmentação de Imagens.

Abstract

Image segmentation is the process of splitting images into regions or groups that allows the identification of features and objects. The image segmentation field has several methods that segment the most diverse types of images. Among them, there are those that employ analysis of texture patterns, which facilitates the interpretation of relevant information in an image. Texture information is commonly extracted pixel by pixel from an image. Extraction from predefined regions remains a challenge that often requires the adaptation of texture descriptors. This work offers two contributions to image segmentation using texture analysis. The first is to segment stomata images through the combination of the *à trous* Wavelet and the Watershed transforms. The second consists in the combination of the Gray-Level Cooccurrence Matrix, the Watershed Transform and Normalized Cut to segment general images. The datasets were obtained from the ScianLab group (University of Chile) and the BSDS500 (University of California-Berkeley), respectively. The findings in both applications were evaluated by the well-known F-measure. The comparisons include different techniques to approach both segmentation and texture features. The results obtained were promising in both cases. For example, the first application achieved an accuracy of 98% and 70% in the identification and segmentation of stomata structures, respectively. In the second application, the F-measure accuracy outperformed other techniques.

Keywords: Texture Features; Watershed Transform; Wavelet Transform; Gray-Level Cooccurrence Matrix; Normalized Cut; Graph Theory; Image Segmentation.

*“Aqueles que se sentem satisfeitos sentam-se e nada fazem. Os insatisfeitos são os
únicos benfeitores do mundo”
(Walter Savage Landor)*

Lista de ilustrações

Figura 1 – Tecido vegetal da planta <i>Ugni Molinae</i>	16
Figura 2 – Representação de grafos: (a) Representação gráfica; (b) Matriz de Adja- cência; (c) Matriz de Similaridade; (d) Matriz de Graus; (e) Matriz La- placiana	21
Figura 3 – Imagem Digital: (a) Imagem em tons de cinza; (b) Representação Ma- tricial, com valores dispostos em um intervalo de 0 a 255 (256 níveis)	22
Figura 4 – Etapas para o processamento de imagens. (PEDRINI; SCHWARTZ, 2008)	22
Figura 5 – Segmentação de Imagens: (a) Imagem Original; (b) Imagem segmentada em 2 regiões.	24
Figura 6 – Definição dos quatro operadores.	26
Figura 7 – Operações em conjunto: (a) Imagem original; (b) Imagem Transladada; (c) Imagem refletida.	28
Figura 8 – Aplicação dos Operadores de Dilatação e Erosão: (a) Objeto inicial; (b) Elemento estruturante; (c) Dilatação; (d) Erosão.	29
Figura 9 – Modelos de espaços de cores: (a) Espaço aditivo; (b) Espaço subtrativo.	30
Figura 10 – Espaço de cores RGB	30
Figura 11 – Espaço de cores $L^*a^*b^*$	31
Figura 12 – Processo de inundação em uma imagem cortada transversalmente (m representa os mínimos regionais): (a) Imagem sem a inundação; (b) Imagem com o processo de inundação iniciado com vizinhança 8; (c) Imagem com o processo de inundação concluído	32
Figura 13 – Processo de inundação aplicado em uma imagem 2D: (a) Imagem Ori- ginal #42049 da BSDS500; (b) Gradiente Morfológico; (c) Aplicação da Transformada Watershed	34
Figura 14 – Imagem da planta <i>Ugni Molinae</i> (DUARTE <i>et al.</i> , 2017): (a) Imagem Original; (b) Marcadores que representam a localidade dos estômatos; (c) Imagem Segmentada	34
Figura 15 – Geração da Árvore de Lagos Críticos: (a) Imagem cortada transversal- mente; (b) Imagem preenchida; (c) Árvore gerada para definir o número de regiões da imagem.	34
Figura 16 – Obtenção da IFT a partir de um grafo ponderado (pontos escuros re- presentam os <i>seed-pixels</i>): (a) Grafo Ponderado; (b) Representação via IFT	35

Figura 17 – Modelagem Imagem-Grafo: (a) Imagem com 15 regiões obtidas através da aplicação do Watershed Hierárquico; (b) Grafo de Similaridade . . .	38
Figura 18 – Tipos de Cortes: (a) Grafo cíclico com 6 nós; (b) Corte Máximo; (c) Corte Mínimo; (d) Corte Normalizado	39
Figura 19 – Matrizes de Coocorrência com $\theta = 0$ e $d = 1$: (a) Matriz Original; (b) Matriz de Coocorrência sem Normalização; (c) Matriz de Coocorrência com Normalização (valores arredondados)	43
Figura 20 – Orientações referentes aos ângulos para a geração da Matriz de Coocorrência.	44
Figura 21 – Matrizes de Coocorrência: (a) Matriz Original; (b) Matriz de Coocorrência com 0° ; (c) Matriz de Coocorrência com 45° ; (d) Matriz de Coocorrência com 90° ; (e) Matriz de Coocorrência com 135°	45
Figura 22 – (a) Aproximação da Pirâmide que representa as escalas da imagem; (b) Sistema simples para geração das pirâmides de aproximação e residual de previsão (GONZALEZ; WOODS, 2010).	47

Lista de Acrônimos e Abreviações

ASM *Angular Second Moment*

BSDS500 *Berkeley Segmentation Data Set and Benchmarks 500*

C *Contrast*

CIELab *Commission Internationale de l'Eclairage L*a*b**

E *Energy*

F *F-Measure*

FN *False Negative*

FP *False Positive*

GLCM *Gray-Level Cooccurrence Matrix*

H *Homogeneity*

IFT *Image Foresting Transform*

NCut *Normalized Cut*

PDI *Processamento Digital de Imagens*

RG *Region Gray level*

RGB *Red-green-blue*

SCIAN-Lab *Laboratory of Scientific Image Analysis*

SGT *Spectral Graph Theory*

TEG *Teoria Espectral de Grafos*

TN *True Negative*

TP *True Positive*

TX *Texture descriptors*

Sumário

1	Introdução	15
1.1	Objetivos	17
1.2	Organização do Texto	17
2	Fundamentação Teórica	19
2.1	Conceitos Gerais de Grafos	19
2.1.1	Definições Matemáticas	19
2.2	Imagem Digital	21
2.3	Processamento Digital de Imagens	22
2.3.1	Etapas para Processamento Digital de Imagens	22
2.4	Segmentação de Imagens	23
2.4.1	Medida F	24
2.5	Morfologia Matemática	26
2.5.1	Operadores Morfológicos	27
2.6	Espaço de Cores	29
2.6.1	Espaço de Cores RGB	29
2.6.2	Espaço de Cores XYZ	31
2.6.3	Espaço de Cores $L^*a^*b^*$	31
2.7	Transformada Watershed	32
2.7.1	Transformada Imagem Floresta	33
2.7.2	Transformada Watershed via IFT	35
2.8	Representação da Imagem por Grafos	37
2.9	Particionamento em Grafos	38
2.9.1	Corte Normalizado	39
2.10	Textura	41
2.10.1	Matriz de Coocorrência	42
2.10.1.1	Construção da Matriz de Coocorrência	42
2.10.1.2	Cálculo dos Descritores de Textura	44
2.10.2	Transformada Wavelet	46
2.10.2.1	Transformada Wavelet <i>à trous</i>	48
2.10.3	Função de Distância	49
3	Considerações Finais	51
	Referências	54
	Apêndices	58

A	<i>Segmenting High-Quality Digital Images of Stomata Using the Wavelet Spot Detection and the Watershed Transform</i>	59
B	<i>Adding GLCM Texture Analysis to a Combined Watershed Transform and Graph Cut Model for Image Segmentation</i>	68
	Anexos	81
A	Carta de Permissão para Inclusão do Apêndice A no Trabalho	82

1 Introdução

Imagem, advinda do termo em latim *imago*, significa a representação visual da informação. Desde os tempos da Grécia antiga este termo vem sendo debatido, inicialmente com a teoria desenvolvida por Platão, o idealismo, ao qual uma imagem era considerada uma ideia projetada pela mente e, após com Aristóteles, que a considerava como uma aquisição dos sentidos, também como uma representação mental. Desta forma, apesar do conceito ter sido bem difundido com o passar dos séculos, a importância da imagem mantém-se alta. Atualmente, as imagens são bastante utilizadas para servir de suporte ao veículo da informação, ao qual, digitalmente, podem ser acessadas em diversos lugares no mundo.

Neste sentido, a grande área de Visão Computacional provê modos para solucionar problemas mais complexos, imitando assim a cognição e habilidade humana de interpretar quase instantaneamente o propósito das imagens. Porém, mesmo com os recentes avanços da comunidade científica, realizar processos focados em imagens digitais ainda é desafiador.

De acordo com Pedrini e Schwartz (PEDRINI; SCHWARTZ, 2008), existem dois níveis de abstração que simplificam as tarefas de visão computacional: o processamento digital de imagens (baixo nível) e a análise de imagens (alto nível).

O *Processamento Digital de Imagens* (PDI) é um conjunto de técnicas que objetivam capturar, representar e transformar imagens com o auxílio de uma máquina. Ou seja, processos como melhoria da qualidade visual de determinadas imagens, que visem facilitar a interpretação humana da informação, estão classificadas neste conjunto.

Ao tratar da interpretação de alto nível, a *análise de imagens* está diretamente relacionada às informações características contidas, tais como: forma (LUCCHI *et al.*, 2012), textura (O; KIM, 2016) e coloração (ISPIRYAN *et al.*, 2013). Tarefas como segmentação de imagens em regiões ou ROIs (*Regions Of Interest*, em inglês) e simplificação de objetos para representar um determinado conteúdo são encontradas entre seus propósitos. Além das características acima citadas, contornos, bordas e tamanhos, também são consideradas características comuns utilizadas na descrição de um objeto. Diante de tais definições expostas, realizar a segmentação de imagens, ainda que utilizando algoritmos mais complexos, é um desafio.

Conforme avançam as áreas de PDI e Segmentação de Imagens, evolui também a complexidade e o número de técnicas para segmentar ou mesmo reconhecer padrões. Dentre estas técnicas, pode-se citar a Transformada Watershed (BEUCHER; MATHMA-

TIQUE, 1991), as baseadas em grafos (CARVALHO *et al.*, 2010a) e Corte Normalizado (SHI; MALIK, 2000), estas empregadas neste trabalho.

Porém, independentemente das técnicas empregadas para realizar a segmentação, alguns desafios mantêm-se para todas as imagens; um destes é a necessidade de algoritmos que definam, cada vez melhor, os contornos e as regiões. Quanto mais automatizado for o método proposto, maior será a complexidade na implementação de bons descritores para captação de determinados padrões. A robustez, parâmetro que indica a viabilidade do método em diversos contextos, também necessita ser mais estudada, pois os métodos que são empregados para realizar a segmentação ainda utilizam as características de um grupo específico de imagens, tornando-os inviáveis a outros.

Duas contribuições às áreas de segmentação de imagens e análise de textura são declaradas neste trabalho, ambas disponíveis em artigos publicados em eventos internacionais, produzidos durante a execução deste mestrado.

Como primeira contribuição, é proposto um método de segmentação de estômatos. Estômatos são pequenas estruturas localizadas abaixo das folhas nas plantas, possuem um formato celular e são responsáveis pela troca gasosa e pelo auxílio a fotossíntese. Na Figura 1 é ilustrado um tecido vegetal com essas estruturas. Este método é composto por três etapas principais: 1) Aquisição e Extração; 2) Detecção celular; 3) Segmentação. No processo de aquisição e extração, a imagem é obtida e pré-processada, com a redução do tamanho da imagem e a alteração do espaço de cores, a fim de viabilizar a identificação dos estômatos nos passos seguintes. Na detecção celular é utilizada a técnica *Wavelet Spot Detection* para localizar os estômatos. Na segmentação, é aplicada a Transformada Watershed via Marcadores (identificados no passo anterior) que visa encontrar as regiões correspondentes aos estômatos. A avaliação foi realizada através da medida F, obtendo uma acurácia aproximada de 70% ao fim do processo.

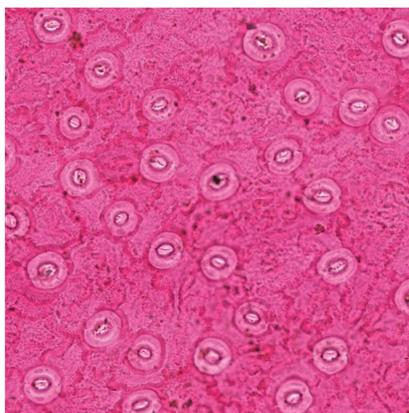


Figura 1 – Tecido vegetal da planta *Ugni Molinae*.

Como segunda contribuição, é proposto um método para segmentação de ima-

gens naturais. O método consiste basicamente de quatro etapas: 1) Segmentação Inicial; 2) Geração do Grafo de Similaridade; 3) Extração de Características; 4) Particionamento do Grafo. Na primeira etapa, a imagem é pré-processada, obtendo assim o Gradiente Morfológico da imagem. Este gradiente servirá de entrada para a aplicação da Transformada Watershed, que gerará as regiões iniciais. É aplicado o Watershed Hierárquico para reduzir o número de regiões. Um grafo de similaridade é gerado, com as regiões sendo vinculadas aos nós do grafo, e a afinidade entre elas sendo vinculada às arestas. Para cada região, são extraídas as características de textura via Matriz de Coocorrência. A afinidade entre regiões é calculada por uma função que integra três características (Contraste, Energia e Homogeneidade). Para finalizar, é aplicado o Corte Normalizado para particionar o grafo, gerando assim a imagem segmentada. As comparações realizadas neste trabalho vão desde o método sem tais descritores, a outros métodos disponíveis na literatura.

1.1 Objetivos

O objetivo deste trabalho é demonstrar que a inserção de características de textura melhora os resultados da segmentação de imagens, através de duas abordagens em duas aplicações distintas. Para tal feito, foram utilizadas técnicas bastante conhecidas na literatura, a Matriz de Coocorrência e a Transformada Wavelet. A Matriz de Coocorrência é combinada com a Transformada Watershed e o Corte Normalizado para a segmentação de imagens naturais. A Transformada Wavelet é combinada com a Transformada Watershed para a segmentação de imagens biológicas.

Inicialmente, este trabalho tinha como objetivo inserir características de textura em um framework que continha a Transformada Watershed e o Corte Normalizado para segmentação de imagens naturais. Contudo, durante a execução do mestrado, foi possível participar de um curso de treinamento em Montevideu-Uruguai chamado *Processing and Analysis of Fluorescence Microscopy Images*. Este curso possibilitou a definição de um novo objetivo, que era inserir as características de textura em imagens biológicas, mais especificamente, de estômatos.

1.2 Organização do Texto

O texto desta dissertação segue o modelo de coletânea de artigos. Durante o mestrado, dois artigos foram elaborados, contendo tanto a abordagem proposta, quanto resultados, discussões, imagens e conclusões específicas utilizadas. Portanto, na seção Apêndice são apresentados os trabalhos que foram publicados, gerados no decorrer deste trabalho de mestrado. Estas publicações são apresentadas nos Apêndices A e B, contendo

respectivamente, o artigo *Segmenting High-Quality Digital Images of Stomata Using the Wavelet Spot Detection and the Watershed Transform*, apresentado na conferência *12th International Conference on Computer Vision Theory and Applications*, e o artigo intitulado *Adding GLCM Texture Analysis to a Combined Watershed Transform and Graph Cut Model for Image Segmentation*, aceito na conferência *Advanced Concepts for Intelligent Vision Systems*.

Os outros capítulos desta dissertação são detalhados conforme segue: O Capítulo 1 trouxe uma breve introdução do tema que será trabalhado no decorrer da leitura deste trabalho; As fundamentações teóricas necessárias para o compreensão deste trabalho estão relatadas no Capítulo 2; Na sequência, as Considerações Finais trazem as últimas conclusões, bem como sugestões de trabalhos futuros.

2 Fundamentação Teórica

Neste Capítulo serão descritos os conceitos necessários para a compreensão deste trabalho. A abordagem utilizada para descrição dos tópicos que seguem é apresentada partindo de um tópico macro, iniciando pela definição de conceitos gerais de grafos, até um tópico mais pontual relativo a uma técnica utilizada, como o Corte Normalizado.

2.1 Conceitos Gerais de Grafos

Para descrever os principais conceitos de grafos, ou minimamente, os conceitos necessários para o entendimento deste trabalho, é pertinente entender um pouco sobre a visão macro de grafos, ou seja, a Teoria dos Grafos.

2.1.1 Definições Matemáticas

Grafo é uma estrutura algébrica $G = (V, E)$ que possui o conjunto de *vértices* $i \in V$, encontrado na literatura também como conjunto de *nós* ou *vértices*, e por um conjunto de *arestas* $e(i, j) \in E$, descrito também como conjunto de *arcos* ou *linhas*, tal que $i, j \in V$ (WILSON; WATKINS, 1990). Dados dois vértices $i, j \in V$, pode-se dizer que são adjacentes, denotado por $i \sim j$, se eles mantiverem ao menos uma referência no conjunto E (WILSON; WATKINS, 1990). Um *grafo ponderado* possui um peso $w(i, j)$ referente à aresta que interconecta dois nós. Define-se então como grafo ponderado $G = (V, E, W)$ todo grafo que necessita do valor(peso) para cada aresta (i, j) , representado então como $w(i, j) \in W$.

Um grafo em que todo par distinto de vértices define uma aresta é chamado de *Grafo Completo*. O Grafo Completo com n vértices é denotado por K_n . Denota-se número de vértices em um grafo por $|V|$, e o número de arestas por $|E|$ (GIBBONS, 1985; WILSON; WATKINS, 1990). Em um Grafo Completo, o número de arestas é dado por $n * (n - 1)/2$.

Dado um grafo G , é permitido definir como *subgrafo* de G todo grafo obtido pela remoção de um número (diferente de 0) de arestas ou vértices de G . A remoção de vértices necessariamente implica na remoção de qualquer aresta que seja incidente neles. Porém, se forem removidas arestas $e \in E$ ou vértices $v \in V$, então o grafo resultante pode ser definido como $(G - e)$ ou $(G - v)$. A análise inversa também é permitida, se H é um subgrafo de G , então G é um supergrafo de H , definido por $H \subset G$. Um subgrafo de G é

induzido por um subconjunto de vértices, $V' \subset V$. Portanto, o grafo G consiste de V' e as arestas de G são ambos pontos finais de V' (GIBBONS, 1985).

Em Teoria dos Grafos, um *caminho* de tamanho N é uma lista com o conjunto de vértices ordenados $\pi = \{1, 2, 3, \dots, N - 1, N\}$ e o conjunto de arestas determinado por $\{\{1, 2\}, \{2, 3\}, \dots, \{N - 1, N\}\}$, ao qual a orientação dos arcos é sempre a mesma (PEDRINI; SCHWARTZ, 2008). Ou seja, caminho é um trajeto entre vértices adjacentes.

As estruturas em grafos podem ser utilizadas também para processos mais simples, como por exemplo, a criação de um percurso. *Percurso* em grafos, também conhecido como itinerário ou cadeia, representa a ligação sequencial entre nós adjacentes. Em um grafo podem existir diversos percursos, bastando apenas gerar uma ordem sequencial de nós para que um caminho seja criado (NETTO, 2012).

Wilson e Watkins (WILSON; WATKINS, 1990) relatam que é possível representar um grafo através de algumas formas:

- *Listas*: Um diagrama consistindo de pontos ligados por linhas, estabelecendo vértices e listas de arestas que partem daquele vértice, representado de forma gráfica. Uma das vantagens das listas é a capacidade de visualizar o grafo como um todo. Este modo é bastante utilizado para armazenar grafos esparsos, os quais possuem diversos vértices, mas sem um grande conjunto de arestas.
- *Matrizes*: A representação matricial é responsável por armazenar os dados de um grafo. Para isso, são definidos diversos tipos de matrizes, dentre elas, será citado a de adjacência, de graus e a laplaciana.

A *matriz de adjacência* A tem a finalidade de mapear todas as conexões de todos os vértices de G . O tamanho desta matriz é $N \times N$, sendo N a quantidade de vértices no grafo G . A matriz de adjacência é definida conforme a Equação 2.1.

$$A(i, j) = \begin{cases} 1, & \text{se } i \sim j; \\ 0, & \text{caso contrário} \end{cases} \quad (2.1)$$

Entretanto, para representar um grafo ponderado, é necessário informar também o peso das arestas. O valor da ponderação pode ser advindo de uma função de similaridade, denominando a matriz gerada, para este caso em específico, como *Matriz de Similaridade* S . Esta matriz pode ser definida conforme a Equação 2.2.

$$S(i, j) = \begin{cases} w(i, j), & \text{se } i \sim j; \\ 0, & \text{caso contrário} \end{cases} \quad (2.2)$$

Pode-se definir o grau de um vértice como a soma dos pesos nas arestas que incidem nele. Para a Matriz de Adjacência não ponderada, cada conexão tem peso 1. Outro modo de calcular o grau de um vértice é realizar a soma, elemento a elemento, da linha, ou coluna, na Matriz de Adjacência em uma posição correspondente ao nó. Pode-se determinar o grau do vértice portanto, através da Equação 2.3.

$$d(i, i) = \sum_j w(i, j) \tag{2.3}$$

Como é desejado saber apenas a quantidade de conexões de cada vértice em si, a *Matriz de Graus* D é uma matriz diagonal com $d(i, i)$ na sua diagonal principal.

A *Matriz Laplaciana* L está relacionada à importância do espectro do grafo (explicado na Seção 2.9.1). Tal matriz pode ser definida conforme a Equação 2.4.

$$L(G) = D - S \tag{2.4}$$

A Figura 2 ilustra as Equações 2.1, 2.2, 2.3, 2.4 aplicadas a um grafo conexo.

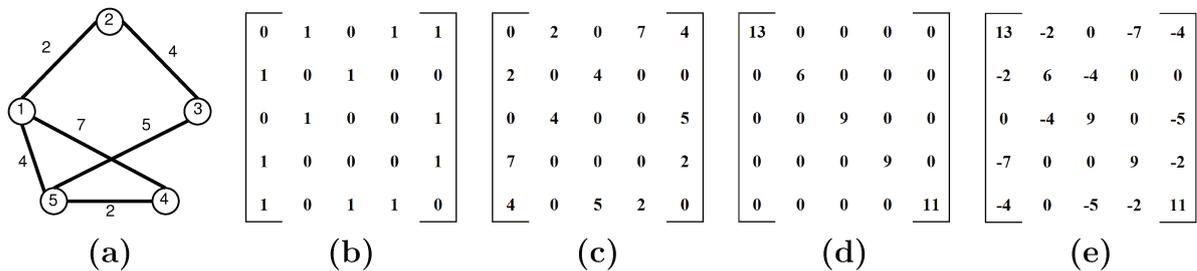


Figura 2 – Representação de grafos: (a)Representação gráfica; (b)Matriz de Adjacência; (c)Matriz de Similaridade; (d)Matriz de Graus; (e)Matriz Laplaciana

2.2 Imagem Digital

Imagem é definida como toda função bidimensional $f(x, y)$, sendo x e y coordenadas espaciais, caracterizando plano, e a amplitude $f(x, y)$ como um valor para a intensidade em tons do ponto. Porém, quando as coordenadas e amplitudes são limitadas, este plano passa a ser nomeado como *imagem digital*. Cada localização (x, y) específica é chamada de pixel. O que é determinado como intensidade do pixel é dada pela amplitude $f(x, y)$. A Figura 3 ilustra um exemplo para facilitar a compreensão sobre o que é uma imagem digital.

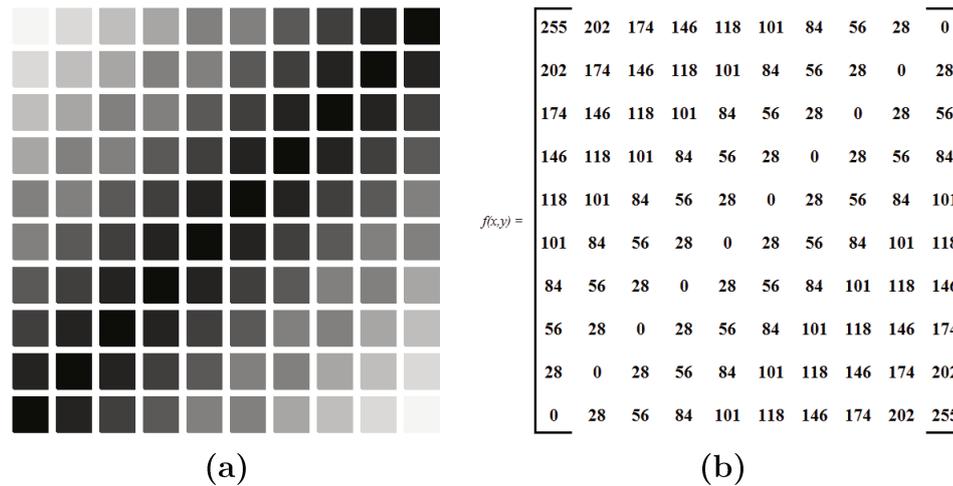


Figura 3 – Imagem Digital: (a) Imagem em tons de cinza; (b) Representação Matricial, com valores dispostos em um intervalo de 0 a 255 (256 níveis)

2.3 Processamento Digital de Imagens

Processamento Digital de Imagens-PDI é qualquer processamento que utilize imagens como parâmetro de entrada e saída (PEDRINI; SCHWARTZ, 2008). Segundo Câmara (CÂMARA *et al.*, 1996), o objetivo de PDI é oferecer a melhoria de aspectos visuais de modo que facilite tanto a análise humana, quanto funcione de entrada para outros processamentos que venham a ser realizados na sequência.

2.3.1 Etapas para Processamento Digital de Imagens

A proposta desta seção é definir o conjunto de etapas necessárias para caracterizar a área de PDI. As etapas propostas fazem direta referência ao modelo definido em Pedrini e Schwartz (PEDRINI; SCHWARTZ, 2008), no qual dividem o PDI em cinco etapas. Estas etapas podem ser identificadas, ordenadamente, na Figura 4.

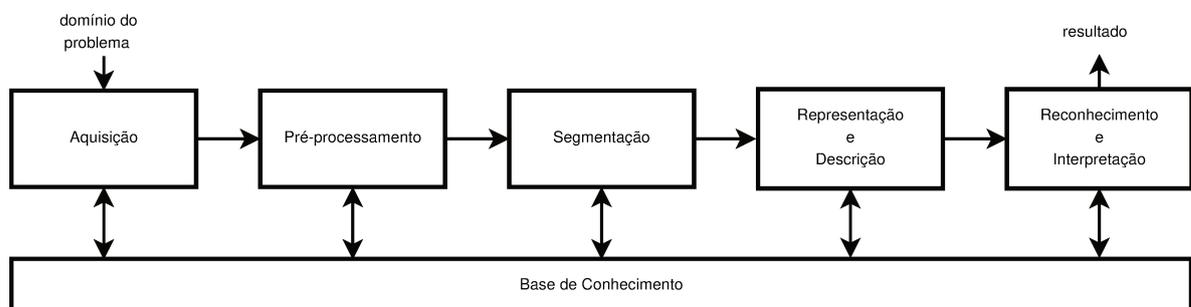


Figura 4 – Etapas para o processamento de imagens. (PEDRINI; SCHWARTZ, 2008)

A etapa de *Aquisição* corresponde à obtenção da imagem, seja por uma câmera ou qualquer outro dispositivo, que a capture e converta-a em uma representação digital. Alguns fatores nesta etapa podem ser factíveis de estudo a fim de melhorar a imagem

obtida, como por exemplo, ajustes em iluminação, resolução, número de tons de cinza, coloração, entre outros.

Após a obtenção da imagem na etapa de aquisição, é realizada a etapa de *Pré-processamento*. Esta etapa visa retirar, ou no mínimo reduzir, possíveis degradações e imperfeições que possam ter ocorrido em decorrência de alguns aspectos, entre eles a iluminação, defeito do equipamento (e.g. câmera, filmadora, *scanner*), ambiente e outros. É concebível nesta etapa a aplicação de filtros para redução de ruídos, alteração do contraste, ajuste no brilho e outras propriedades.

Na sequência, é empregada a etapa de *Segmentação*. Em suma, a Segmentação visa dividir a imagem em regiões, ou áreas de interesse, a fim de facilitar a análise, tanto do computador quanto do ser humano. Dentre as categorias de técnicas de segmentação que podem ser utilizadas nesta etapa, é possível ressaltar a detecção de bordas e a similaridade entre pixels. Devido à importância da etapa de Segmentação para o entendimento deste trabalho, a mesma será abordada com mais detalhes posteriormente.

É importante, após realizar a segmentação, armazenar as características obtidas da imagem. Para isso, devem ser definidas uma ou mais estruturas que comportem essa informação, caracterizando assim a etapa de Representação. O processo de Descrição é responsável pela extração de características que possam ser utilizadas para diversos fins.

Finalmente, a etapa de *Reconhecimento e Interpretação* é empregada. Reconhecimento é o processo que valora, ou rotula, as regiões. Essa rotulação pode ser baseada, por exemplo, nas características que foram oferecidas pelos descritores. A etapa de Interpretação é responsável por atribuir um significado ao conjunto de informações obtidas.

2.4 Segmentação de Imagens

A interpretação de dados, partindo de uma imagem original, é uma atividade muito custosa e complexa. Por conseguinte, é comum na literatura encontrar trabalhos que realizem passos anteriores à interpretação. Um destes passos é a Segmentação de Imagens. A Segmentação de Imagens subdivide a imagem em regiões, ou agrupamento de pixels.

Computacionalmente, a *segmentação* é o processo que associa um pixel a um identificador de uma região. Este passo é utilizado para facilitar a interpretação de dados, assim como para encontrar objetos de interesse.

Paulus e Hornegger (PAULUS; HORNEGGER, 2003) declaram que os algoritmos de segmentação podem ser utilizados em diversas aplicações, como imagens médicas, reconhecimento facial, entre outros, apenas ajustando uma amostra de parâmetros.

Em Gonzalez e Woods (GONZALEZ; WOODS, 2010) é citado que se deve ter cautela ao parametrizar o nível de detalhe que o algoritmo se propõe a segmentar. Estipular a precisão do algoritmo não é uma tarefa trivial, pois existe a possibilidade de perder o objeto de interesse, tornando a segmentação um passo ineficiente neste contexto.

Na literatura, vários autores (PAULUS; HORNEGGER, 2003; GONZALEZ; WOODS, 2010; PEDRINI; SCHWARTZ, 2008) consideram como abordagens comuns, encontradas em boa parte dos algoritmos de segmentação de imagens, a definição de descontinuidade e a similaridade entre pixels. Na descontinuidade, a abordagem é dividir a imagem conforme a identificação de alterações abruptas nas intensidades de tons de cinza, podendo caracterizar a existência de uma linha, um ponto ou mesmo a borda de uma região. Na similaridade, a proposta é encontrar agrupamentos pertinentes entre pixels de acordo com um certo grau de afinidade entre eles.

Apesar da diversidade de técnicas de segmentações utilizadas na literatura, é possível agrupá-las em três categorias: 1) *segmentação não supervisionada*, que compreende algoritmos em que as regiões são associadas a uma classe, definida através da similaridade entre os elementos, não necessitando, por exemplo, de dados já rotulados para etapas de treinamento (e.g. (YU *et al.*, 2012; ZHANG *et al.*, 2012)); 2) *segmentação supervisionada*, a qual ocorre quando as regiões se associam a uma ou mais classes de interesse na imagem, e desta forma pode-se utilizar dados já rotulados para treinamento (e.g. (DERIVAUX *et al.*, 2010; MARIN *et al.*, 2011)); 3) *segmentação semi-supervisionada*, na qual ocorre a utilização de dados rotulados e não rotulados (e.g. (CHEW; CAHILL, 2015; PORTELA *et al.*, 2014)).

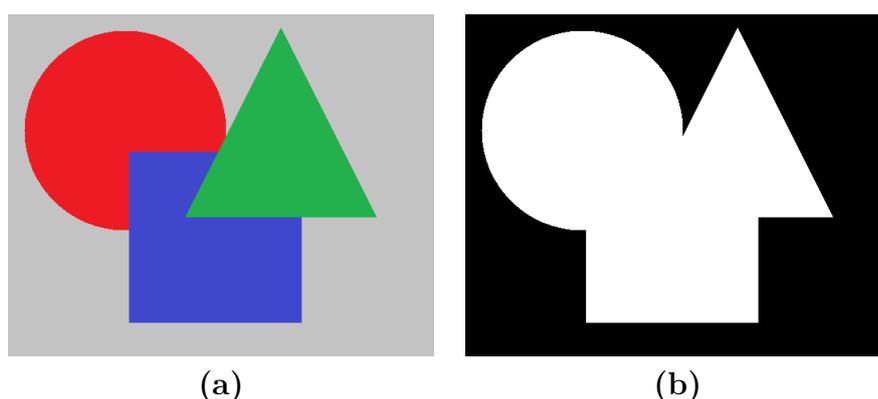


Figura 5 – Segmentação de Imagens: (a)Imagem Original; (b)Imagem segmentada em 2 regiões.

2.4.1 Medida F

Medida F (*f-measure*, em inglês) é uma métrica que avalia a acurácia da abordagem proposta. Porém, antes de formalmente defini-la em termos matemáticos, alguns

conceitos devem ser compreendidos. De certa forma, a medida F não avalia apenas segmentações, podendo ser utilizada também em tarefas como classificação e recuperação de informação. Assim, esta seção poderia ser alocada como uma seção independente, mas como neste trabalho ela ocorre sempre como meio de avaliar a segmentação, por finalidade lógica, estipulou-se-a como uma subseção.

Como toda a métrica que realiza a avaliação de acurácia, a busca é realizada para identificar se os elementos relevantes foram identificados ou não. Em tarefas como segmentação e classificação de imagens, os termos *Verdadeiros Positivos*, *Verdadeiros Negativos*, *Falsos Positivos* e *Falsos Negativos* (*true positive*, *true negative*, *false positive*, *false negative*, respectivamente em inglês) definem se os elementos relevantes foram de fato obtidos. Normalmente, a definição destes operadores ocorre com dois parâmetros de entrada, a imagem final obtida por um método a ser avaliado e o *gold standard* daquela imagem. Os termos Positivo e Negativo são referentes a predição do algoritmo sobre o elemento que foi identificado, se ele é relevante (positivo) ou não (negativo). Os termos Verdadeiro e Falso representam se as predições foram satisfeitas. De forma mais específica, os quatro operadores podem ser definidos da seguinte maneira:

- **Falso Negativo (FN):** Indica os elementos que foram identificados pela aplicação proposta como não relevantes, mas de acordo com o *gold standard* são relevantes.
- **Falso Positivo (FP):** Inversamente ao FN, os elementos são identificados como não relevantes pelo *gold standard*, mas como relevantes pela aplicação proposta.
- **Verdadeiro Negativo (VN):** Indica que um elemento foi identificado tanto pela aplicação proposta quanto pelo *gold standard* como não relevante.
- **Verdadeiro Positivo (VP):** Indica que um elemento foi identificado como relevante tanto pelo *gold standard* quanto pela aplicação proposta.

Além da definição dos termos, a Figura 6 ilustra como eles seriam representados em uma aplicação.

O conceito de elemento e relevante é adaptado de acordo com a aplicação. Em segmentação de imagens, elemento normalmente é definido como um pixel ou região, e relevante, como utilizado pelo Apêndice A e exemplo de utilização, é se aquele pixel é de um estômato ou não.

Existem duas métricas mais comumente utilizadas na literatura que avaliam a identificação dos elementos relevantes na segmentação, são elas: Precisão e Revocação.

A *Precisão* (*precision*, em inglês) é uma métrica que avalia quantos elementos que foram selecionados são de fato relevantes. O conceito de Precisão, conforme definido

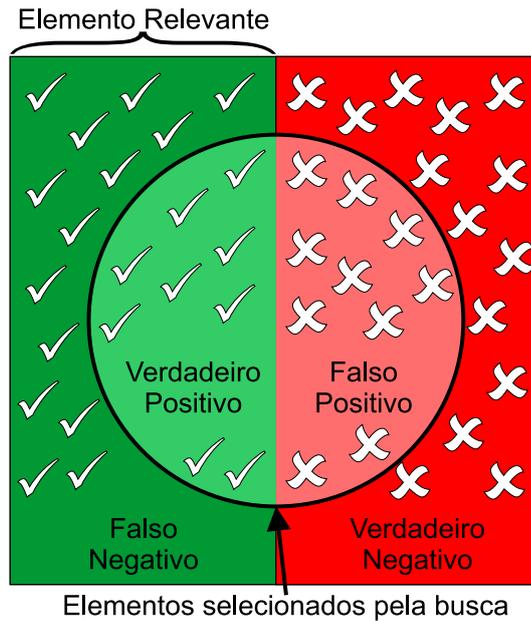


Figura 6 – Definição dos quatro operadores.

neste parágrafo, difere da definição de acurácia e precisão encontradas na literatura. Esta métrica é obtida através da Equação 2.5.

$$Precisão = \frac{VP}{VP + FP} \quad (2.5)$$

A *Revocação* (*recall*, em inglês) é a métrica que avalia quantos elementos relevantes foram selecionados. Entretanto, é comum encontrar a acurácia de 100% de Revocação, para isso bastando apenas preencher como relevante toda a imagem que está sendo processada. Conseqüentemente, não é recomendado utilizar apenas a Revocação como métrica de acurácia. Esta métrica pode ser obtida através da Equação 2.6.

$$Revocação = \frac{VP}{VP + FN} \quad (2.6)$$

A Medida F combina as métricas de precisão e revocação através de uma relação harmônica entre elas. Matematicamente, a medida F é obtida através da Equação 2.7.

$$F = 2 \cdot \frac{Precisão \cdot Revocação}{Precisão + Revocação} \quad (2.7)$$

2.5 Morfologia Matemática

Morfologia Matemática é definida como uma metodologia para analisar imagens através da *teoria dos conjuntos*, no qual permite a construção de operadores para a

descrição e identificação de objetos (MATHERON, 1967). Para isso, os pixels encontrados na imagem pertencem a um espaço bidimensional de números inteiros $\mathbb{Z} \times \mathbb{Z}$ (\mathbb{Z}^2), em que cada elemento do conjunto é um vetor bidimensional de coordenadas (x, y) .

A Morfologia Matemática é utilizada tanto para imagens binárias (espaço \mathbb{Z}^2) quanto para imagens monocromáticas (espaço \mathbb{Z}^3), no qual além de sua coordenada cartesiana (x, y) é também considerado o valor discreto da intensidade do pixel. Algumas aplicações desta metodologia são: 1) extração de componentes conexos; 2) busca de padrões específicos na imagem; e 3) extração de bordas.

Os conceitos de translação e reflexão de conjuntos são amplamente utilizados em Morfologia Matemática, além de serem a base para a geração dos operadores (seção 2.5.1). A *translação* de A pelo ponto $p = (p_1, p_2)$, denotada por $A + p$, é definida conforme a Equação 2.8.

$$A + p = \{a + p | a \in A\} \quad (2.8)$$

Se o conjunto de pixels representar um objeto em uma imagem, então $A + p$ será o processo de substituição das coordenadas (x, y) por $(x + p_1, y + p_2)$. O processo de translação está detalhado na Figura 7(b) de acordo com o objeto inicial da Figura 7(a).

A *reflexão* de um conjunto A , denotada por \hat{A} , é definida conforme a Equação 2.9.

$$\hat{A} = \{-a | a \in A\} \quad (2.9)$$

Se A for um conjunto de pixels que representam um objeto em uma imagem, então a reflexão \hat{A} será a substituição das coordenadas (x, y) por $(-x, -y)$. O processo de reflexão é ilustrado na Figura 7(c) de acordo com um objeto inicial da Figura 7(a).

2.5.1 Operadores Morfológicos

Um *operador morfológico* é um mapeamento entre o conjunto A que define a imagem e um conjunto b , chamado de *elemento estruturante*. Este elemento é expresso a respeito a uma origem local. Os operadores básicos de Morfologia Matemática são: Dilatação, Erosão, Abertura e Fechamento.

No processo de *Dilatação*, é aplicado a translação do elemento estruturante sobre todas as posições da imagem e, para cada posição transladada, os valores do elemento estruturante são somados aos valores dos pixels da imagem, tomando-se o valor máximo.

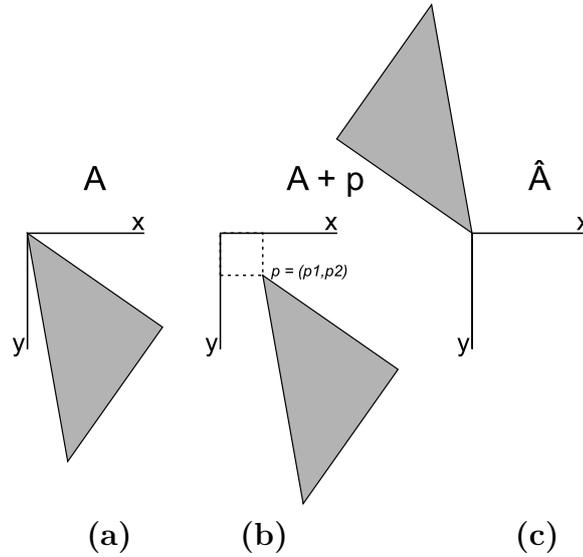


Figura 7 – Operações em conjunto: (a) Imagem original; (b) Imagem Transladada; (c) Imagem refletida.

A Dilatação de uma imagem monocromática f por um elemento estruturante b é definida conforme a Equação 2.10.

$$f \oplus b(x, y) = \max\{f(x - m, y - n) + b(m, n)\} \quad (2.10)$$

em que $(m, n) \in b$, tal que $(m, n) = (0, 0)$ é a origem de b .

Na *Erosão*, de forma análoga a dilatação, ocorre a translação do elemento estruturante sobre as posições da imagem. Para cada posição transladada, os valores do elemento estruturante são subtraídos dos valores dos pixels, levando em consideração o valor de mínimo. A erosão de uma imagem monocromática f por um elemento estruturante b é definida conforme a Equação 2.11.

$$f \ominus b(x, y) = \min\{f(x - m, y - n) - b(m, n)\} \quad (2.11)$$

Na Figura 8 é ilustrado os conceitos de Dilatação e Erosão de acordo com um determinado elemento estruturante.

A *Abertura* é a união de todas as translações de um elemento estruturante b em uma imagem f . Essa operação é dada conforme a Equação 2.12.

$$f \circ b(x, y) = (f \ominus b) \oplus b \quad (2.12)$$

O *Fechamento* da imagem f pelo elemento estruturante b é definido como a

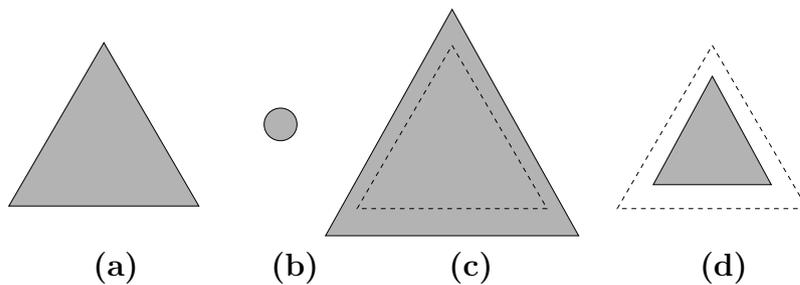


Figura 8 – Aplicação dos Operadores de Dilatação e Erosão: (a) Objeto inicial; (b) Elemento estruturante; (c) Dilatação; (d) Erosão.

dilatação de f por b , seguida pela erosão do resultado por b .

$$f \bullet b(x, y) = (f \oplus b) \ominus b \quad (2.13)$$

2.6 Espaço de Cores

Espaços, ou *modelos de cores*, são representações que viabilizam a especificação de cores em um formato padronizado, a fim de atender diferentes dispositivos gráficos (PEDRINI; SCHWARTZ, 2008).

A utilização de um espaço de cores varia conforme a aplicabilidade do mesmo para tipos específicos de problemas. Isso deve-se ao fato de não existir um modelo que descreva todos os aspectos referentes a coloração de uma imagem. De forma geral, os modelos de cores podem ser agrupados em dois grupos: *aditivos* e *subtrativos*. Em modelos aditivos, as cores são geradas através da combinação dos comprimentos de ondas luminosas, e.g. a cor branca é a adição das cores primárias vermelho, verde e azul e, em contrapartida, a cor preta é representada pela ausência destas três cores. Já em modelos subtrativos, a cor preta é representada pela adição das cores ciano, magenta e amarelo. Na Figura 9 é ilustrado estes dois grupos.

Definiremos, mesmo neste vasto espectro de espaços de cores, três deles: RGB, XYZ, e $L^*a^*b^*$, referentes aos que foram utilizados por esta dissertação.

2.6.1 Espaço de Cores RGB

O Espaço de Cores *RGB* é composto pelas três cores primárias: vermelho (R, red), verde ($G, green$) e azul ($B, blue$). Estas três cores são dispostas em uma plano de coordenadas cartesianas tridimensional. Desta forma, pode-se representar este espaço de cores como um cubo, no qual as três cores primárias estão em três vértices do cubo, e as cores primárias complementares (ciano, magenta e amarelo) estão dispostas em outros

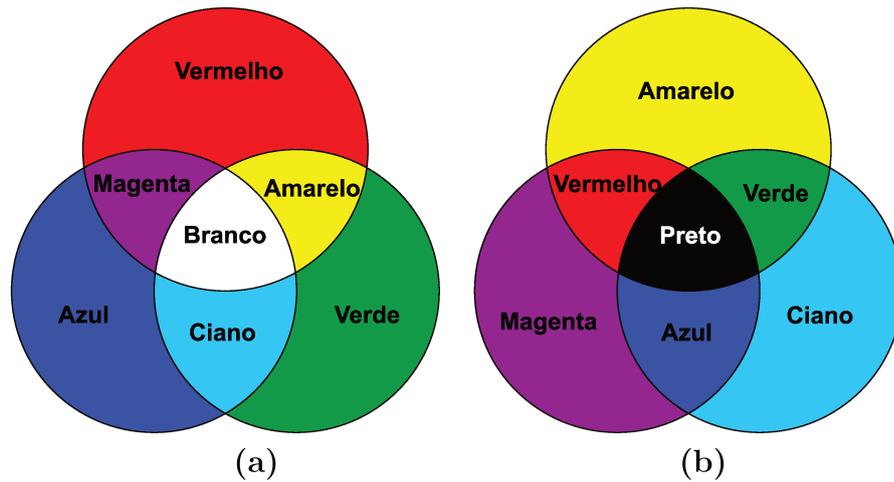


Figura 9 – Modelos de espaços de cores: (a) Espaço aditivo; (b) Espaço subtrativo.

três. Os dois vértices restantes são preenchidos pela combinação das cores primárias, ou ausência delas (branco e preto, respectivamente).

Portanto, a proposta deste cubo é apresentar a combinação de cada tonalidade. Algumas inferências podem ser realizadas diante deste modelo: A tonalidade de cinza é encontrada na diagonal que interconecta a cor branca e a cor preta. Em imagens, é comum encontrar as cores em intervalos de 0 a 254 tons. Na Figura 10 é ilustrado o cubo que caracteriza este espaço de cores.

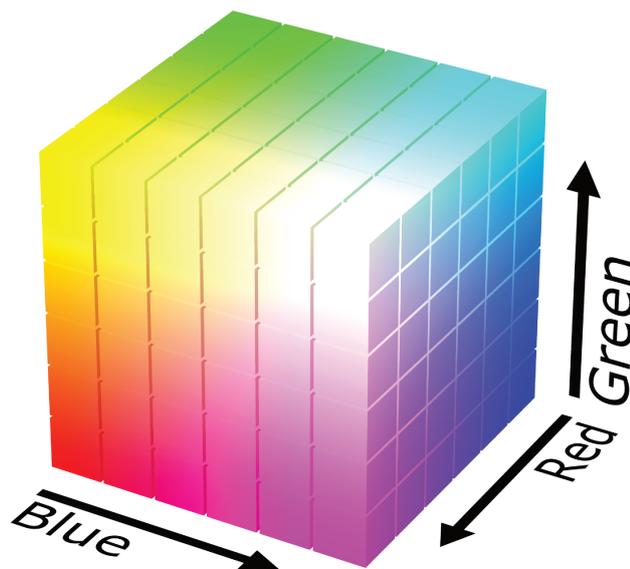


Figura 10 – Espaço de cores RGB

2.6.2 Espaço de Cores XYZ

O Espaço de Cores XYZ foi proposto pela *Commission internationale de l'éclairage-CIE*, normalmente como passo intermediário para obtenção de outros espaços de cores. Este espaço utiliza a combinação aditiva das cores imaginárias X , Y , Z , obtidas através da relação linear com o espaço RGB . Na Equação 2.14 é declarada a conversão do espaço de cores RGB para XYZ .

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4303 & 0.3416 & 0.1784 \\ 0.2219 & 0.7068 & 0.0713 \\ 0.0202 & 0.1296 & 0.9393 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.14)$$

2.6.3 Espaço de Cores $L^*a^*b^*$

O espaço de cores $L^*a^*b^*$, proposto pela CIE em 1976(CONNOLLY; FLEISS, 1997), é derivado do espaço XYZ . O componente L^* corresponde à luminância, os componentes a^* e b^* correspondem à crominância. Na Figura 11 é ilustrada um modelo de interpretação deste espaço. Os canais deste espaço de cores é obtido através das seguintes Equações:

$$L^* = 116f\left(\frac{Y}{Y_0}\right) - 16 \quad (2.15)$$

$$a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right] \quad (2.16)$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right] \quad (2.17)$$

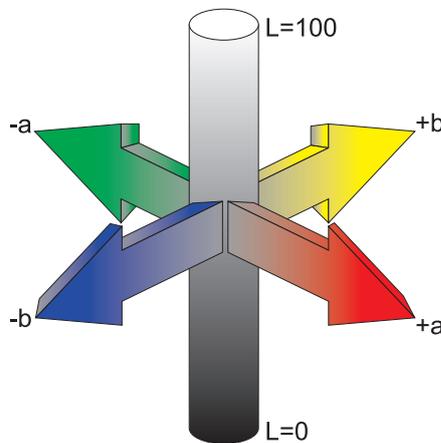


Figura 11 – Espaço de cores $L^*a^*b^*$

2.7 Transformada Watershed

Transformada Watershed é uma técnica de segmentação de imagens ao qual é feita uma analogia à inundação de um relevo topográfico (BEUCHER; MATHMATIQUE, 1991). Uma imagem, portanto, pode ser comparada a esse relevo, no qual o valor de cada pixel da imagem corresponda a uma determinada altitude deste relevo. Assim, os valores mais baixos e mais altos representam, respectivamente, os *vales* e os *picos*. Os valores mais baixos, se comparado com os valores vizinhos, são chamados de *mínimos locais*, e o valor mais baixo em toda a imagem é denominado *mínimo global*. A mesma denominação é aplicada inversamente aos máximos. A Figura 12 apresenta o processo de inundação em um relevo cortado transversalmente.

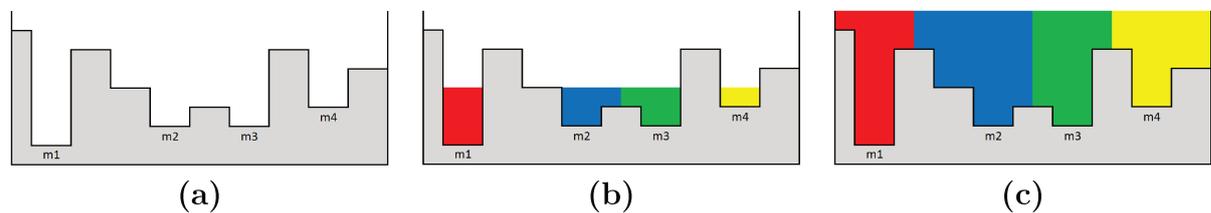


Figura 12 – Processo de inundação em uma imagem cortada transversalmente (m representa os mínimos regionais): (a) Imagem sem a inundação; (b) Imagem com o processo de inundação iniciado com vizinhança 8; (c) Imagem com o processo de inundação concluído

O processo de imersão, ou inundação, é realizado para preencher este relevo, a fim de encontrar as regiões, ou segmentos, da imagem (BEUCHER; MATHMATIQUE, 1991). Cada ponto de inundação é associado a um mínimo encontrado na imagem. Conforme os mínimos (vales) começam a ser inundados, a imagem começa a ser segmentada. As áreas inundadas são chamadas de bacias de retenção. Porém, conforme o processo de inundação ocorre, as bacias de retenção se encontrarão, e quando isso ocorrer, será gerada uma linha de contenção, ou linha de watershed, que separará estas áreas. As linhas são interpretadas como delimitações de regiões na imagem. O processo de imersão acaba somente quando toda a superfície estiver inundada, significando que todas as linhas de contenção foram definidas. Logo, é possível afirmar que ao final do processo será obtido uma imagem segmentada via Transformada Watershed. Na Figura 13 é ilustrada a segmentação de uma imagem digital.

Vários fatores podem afetar a qualidade da Transformada Watershed, como por exemplo, as informações ruidosas. Essas informações corrompem a informação da imagem, fazendo a técnica identificar valores duvidosos e que prejudicam a qualidade final da segmentação.

Na literatura é bastante comum encontrar a Transformada Watershed com algumas modificações, seja na implementação da técnica ou no pré-processamento realizado.

É pertinente relembrar que a Transformada Watershed visa extrair objetos uniformes da imagem, ou seja, com as regiões apresentando leves variações de intensidade (baixos valores em seu gradiente). As modificações da Transformada Watershed serão categorizadas conforme segue:

- *Clássica*: Modelo clássico da Transformada Watershed. A imagem de entrada não necessita, necessariamente, ser a imagem original. As variações, portanto, são referentes a imagem utilizada pela Transformada Watershed, podendo ser, por exemplo, o gradiente da imagem. Na Transformada Clássica é comum ocorrer a chamada supersegmentação, no qual a imagem fica com mais regiões do que ela necessita (Figura 13(c)).
- *Marcadores*: A utilização da Transformada Watershed via marcadores é utilizada para reduzir a supersegmentação. Marcadores são componentes conectados que pertencem a imagem, podendo ser indicados nos objetos de interesse ou no fundo. Estes marcadores podem ser gerados de forma supervisionada, com a interação ou rotulação prévia dos elementos, ou através de algoritmos que os definam através da análise de padrões na imagem. A indicação desses marcadores pode limitar o número de regiões que a Transformada Watershed irá delinear até o fim do processo de segmentação.
- *Hierárquico*: O Watershed Hierárquico é uma modificação da Transformada Watershed que visa controlar e estipular um número de regiões finais para o processo de segmentação. Esse controle é realizado através da Árvore de Lagos Críticos (MEYER, 1996). Detalhadamente, o processo de geração da Árvore de Lagos Críticos é bastante semelhante com a Transformada Clássica, ou seja, o processo de inundação ocorre da mesma forma. Paralelamente, nós vão sendo adicionados a uma árvore a cada vez que uma região é identificada. Desta forma, a árvore gerada armazenará as regiões de forma hierárquica, ou seja, conforme sua formação. Assim, a quantidade de regiões finais dependerá do nível de detalhe que a imagem necessita, estabelecendo para este fim um limiar horizontal, que cortará esta árvore. Tal árvore é ilustrada na Figura 15(c).

Nas Figuras 13, 14 e 15 são mostrados, respectivamente, o uso do gradiente da imagem, dos marcadores e do Watershed Hierárquico.

2.7.1 Transformada Imagem Floresta

Um dos modos de implementar a Transformada Watershed é através do *framework Transformada Imagem Floresta (Image Foresting Transform-IFT*, em inglês). A

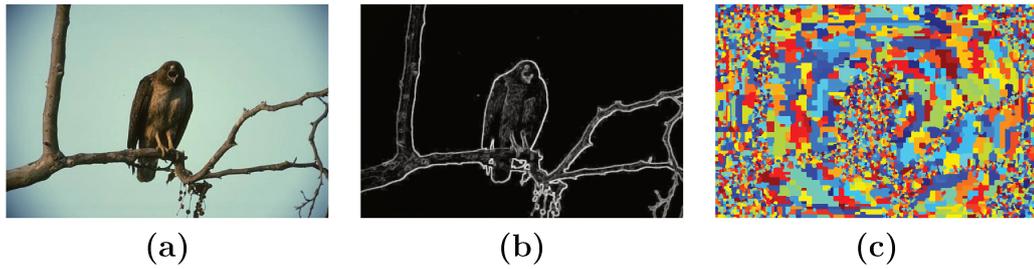


Figura 13 – Processo de inundação aplicado em uma imagem 2D: (a) Imagem Original #42049 da BSDS500; (b) Gradiente Morfológico; (c) Aplicação da Transformada Watershed

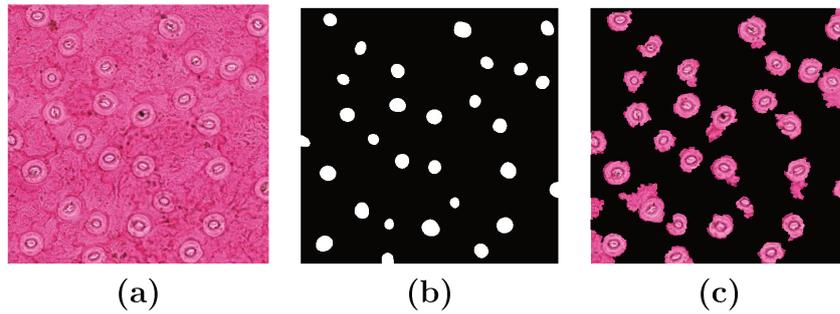


Figura 14 – Imagem da planta *Ugni Molinae* (DUARTE *et al.*, 2017): (a) Imagem Original; (b) Marcadores que representam a localidade dos estômatos; (c) Imagem Segmentada

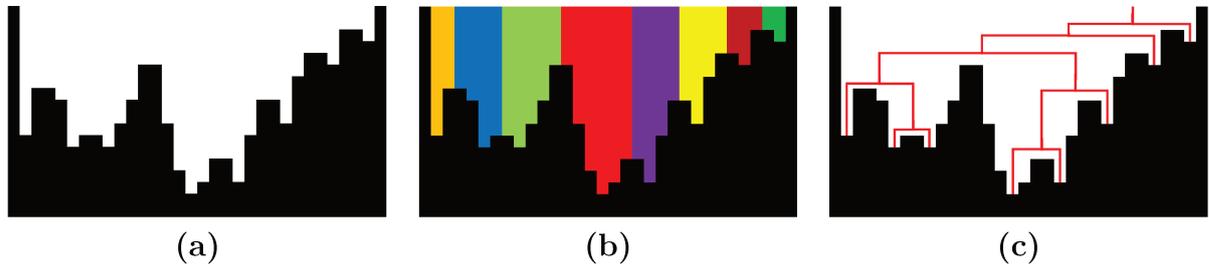


Figura 15 – Geração da Árvore de Lagos Críticos: (a) Imagem cortada transversalmente; (b) Imagem preenchida; (c) Árvore gerada para definir o número de regiões da imagem.

IFT, proposta por Falcão (FALCÃO *et al.*, 2004), é uma ferramenta para avaliação, implementação e modelagem dos operadores matemáticos. Além disso, a IFT visa definir uma floresta de caminho com custo mínimo, interpretando a imagem como um grafo, ao qual os pixels são vértices e a relação de adjacência entre eles são as arestas.

Um caminho π simples em grafos representa a rota ao qual todos os nós são diferentes e adjacentes. Além disso, cada caminho definido em um grafo possui um valor associado ao custo, matematicamente expresso como $c(\pi)$. O custo $c(\pi)$ é determinado através de uma função de custo, que pode estar relacionada, por exemplo, à tonalidade dos pixels na imagem, ao gradiente extraído da imagem, a distância unitária dos nós, entre outros. Em IFT, os caminhos são iniciados em nós particulares, chamados de nós

sementes, ou *seed pixels*, que estão associados a um conjunto de nós sementes S .

A IFT pode ser associada ao *Algoritmo de Dijkstra* (DIJKSTRA, 1959), algoritmo este responsável por identificar o caminho de custo mínimo em um grafo partindo de um nó inicial. A maior alteração está no fato da IFT possuir diversos nós/vértices semente, e não apenas um único. Na Figura 16 é ilustrada uma IFT gerada a partir de um grafo ponderado.

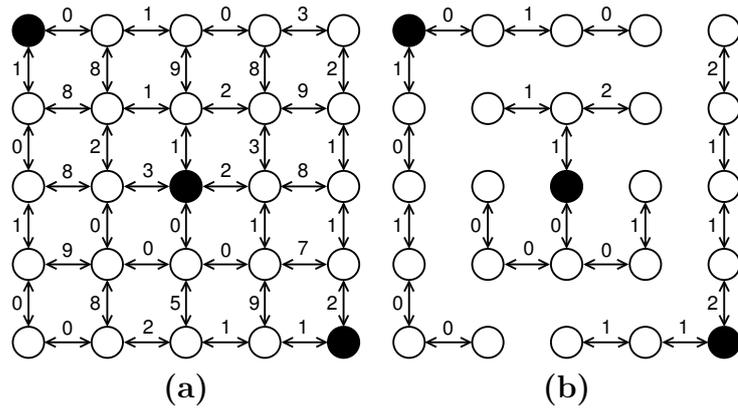


Figura 16 – Obtenção da IFT a partir de um grafo ponderado (pontos escuros representam os *seed-pixels*): (a) Grafo Ponderado; (b) Representação via IFT

2.7.2 Transformada Watershed via IFT

Após a definição das técnicas Transformada Watershed e Transformada Imagem Floresta terem sido relatadas, nesta seção será mostrado o modo de implementar a Transformada Watershed via IFT. Para isso, essa seção apresentará, além do algoritmo, o conjunto de fatores que viabilizam sua utilização.

Para efetivar a implementação da Transformada Watershed pela IFT, é necessário, previamente, definir uma função de custo de caminho. Esta função, também conhecida como função de máximo, está definida conforme a Equação 2.18 (AUDIGIER, 2004).

$$c_{max(\pi)} = \begin{cases} \max_{\forall (p,q) \in \pi} \{\delta(p,q)\}, & \text{se a origem de } \pi \in S \\ +\infty, & \text{caso contrário} \end{cases} \quad (2.18)$$

Tal que π está associado ao caminho no grafo, S representa o conjunto com todos os nós sementes, e $\delta(p,q)$, é a função de dissimilaridade entre os vértices p e q (FALCÃO *et al.*, 2004).

O Algoritmo 2.1 apresenta a representação da Transformada Watershed através da IFT. A definição de um grupo determinado de variáveis, necessários para compreen-

são do Algoritmo, está dividido em três grupos: 1) Variáveis de Entrada; 2) Variáveis Auxiliares; e 3) Variáveis de Saída.

1. Variáveis de Entrada

- a) I : Imagem utilizada (matriz de pixels)
- b) V : Conjunto de vizinhança, onde $V(p)$ apresenta todos os vizinhos do nó p
- c) S : Conjunto de nós sementes.
- d) λ : Função responsável por rotular cada nó semente no conjunto S .
- e) f : Função de custo de caminho que associa o custo $f(\pi_p)$ ao caminho π_p .

2. Variáveis Auxiliares

- a) Q : Fila de prioridade, representando a escolha de um determinado *pixel* para análise.
- b) e : Tabela de estado dos *pixels*, onde associa cada *pixel* a um valor $e(p)$. O valor $e(p)$ é obtido através da seguinte condicional: 1) Se o *pixel* p não foi utilizado e não pertence ao conjunto Q , atribua 0; 2) Se p pertence ao conjunto Q , atribua 1; 3) Se p foi utilizado e não pertence ao conjunto Q , atribua 2.
- c) c : Armazena temporariamente o valor do custo.

3. Variáveis de Saída

- a) C : Mapa de custos (matriz com os custos acumulados do caminho até cada *pixel*).
- b) R : Mapa de rótulos.
- c) P : Mapa de predecessores.

Além das variáveis definidas, é necessário também ressaltar o grupo de funções, todas responsáveis pelo gerenciamento do conjunto Q . A definição de tais funções será detalhada a seguir:

1. $RemoveMin(Q)$: remove o *pixel* de menor custo da fila Q .
2. $Remove(p, Q)$: remove o *pixel* p da fila Q .
3. $Inser(p, Q, c)$: insere o *pixel* p com custo c na fila Q .
4. $EstáVazia(Q)$: Retorna um operador booleano (1 - Fila Q está vazia, 0 - caso contrário).

Algoritmo 2.1 Transformada Watershed via IFT (AUDIGIER, 2004)

```

1:  $\forall p \in I, C[p] \leftarrow \infty; R[p] \leftarrow nil; P[p] \leftarrow nil; e(p) \leftarrow 0$ 
2:  $\forall s \in S, C[s] \leftarrow 0; R[s] \leftarrow \lambda(s); Insere(s, Q, 0); e(s) \leftarrow 1$ 
3: while EstáVazia(Q) = FALSO do
4:    $p \leftarrow RemoveMin(Q); e(p) \leftarrow 2;$ 
5:    $\forall q \in V(p),$ 
6:   if  $e(q) \neq 2$  then
7:      $c \leftarrow MAX(C[p], I[q]);$ 
8:     if  $c < C[q]$  then
9:       if  $e(q) = 1$  then
10:         $Remove(q, Q);$ 
11:       end if
12:        $Insere(q, Q, c); e(q) \leftarrow 1;$ 
13:        $C[q] \leftarrow c; R[q] \leftarrow R[p]; P[q] \leftarrow p;$ 
14:     end if
15:   end if
16: end while

```

Consequentemente, é necessário apresentar um acompanhamento sobre o funcionamento do Algoritmo 2.1. Inicialmente, é realizada a inicialização das variáveis (i.e. custos, mapa de estados, rótulos, predecessores). O algoritmo é executado até que todos os nós tenham sido rotulados e seus predecessores calculados. O processo irá sempre retirar da fila o nó de menor custo (inicialmente advindos dos pixels sementes) para realizar os cálculos. Os nós que são vizinhos ao que está sendo trabalhado, e que ainda não passaram pelo processo, serão também adicionados à fila. Os predecessores sempre adicionam o nó com custo mínimo, o que significa que o caminho até aquele determinado ponto é mínimo. Os custos de um nó serão sempre mínimos. Na sequência, após o nó ter sido trabalhado, o mesmo será retirado da fila.

2.8 Representação da Imagem por Grafos

Na literatura pode-se encontrar as imagens sendo representadas por diversas estruturas computacionais e matemáticas. Uma das estruturas de representação utilizadas são os grafos. Para compreender melhor essa relação, é necessário antes fazer uma breve associação dos termos de ambas as estruturas (conforme as definições da seção 2.2). Em imagens, existem os chamados pixels, que são responsáveis por armazenar as informações de intensidades ou colorações. Portanto, relacionando essa informação com grafos, é possível determinar que os pixels da imagem estão associados aos nós ou vértices do grafo, e as relações de afinidade entre estes pixels na imagem são associadas às arestas do grafo.

Porém, realizar a representação da imagem por grafos por meio da associação

de pixels é um processo custoso. Para isso, em alguns trabalhos (PINTO *et al.*, 2014; CARVALHO *et al.*, 2010b; CARVALHO *et al.*, 2010a) é realizado um pré-processamento inicial na imagem, visando agrupar os pixels em regiões, diminuindo assim o conjunto de nós que serão gerados. Conseqüentemente, os nós, que antes representavam os pixels, passam a representar as regiões. As arestas passam a representar a relação de vizinhança entre as regiões. Na Figura 17 é ilustrada a representação via grafo de uma imagem segmentada via Watershed Hierárquico.

O processo de representação da imagem por um grafo é responsável por associar os *pixels* ou regiões da imagem aos nós do grafo. Os pesos entre os nós são determinados através de critérios no qual, em um grafo de similaridade, quanto maior o valor da aresta, maior a proximidade entre os nós. Em um grafo de distância, quanto maior o valor da aresta, maior será a distância ou dissimilaridade entre os nós.

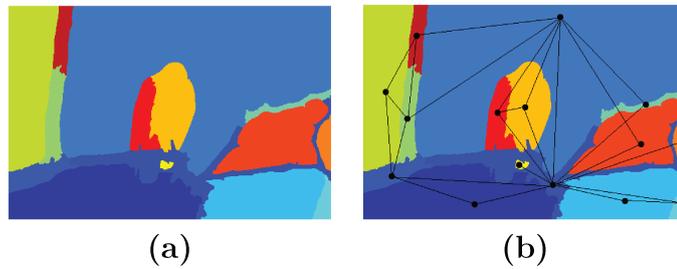


Figura 17 – Modelagem Imagem-Grafo: (a) Imagem com 15 regiões obtidas através da aplicação do Watershed Hierárquico; (b) Grafo de Similaridade

2.9 Particionamento em Grafos

Na Teoria dos Grafos, particionar um grafo conexo G é o processo responsável por dividi-lo em subgrafos obedecendo um determinado conjunto de regras. Esse processo de divisão realiza, portanto, a remoção de determinadas arestas do grafo. Matematicamente, o *corte* em grafos Cut é definido conforme a Equação 2.19.

$$Cut(S_1, S_2) = \sum_{i \in S_1, j \in S_2} w(i, j) \quad (2.19)$$

Tal que S_1 e S_2 são subgrafos de G , $w(i, j)$ é o peso da aresta que conecta o nó i da partição S_1 com o nó j da partição S_2 .

Uma das aplicações de corte em grafos é a segmentação de imagens, ao qual a imagem é representada por um grafo ponderado (CARVALHO *et al.*, 2010a; CARVALHO *et al.*, 2010b; SHI; MALIK, 2000).

Algumas abordagens sobre os critérios de utilização do corte são encontrados na literatura, tais como:

- *Corte Máximo*: Ao invés de analisar o peso total de arestas, deve-se calcular o corte como a fração do peso das arestas de cada partição com o grafo original, removendo a soma do conjunto de arestas de maior peso a fim de se formar dois subgrafos (HADLOCK, 1975);
- *Corte Mínimo*: Obtido de forma analogamente inversa ao Corte Máximo, é removido a soma do conjunto de arestas de menor peso (WU; LEAHY, 1993);
- *Corte Normalizado*: Visa gerar dois subgrafos no qual os pesos de ambos estejam balanceados, removendo o mínimo de arestas possíveis (SHI; MALIK, 2000).

A Figura 18 ilustra a aplicação dos cortes máximo, mínimo e normalizado.

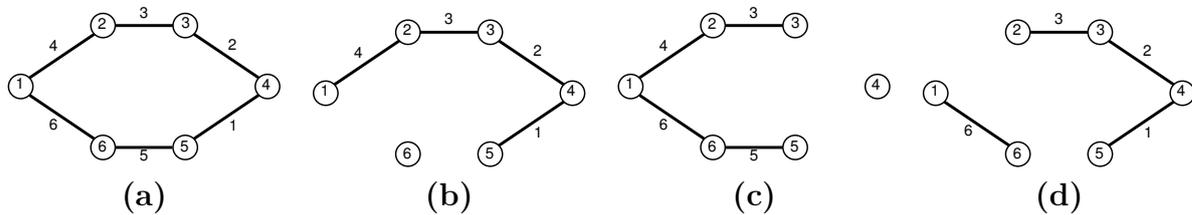


Figura 18 – Tipos de Cortes: (a) Grafo cíclico com 6 nós; (b) Corte Máximo; (c) Corte Mínimo; (d) Corte Normalizado

O Corte Normalizado foi utilizado por este trabalho (Apêndice B) como forma de particionar o grafo. Portanto, será dedicada uma seção apenas para seu detalhamento.

2.9.1 Corte Normalizado

O *Corte Normalizado* foi proposto por Shi e Malik (SHI; MALIK, 2000) como uma alternativa ao uso do corte mínimo, sendo que esse último favorecia, através de seu critério, o desbalanceamento dos números de vértices (ordens) dos subgrafos gerados (WU; LEAHY, 1993). As propriedades algébricas da Matriz Laplaciana são utilizadas pelo Corte Normalizado, objetivando separar os vértices de um grafo de acordo com a dissimilaridade entre eles. O valor do Corte Normalizado ($NCut$) é calculado conforme a Equação 2.20.

$$NCut(S_1, S_2) = \frac{Cut(S_1, S_2)}{SumCon(S_1, V)} + \frac{Cut(S_1, S_2)}{SumCon(S_2, V)} \quad (2.20)$$

tal que $SumCon(S_1, V) = \sum_{i \in S_1, j \in V} w(i, j)$, ou seja, a soma da dissimilaridade de todos os vértices pertencentes ao grafo original V .

Quanto menor for o valor do corte normalizado, melhor será a qualidade da segmentação (SHI; MALIK, 2000). Porém, encontrar tal valor é um problema de complexidade NP-Completo.

A fim de alcançar o valor ótimo do corte normalizado (ou minimamente se aproximar dele), Shi e Malik (SHI; MALIK, 2000) declaram que é possível realizar a expansão da Equação 2.20. Tal expansão explora as propriedades espectrais do grafo, utilizando definições da Teoria Espectral de Grafos (TEG).

A *Teoria Espectral de Grafos* tem por objetivo relacionar as propriedades algébricas do espectro de matrizes representativas dos grafos e suas propriedades estruturais (CVETKOVIC *et al.*, 2009).

O espectro de um grafo G , portanto, pode ser definido como um conjunto dos seus autovalores(λ), geralmente apresentados de forma decrescente, que satisfazem a seguinte Equação:

$$\lambda x = Ax \tag{2.21}$$

Em que A é a Matriz de Adjacência que representa G e x um determinado autovetor. Um determinado autovalor de A é um escalar λ . Todo x que satisfizer esta equação será denominado autovetor de A associado ao autovalor λ . Consequentemente, para calcular os autovalores é necessário encontrar as raízes da equação polinomial $p_G(\lambda) = \det(\lambda I - A)$, onde I é a Matriz Identidade e $p_G(\lambda)$ é o polinômio característico (ABREU, 2005).

Em Shi e Malik (SHI; MALIK, 2000), adaptando o que foi descrito, o Corte Normalizado pode ser minimizado através da resolução da equação $Lx = \lambda Dx$, onde L é a Matriz Laplaciana e D é a Matriz de Graus. Devido ao fato do menor autovalor de L ser sempre 0, o autovetor correspondente ao segundo menor autovalor determina como particionar o grafo.

É pertinente salientar que todos os autovetores associados aos autovalores apresentam informações sobre o particionamento do grafo. Porém, quanto maior o autovetor, menor a confiabilidade das informações (SHI; MALIK, 2000).

Em Shi e Malik (SHI; MALIK, 2000) são propostas duas formas para particionar um grafo via Corte Normalizado:

- *Recursive 2-way NCut (2-way)*: Cada região da imagem é recursivamente particionada utilizando o autovetor correspondente ao segundo menor autovalor do grafo em duas novas regiões. Este cálculo é recursivamente recalculado com as novas matrizes que representariam os subgrafos. Este modo visa aumentar a precisão.
- *Simultaneous k-way Cut with Multiple Eigenvectors (k-way)*: É inserido como parâmetro inicial um número k de regiões. O particionamento ocorre através da utilização dos k autovetores calculados em uma única execução. Este modo visa reduzir o esforço computacional.

Para compreender melhor tais abordagens aplicadas à segmentação de imagens, a seguir será definido um passo-a-passo das abordagens:

1. Construir um grafo $G = (V, E)$ associado a imagem I ; calcular o peso de acordo com uma função de similaridade e armazenar os dados gerados em S e D ;
2. Resolver a equação $Lx = \lambda Dx$ para os autovetores com os menores autovalores;
3. Usar o autovetor correspondente ao segundo menor autovalor para dividir o grafo em duas partições. O particionamento portanto separa os vértices representados neste autovetor através de um limiar t estabelecido, e.g. $t = 0$. O particionamento separará as entradas do autovetor entre valores positivos e negativos;
4. **2-way**: Decidir se as partições correntes devem ser subdivididas novamente e recursivamente particioná-las;
k-way: Faz uso dos demais autovetores para particionar o grafo em k regiões;
5. Mapear as partições para o domínio da imagem I .

2.10 Textura

Um dos principais desafios encontrados ao analisar imagens é definir um conjunto de características que sejam capazes de representar fielmente as regiões contidas em uma imagem. Porém, extrair tais características, apesar de possuir diversas técnicas disponíveis na literatura para fazê-lo, não é algo trivial. Um dos tipos de características que podem ser extraídas e analisadas nas imagens é a textura. Apesar da textura ser um conceito bastante difundido na literatura, ainda não há uma definição padronizada sobre o termo. Cada trabalho encontrado na literatura define textura de acordo com a aplicação que será realizada em um domínio específico. Para Gonzalez e Woods (GONZALEZ; WOODS, 2010), a *análise de textura* pode ser classificada como uma etapa intermediária em uma abordagem de segmentação de imagens ao qual, após estipulados os dados de entrada, são aplicados os descritores de textura a fim de reduzir a quantidade de regiões redundantes na imagem ou mesmo agrupar regiões semelhantes. Embora não exista na literatura um consenso, os métodos que lidam com textura podem ser divididos em abordagens, conforme especificado a seguir:

- **Abordagem estatística**: Essas abordagens buscam representar as texturas através de propriedades não-determinísticas, que normalmente com informações como a distribuição e relacionamento entre as tonalidades ou níveis de cinza na imagem. Através desta breve definição, compreende-se que abordagens estatísticas são uma

forma simples de manter as informações de textura, não mantendo, por exemplo, sua estrutura hierárquica (e.g. Matriz de Coocorrência, Função de Autocorrelação).

- **Abordagem espectral:** Também encontrada como Abordagem Baseada em Processamento de Sinais (PEDRINI; SCHWARTZ, 2008), os métodos de análise de textura são voltados a extrair descritores a partir das representações obtidas através da aplicação de transformadas (e.g. Espectro de Fourier, Transformada Wavelet).

As estruturas que armazenam determinadas características, advindas das imagens, são denominadas *descritores*. Os descritores podem ser qualquer tipo de característica, seja tonalidades, textura, forma, entre outros. Eles são divididos em dois grupos: 1) Extração de características, que representam as propriedades da imagem; e 2) Função de Distância, que representa a similaridade entre regiões ou imagens (ANDALO, 2007).

No presente trabalho foram utilizados dois tipos de características de textura, em duas abordagens distintas, a Matriz de Coocorrência (abordagem estatística) e a Transformada Wavelet (abordagem espectral), definidas a seguir. Logo após, será definido o conceito de Função de Distância ou Similaridade.

2.10.1 Matriz de Coocorrência

A *Matriz de Coocorrência* (*Gray-Level Co-occurrence Matrix-GLCM*, em inglês) é uma estrutura que armazena em matrizes as relações entre os pixels (HARALICK *et al.*, 1973). Essas matrizes possuem um alto índice discriminatório, permitindo a identificação de padrões de áreas texturizadas.

Cada ocorrência de um pixel com tonalidade $m \in I$ para um pixel de tonalidade $n \in I$ é armazenado na GLCM, sendo I o conjunto de pixels da imagem. O tamanho da imagem que está sendo processada é irrelevante, pois o que está sendo armazenado são as informações da tonalidade de cada pixel. Logo, pode-se definir que cada Matriz de Coocorrência possui tamanho $N \times N$, sendo N o maior tom de cinza encontrado na imagem. Entretanto, assim como relatado em Pedrini e Schwartz (PEDRINI; SCHWARTZ, 2008), essa matriz irá perder as informações espaciais da imagem.

É dividida, para fins didáticos, esta seção em duas partes: 1) Construção da Matriz de Coocorrência; 2) Cálculo dos Descritores de Textura.

2.10.1.1 Construção da Matriz de Coocorrência

Define-se $P(m,n)$ todo elemento da Matriz de Coocorrência responsável por armazenar o número de transições de um pixel com tonalidade m para um pixel com tonalidade n , sendo $1 \leq m, n \leq N$ (PEDRINI; SCHWARTZ, 2008). Para isso, define-se

um conjunto V , composto por pares ordenados (m,n) que representam as coordenadas de cada pixel. Após a definição do conjunto V , é necessário percorrer toda a imagem buscando as transições de um pixel m à um pixel n . A função $f(x,y)$ representa a tonalidade de um pixel na posição (x,y) . $P(m,n)$ é definido conforme a Equação 2.22.

$$P(m,n) = \#\{[(i,j), (k,l)] \subset V | f(i,j) = m \text{ e } f(k,l) = n\} \quad (2.22)$$

Após a Equação 2.22 determinar o número de ocorrências das transições, é construída uma matriz conforme ilustrado na Figura 19(b).

As características de textura são obtidas através da *matriz de coocorrência normalizada*, definida através da Equação 2.23. Cada elemento da matriz calculada pela Equação 2.22 é dividido pela soma de seus componentes.

$$p_{m,n} = \frac{P(m,n)}{\sum_{i=0}^{H_g} \sum_{j=0}^{H_g} P(i,j)} \quad (2.23)$$

Na Figura 19(c) é ilustrado um exemplo da matriz de coocorrência normalizada.

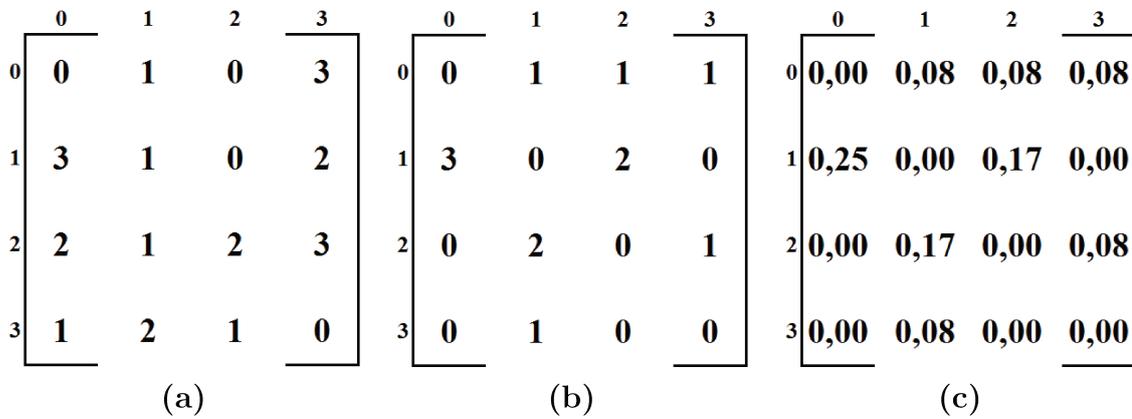


Figura 19 – Matrizes de Coocorrência com $\theta = 0$ e $d = 1$: (a) Matriz Original; (b) Matriz de Coocorrência sem Normalização; (c) Matriz de Coocorrência com Normalização (valores arredondados)

Como declarado em Haralick (HARALICK *et al.*, 1973), outros dois fatores têm grande importância para a geração da matriz de coocorrência: o ângulo θ e a distância d . Ambos os fatores são definidos arbitrariamente e podem ajustar o GLCM de acordo com cada aplicação. Um *ângulo* θ representa a direção que será calculada a ocorrência entre um pixel x e um pixel y . Este ângulo pode ser definido como um entre quatro valores ($0^\circ, 45^\circ, 90^\circ, 135^\circ$), conforme ilustrado na Figura 20. Para cada valor é gerada uma matriz de coocorrência independente. A *distância* d , por sua vez, representa o intervalo de pixels entre a ocorrência de um pixel x para um pixel y . Para fins de exemplo, com $d = 1$, a

ocorrência entre um pixel x e um pixel y deve ser diretamente adjacente, pois ele efetua apenas uma transição.

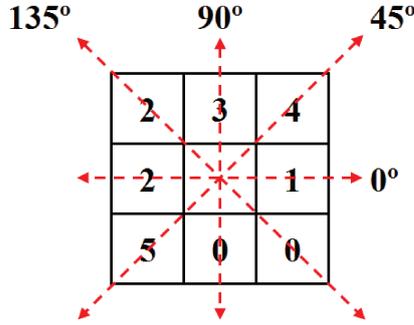


Figura 20 – Orientações referentes aos ângulos para a geração da Matriz de Coocorrência.

Matematicamente, cada ângulo pode ser definido através das seguintes equações:

$$P(i, j, d, 0^\circ) = \#\{[(k, l), (m, n)] \subset V \mid k - m = 0, |l - n| = d, f(k, l) = i, f(m, n) = j\} \quad (2.24)$$

$$P(i, j, d, 45^\circ) = \#\{[(k, l), (m, n)] \subset V \mid (k - m = -d, l - n = d) \vee (k - m = d, l - n = -d), f(k, l) = i, f(m, n) = j\} \quad (2.25)$$

$$P(i, j, d, 90^\circ) = \#\{[(k, l), (m, n)] \subset V \mid |k - m| = d, l - n = 0, f(k, l) = i, f(m, n) = j\} \quad (2.26)$$

$$P(i, j, d, 135^\circ) = \#\{[(k, l), (m, n)] \subset V \mid (k - m = d, l - n = d) \vee (k - m = -d, l - n = -d), f(k, l) = i, f(m, n) = j\} \quad (2.27)$$

Para fins de exemplo, a Figura 21 apresenta a geração das matrizes de coocorrência para cada ângulo definido anteriormente.

A quantidade de matrizes de coocorrência é definida através do produto entre a quantidade de ângulos θ e a quantidade de distâncias d .

2.10.1.2 Cálculo dos Descritores de Textura

Na sequência, após a geração das matrizes de coocorrência, é necessário estipular um meio para extrair as informações de textura nelas contidas. Para isso, Haralick

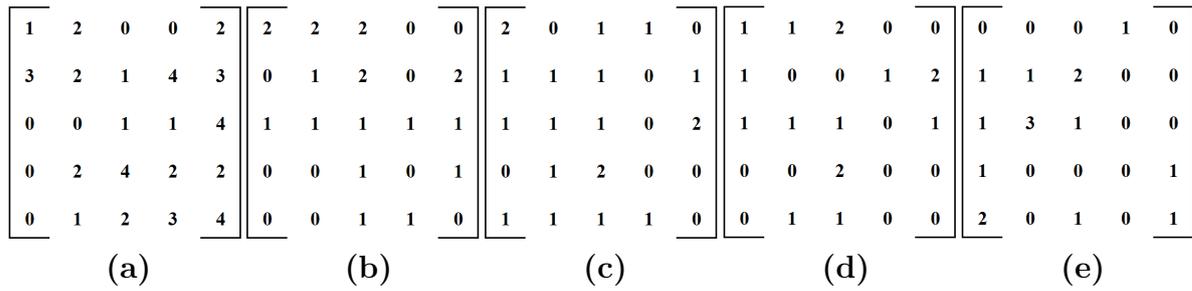


Figura 21 – Matrizes de Coocorrência: (a) Matriz Original; (b) Matriz de Coocorrência com 0°; (c) Matriz de Coocorrência com 45°; (d) Matriz de Coocorrência com 90°; (e) Matriz de Coocorrência com 135°.

(HARALICK *et al.*, 1973) propôs 14 medidas estatísticas que realizam determinados cálculos utilizando estas matrizes. Embora existam diversas medidas para extrair características, Baraldi e Parmiggiani (BARALDI; PARMIGGIANI, 1995) declararam que apenas seis delas apresentavam relevância. Destas seis características, foram utilizadas neste trabalho três delas: Contraste, Energia e Homogeneidade. Essas três características foram as que melhor se enquadraram no modelo de regiões (seção 2.4), ao invés do modelo de pixels.

O *Segundo Momento Angular*, ou *Energia*, determina a uniformidade da textura. Em texturas ásperas, poucos elementos da matriz de coocorrência são diferentes de 0, e se ocorrer, são próximos de 1. O Segundo Momento Angular é definido conforme a Equação 2.28.

$$\text{Segundo Momento Angular} = \sum_i \sum_j (P(i, j))^2 \tag{2.28}$$

O *Contraste* é caracterizado pela diferença entre os tons de cinza. Ou seja, baixo contraste apresenta baixa diferença entre *pixels* adjacentes. Os valores baixos ocorrem quando existe uma grande concentração de valores significativos em volta da diagonal principal. O Contraste pode ser definido conforme a Equação 2.29.

$$\text{Contraste} = \sum_i \sum_j (i - j)^2 * P(i, j) \tag{2.29}$$

Por fim, a *Homogeneidade* assume altos valores quando ocorrem pequenas variações nos níveis de cinza. Essa medida é representada na Equação 2.30.

$$\text{Homogeneidade} = \sum_i \sum_j \frac{P(i, j)}{1 + (i - j)^2} \tag{2.30}$$

2.10.2 Transformada Wavelet

A *Transformada Wavelet* é uma ferramenta para análise em multirresolução caracterizada pela capacidade de localizar, com alto nível de acurácia, representações de tempo, ou frequência, e escala, ou espaço. Em 1989, Mallat (MALLAT, 1989) demonstrou a utilização de *wavelets* ortogonais para definir uma representação completa de uma imagem.

A Transformada Wavelet possui diversas variações encontradas na literatura; dentre elas tem-se a variação *à trous* (HOLSCHNEIDER *et al.*, 1990), que foi utilizada por este trabalho. Porém, apesar de possuir variações, alguns conceitos gerais da Transformada Wavelet devem ser considerados.

A representação wavelet por pirâmides permite analisar a estrutura multiescala, de uma forma bastante completa e simples (BURT; ADELSON, 1983; GONZALEZ; WOODS, 2010). Basicamente, a base da pirâmide apresentará as representações de alta resolução, e o ápice as de baixa resolução. Matematicamente, algumas definições merecem ser enfatizadas, são elas: 1) A base da pirâmide (nível J) possui tamanho $2^J \times 2^J$ e o ápice (nível 0), possui tamanho 1×1 ; 2) Um nível intermediário j possui tamanho $2^j \times 2^j$, com $0 \leq j \leq J$.

Antes de iniciar o passo-a-passo da geração da pirâmide, é necessário definir dois termos, descritos conforme segue:

- **Decimação (ou *downsampling*)**: A operação de decimação descarta termos de uma sequência \vec{u} cujo índice não seja múltiplo de q . Esta operação está formulada na Equação 2.31.

$$(\downarrow q)(\vec{u}) = (\vec{u}_{nq}) \quad (2.31)$$

- **Interpolação (ou *upsampling*)**: A operação de interpolação adiciona zeros intercalados em uma sequência. A definição desta operação está relatada na Equação 2.32

$$(\uparrow q)(\vec{u})(k) = \begin{cases} \vec{u}(k) & \text{se } n = 2k \\ 0 & \text{se } n = 2k+1 \end{cases} \quad (2.32)$$

Na Figura 22(a) é mostrado o modelo da construção da pirâmide de acordo com uma imagem. Na Figura 22(b), a saída da aproximação do nível $j-1$ estabelece as imagens para construir a pirâmide de aproximação. A saída residual de previsão de nível j (para $J-P+1 \leq j \leq J$) é utilizada para a construção da pirâmide de residual de previsão. Em pirâmides residual de previsão, diferente de pirâmides de aproximação, contêm apenas uma resolução reduzida da imagem de entrada, localizada no ápice. É pertinente salientar que ambas as gerações das pirâmides são processos iterativos. Como primeiro passo, a

imagem a ser representada será localizada no nível J . A geração da pirâmide é dividida em três passos, que serão realizados P vezes $(J, J-1, \dots, J-P+1)$:

1. Calcular a aproximação de resolução reduzida da imagem de entrada no nível j . Este processo ocorre pela filtragem e decimação do resultado filtrado por um fator de 2. Posicionar a aproximação resultante um nível acima $(j-1)$ da pirâmide de aproximação.
2. Criar uma estimativa da imagem de entrada no nível j a partir da aproximação gerada no passo anterior, ou seja, realizar a decimação e filtragem. A imagem prevista terá as mesmas dimensões da imagem de entrada do nível j .
3. Determinar a diferença entre a imagem prevista do passo 2 e a entrada do passo 1. Armazenar o resultado no nível j da pirâmide de residual de previsão.

O resultado da aproximação $J-P$ é substituída no mesmo nível na pirâmide de residual de previsão. Caso não haja a necessidade da pirâmide de residual de previsão, os passos 2 e 3, bem como a decimação, a interpolação e o somador podem ser omitidos.

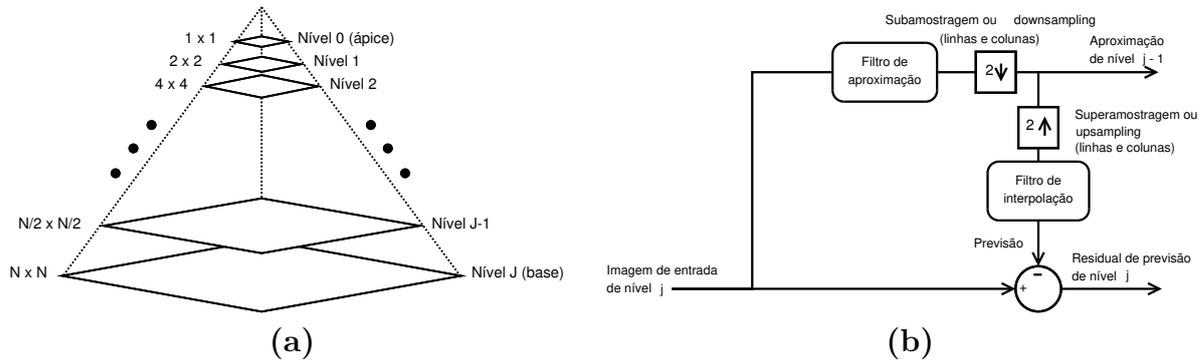


Figura 22 – (a) Aproximação da Pirâmide que representa as escalas da imagem; (b) Sistema simples para geração das pirâmides de aproximação e residual de previsão (GONZALEZ; WOODS, 2010).

Gonzalez e Woods (GONZALEZ; WOODS, 2010) relatam que existe uma variedade de filtros de aproximação e interpolação que podem incrementar ainda mais o processo de geração da pirâmide, tais como média em vizinhança, produzindo pirâmides médias, filtragem passa-baixa gaussiana, produzindo pirâmides gaussianas, entre outros.

A Transformada Wavelet se difere de outras Transformadas por utilizar mais de um núcleo, ou *kernel*, de convolução para obter os coeficientes da transformada. Estes núcleos são obtidos através de funções unidimensionais denominadas função de escala e função wavelet definidas, respectivamente, nas Equações 2.33 e 2.34.

$$\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k) \tag{2.33}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k) \quad (2.34)$$

Na Equação 2.33, para $\forall j, k \in \mathbb{Z}$ e $\phi(x) \in L^2(\mathbb{R})$, com k determinando a posição ao longo do eixo x e j determinando a largura em relação ao eixo x . O termo $2^{j/2}$ controla a amplitude. $L^2(\mathbb{R})$ é o conjunto das funções mensuráveis, de quadrado integrável.

A Equação 2.34 gera basicamente a diferença entre dois subespaços de escalas adjacentes.

Diversas aplicações de Wavelet em PDI utilizam a Transformada Wavelet de Mallat (GONZALEZ; WOODS, 2010). Porém, este modelo não se adequa a certos padrões, por dois motivos principais: 1) Esta transformada não é invariante a rotação; 2) As sub-bandas da imagem, que são a decomposição da mesma em um conjunto de bandas limitadas, não estão relacionadas nos níveis de análise. Assim, para determinadas aplicações, variações da Transformada Wavelet se fazem necessárias.

2.10.2.1 Transformada Wavelet *à trous*

A Transformada Wavelet *à trous* (buracos, em francês) (HOLSCHNEIDER *et al.*, 1990) é uma variação da Transformada Wavelet com determinadas características de interesse, são elas:

- Invariante à translação;
- O conjunto de coeficientes wavelets para cada plano wavelet (ou nível de decomposição) está correlacionado com alto índice de redundância;
- Quase isotrópica;
- As implementações da Transformada Direta e Inversa são bastante simples.

A Transformada Wavelet *à trous* proporciona uma representação em multirresolução, que consiste em imagens de aproximação, ao qual mostra a imagem com resolução grosseira conforme o nível de escala utilizado. Há uma adaptação inerente da análise do tamanho do objeto, com o suporte de um filtro de convolução crescente. Este filtro suaviza a resposta de objetos estreitos em uma dada escala. No primeiro nível, o suporte de filtros é utilizado. A imagem detalhe neste nível tem coeficientes significantes nas localizações quando características significantes de tamanho do pixel estão presentes. Conforme a resolução é diminuída, o filtro de suporte aumenta os coeficientes em tamanho e significância na imagem detalhe. Entretanto, é muito difícil obter as características de interesse da análise em apenas uma imagem detalhe, pois os coeficientes relevantes estão embutidos em coeficientes de detalhe de planos de fundo não especificados.

A decomposição é realizada linha a linha, coluna a coluna, convolucionando cada posição por um número determinado de *kernels*. Em detalhes, partindo da imagem original $A_0(x, y)$, convoluções separáveis apresentam a aproximação $A_1(x, y)$, da qual o plano Wavelet $W_1(x, y)$ é computado por $W_1(x, y) = A_0(x, y) - A_1(x, y)$. Para evitar problemas de descontinuidade das bordas, a imagem é estendida por um espelhamento simétrico. O mesmo processo é repetido recursivamente para as aproximações $A_i(x, y)$, $0 < i \leq J$, com filtro para cada escala i pela inserção de $2^{i-1} - 1$ zeros. Ao final do nível J , tem-se então o conjunto de $J + 1$ imagens, $W = W_1, \dots, W_J, A_J$, também chamada de representação *à trous* wavelet. Tem-se então a Equação de decomposição:

$$W_i(x, y) = A_i(x, y) - A_{i-1}(x, y) \quad 0 < i \leq J \quad (2.35)$$

A decomposição da imagem ocorre de acordo com sua frequência espacial, ao qual cada decomposição representa os atributos de uma determinada faixa de frequência.

A reconstrução é realizada pela soma dos planos, conforme a Equação 2.36.

$$A_0(x, y) = A_J(x, y) + \sum_{i=1}^J W_i(x, y) \quad (2.36)$$

2.10.3 Função de Distância

A *função de distância*, ou *similaridade*, visa medir a distância entre dois pontos (pixels), regiões ou imagens. Esse processo é realizado após a extração das propriedades. Para cada um deles, é definido um valor ou vetor característico, que representará uma determinada informação da imagem.

A definição de uma distância fiel entre dois grupos afetará no quão acurado a abordagem será (BUGATTI *et al.*, 2008). Dentre as funções disponíveis na literatura (e.g. Quadrática, Camberra, Euclidiana), a utilizada por este trabalho (Apêndice B) foi a combinação de duas funções, definidas nas Equações 2.37 e 2.38.

$$W_{ij}^{TC} = \exp\left(\frac{-\|x_i - x_j\|}{\delta^2}\right) \quad (2.37)$$

no qual x_i e x_j são, respectivamente, um determinado critério de similaridade entre as regiões i e j (neste caso, a tonalidade de cinza-TC média das regiões).

$$W_{ij}^{TX} = \exp\left(-\left(\sum_{d=1}^T \frac{\|\tilde{D}_i - \tilde{D}_j\|}{\tilde{D}_i + \tilde{D}_j}\right)\right) \quad (2.38)$$

no qual T representa o número total de descritores de textura (TX), e \tilde{D}_i , \tilde{D}_j representam um determinado descritor aplicado nas regiões i e j , respectivamente.

A combinação entre as funções de similaridade, assim como utilizado em Chang-Ming *et. al.*(CHANG-MING *et al.*, 2008), é realizada conforme descrito na Equação 2.39.

$$W_{ij} = W_{ij}^{TC} \times W_{ij}^{TX} \quad (2.39)$$

3 Considerações Finais

Neste trabalho de dissertação, foram apresentadas duas contribuições para a área de segmentação de imagens. Ambas as contribuições enfatizaram que a inserção das características de textura melhora a segmentação de imagens.

Contribuição 1. Foi proposto um método para segmentação de imagens de estômatos. O método consistia em duas etapas, são elas: 1) a aplicação Transformada Wavelet *à trous* para detecção dos *spots*, que caracterizavam os estômatos; 2) a utilização da Transformada Watershed para que, partindo dos *spots* detectados na etapa anterior, fossem definidas as regiões correspondentes aos estômatos. Os resultados obtidos foram divididos em dois grupos: na detecção dos *spots*, foi obtida a acurácia de 98.25%; na etapa de segmentação dos estômatos, alcançou-se uma média de 70% de acurácia. Em ambas as etapas, a acurácia foi calculada através da medida F. O *gold standard* utilizado necessitou ser implementado. Para isso, foi utilizado o software MATLAB para identificar, pixel a pixel, quão acurada era a segmentação.

Contribuição 2. O método proposto consistia na inserção de características de textura em um *framework* para segmentação de imagens naturais. O processo consistia nos seguintes passos: 1) Aplicar a Transformada Watershed para definir um número de regiões que fosse menor que o *grid* de *pixels*. 2) definir um grafo de similaridade, no qual os nós correspondiam às regiões, e as afinidades entre regiões vizinhas eram definidas através do GLCM. 3) O Corte Normalizado era empregado para particionar o grafo e, por consequência, apresentar a imagem final segmentada. Duas comparações foram efetuadas: a primeira foi comparada com métodos na literatura (e.g. melhoria de 12% em relação a Quadtree); a segunda comparação foi realizada com a versão do *framework* sem o descritor de textura, apresentando uma melhoria de 7%. Ambas as comparações foram realizadas através da medida F. A base de dados utilizada foi a BSDS500, da Universidade da California-Berkeley.

É importante ressaltar que, durante o decorrer desse mestrado, dois descritores de textura foram implementados antes da utilização do GLCM (i.e. Texton, *Local Binary Pattern*). Empregar um descritor em regiões irregulares necessita de ajustes para que o mesmo se adeque a um determinado contexto. Caso contrário, a informação de textura será perdida ou, no mínimo, equivocada. Para ambas as técnicas é necessário a geração de um histograma característico para cada pixel. Ao realizar esta etapa, entretanto, seria necessário definir um histograma para cada região irregular. Portanto, os *bins* (intervalos de valores) necessitariam ser agrupados de tal modo que não comprometessem a informa-

ção de textura nas regiões, além do ajuste necessário na função de similaridade. A soma *bin a bin* dos pixels na região ou a definição de cada região como uma janela irregular não apresentou bons resultados. Portanto, não foi dada continuidade pois identificou-se que as informações estavam comprometidas.

Esta dissertação ainda pode ser explorada em diversos pontos, que serão definidos para as duas abordagens descritas neste trabalho.

No que diz respeito à segmentação de estômatos, as sugestões de trabalhos futuros são:

1. **Alteração no Espaço de Cores:** A conversão utilizada foi CIELab (CONNOLLY; FLEISS, 1997). Porém, é possível utilizar outros espaços de cores, como CMYK, tons de cinza, ou até mesmo manter no espaço RGB.
2. **Detecção dos Pontos:** A detecção foi obtida através da representação multiescala da Transformada Wavelet *à trous*. Entretanto, outros modelos para detecções podem ser implementados, como por exemplo a utilização de Morfologia Matemática (OLIVEIRA *et al.*, 2014).
3. **Segmentação:** Por causa dos marcadores identificados no passo anterior, a técnica que se adequou melhor foi a Transformada Watershed baseada em marcadores. Porém, é possível utilizar os marcadores como parâmetro de entrada para outras técnicas.

Referente à segmentação de imagens naturais, são sugeridos os seguintes trabalhos futuros:

1. **Pré-Processamento:** As imagens foram previamente convertidas para tons de cinza. Porém, é possível utilizar outros espaços de cores, ou mesmo outros gradientes (e.g. Gradiente Morfológico, Gradiente Interno).
2. **Segmentação Inicial:** A segmentação inicial foi obtida pela Transformada Watershed. Esse agrupamento em regiões poderia ter sido realizado por outras técnicas, como *Quadtree* (CARVALHO *et al.*, 2010b) e *Component Tree* (CARVALHO *et al.*, 2010a).
3. **Grafo de Similaridade:** A geração da afinidade entre as regiões foi obtida através da técnica GLCM. Porém, outros descritores poderiam ser implementados, como por exemplo, a Transformada Wavelet.
4. **Particionamento:** A técnica utilizada foi o Corte Normalizado. Outras técnicas que realizam o particionamento de grafos poderá ser implementada neste ponto.

Os tipos das imagens de entrada, para ambas as abordagens, podem ser alterados, sendo necessário apenas alguns ajustes nos parâmetros das abordagens.

De forma geral, as inserções e comparações dos descritores de textura com as abordagens sem os mesmos demonstrou, de forma quantitativa e qualitativa, que existe de fato uma melhora na segmentação. Em ambas as abordagens era possível a aplicação em outros tipos de imagens, mediante a robustez ao qual os métodos apresentavam.

Referências

- ABREU, N. de. Teoria espectral dos grafos: um híbrido entre a Álgebra linear e a matemática discreta e combinatória com origens na química quântica. *Trends in Applied and Computational Mathematics*, v. 6, n. 1, p. 1–10, 2005. ISSN 2179-8451. Disponível em: <<https://tema.sbmac.org.br/tema/article/view/299>>. Citado na página 40.
- ANDALO, F. A. *Descritores de forma baseados em Tensor Scale*. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2007. Citado na página 42.
- AUDIGIER, R. M. M. *Segmentação e Visualização Tridimensional Interativa de Imagens de Ressonância Magnética*. Dissertação (Mestrado) — Universidade Estadual de Campinas, 2004. Citado 2 vezes nas páginas 35 e 37.
- BARALDI, A.; PARMIGGIANI, F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, v. 33, n. 2, p. 293–304, Mar 1995. ISSN 0196-2892. Citado na página 45.
- BEUCHER, S.; MATHMATIQUE, C. D. M. The watershed transformation applied to image segmentation. In: *Scanning Microscopy International*. [S.l.: s.n.], 1991. p. 299–314. Citado 2 vezes nas páginas 16 e 32.
- BUGATTI, P. H.; TRAINA, A. J. M.; TRAINA JR., C. Assessing the best integration between distance-function and image-feature to answer similarity queries. In: *Proceedings of the 2008 ACM Symposium on Applied Computing*. New York, NY, USA: ACM, 2008. (SAC '08), p. 1225–1230. ISBN 978-1-59593-753-7. Disponível em: <<http://doi.acm.org/10.1145/1363686.1363969>>. Citado na página 49.
- BURT, P.; ADELSON, E. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, v. 31, n. 4, p. 532–540, Apr 1983. ISSN 0090-6778. Citado na página 46.
- CARVALHO, M. A. G. d.; COSTA, A. L. d.; FERREIRA, A. C. B.; JUNIOR, R. M. C. Image segmentation using component tree and normalized cut. In: *2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images*. [S.l.: s.n.], 2010. p. 317–322. ISSN 1530-1834. Citado 3 vezes nas páginas 16, 38 e 52.
- CARVALHO, M. A. G. de; FERREIRA, A. C. B.; COSTA, A. L. Image segmentation using quadtree-based similarity graph and normalized cut. In: _____. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 15th Iberoamerican Congress on Pattern Recognition, CIARP 2010, Sao Paulo, Brazil, November 8-11, 2010. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 329–337. ISBN 978-3-642-16687-7. Disponível em: <http://dx.doi.org/10.1007/978-3-642-16687-7_45>. Citado 2 vezes nas páginas 38 e 52.
- CHANG-MING, Z.; GUO-CHANG, G.; HAI-BO, L.; JING, S.; HUALONG, Y. Segmentation of ultrasound image based on texture feature and graph cut. In: *Computer Science and Software Engineering International Conference*. [S.l.: s.n.], 2008. v. 1, p. 795–798. Citado na página 50.

CHEW, S. E.; CAHILL, N. D. Semi-supervised normalized cuts for image segmentation. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 1716–1723. Citado na página 24.

CONNOLLY, C.; FLEISS, T. A study of efficiency and accuracy in the transformation from rgb to cielab color space. *IEEE Transactions on Image Processing*, v. 6, n. 7, p. 1046–1048, Jul 1997. ISSN 1057-7149. Citado 2 vezes nas páginas 31 e 52.

CVETKOVIC, D.; ROWLINSON, P.; SIMIĆ, S. *An Introduction to the Theory of Graph Spectra*. [S.l.]: Cambridge University Press, 2009. (London Mathematical Society Student Texts, Book 75). ISBN 9780521134088. Citado na página 40.

CÂMARA, G.; SOUZA, R. C. M.; FREITAS, U. M.; GARRIDO, J. Spring: Integrating remote sensing and gis by object-oriented data modelling. *Computers Graphics*, v. 20, n. 3, p. 395 – 403, 1996. ISSN 0097-8493. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0097849396000088>>. Citado na página 22.

DERIVAUX, S.; FORESTIER, G.; WEMMERT, C.; LEFÈVRE, S. Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation. *Pattern Recognition Letters*, v. 31, n. 15, p. 2364 – 2374, 2010. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865510002308>>. Citado na página 24.

DIJKSTRA, E. W. A note on two problems in connexion with graphs. *Numer. Math.*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 1, n. 1, p. 269–271, dez. 1959. ISSN 0029-599X. Disponível em: <<http://dx.doi.org/10.1007/BF01386390>>. Citado na página 35.

DUARTE, K. T. N.; CARVALHO, M. A. G. de; MARTINS, P. S. Segmenting high-quality digital images of stomata using the wavelet spot detection and the watershed transform. In: *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, (VISIGRAPP 2017)*. [S.l.: s.n.], 2017. p. 540–547. ISBN 978-989-758-225-7. Citado 2 vezes nas páginas 9 e 34.

FALCÃO, A. X.; STOLFI, J.; LOTUFO, R. de A. The image foresting transform: Theory, algorithms, and applications. *IEEE Trans. Pattern Anal. Mach. Intel.*, IEEE Computer Society, Washington, DC, USA, v. 26, n. 1, p. 19–29, jan. 2004. ISSN 0162-8828. Citado 2 vezes nas páginas 34 e 35.

GIBBONS, A. *Algorithmic Graph Theory*. Cambridge University Press, 1985. ISBN 9780521288811. Disponível em: <<https://books.google.com.br/books?id=Be6t04pgggwC>>. Citado 2 vezes nas páginas 19 e 20.

GONZALEZ, R.; WOODS, R. *Processamento Digital De Imagens*. ADDISON WESLEY BRA, 2010. ISBN 9788576054016. Disponível em: <<https://books.google.com.br/books?id=r5f0RgAACAAJ>>. Citado 6 vezes nas páginas 10, 24, 41, 46, 47 e 48.

HADLOCK, F. Finding a maximum cut of a planar graph in polynomial time. *SIAM Journal on Computing*, Society for Industrial & Applied Mathematics (SIAM), v. 4, n. 3, p. 221–225, sep 1975. Disponível em: <<https://doi.org/10.1137/0204019>>. Citado na página 39.

HARALICK, R. M.; SHANMUGAM, K. *et al.* Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, Ieee, n. 6, p. 610–621, 1973. Citado 3 vezes nas páginas 42, 43 e 45.

HOLSCHNEIDER, M.; KRONLAND-MARTINET, R.; MORLET, J.; TCHAMITCHIAN, P. A real-time algorithm for signal analysis with the help of the wavelet transform. In: _____. *Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990. p. 286–297. ISBN 978-3-642-75988-8. Disponível em: <http://dx.doi.org/10.1007/978-3-642-75988-8_28>. Citado 2 vezes nas páginas 46 e 48.

ISPIRYAN, R.; GRIGORIEV, I.; CASTELL, W. zu; SCHÄFFNER, A. R. A segmentation procedure using colour features applied to images of arabidopsis thaliana. *Functional Plant Biology*, CSIRO Publishing, v. 40, n. 10, p. 1065, 2013. Disponível em: <<https://doi.org/10.1071%2Ffp12323>>. Citado na página 15.

LUCCHI, A.; SMITH, K.; ACHANTA, R.; KNOTT, G.; FUA, P. Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. *IEEE Transactions on Medical Imaging*, v. 31, n. 2, p. 474–486, Feb 2012. ISSN 0278-0062. Citado na página 15.

MALLAT, S. G. A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 11, n. 7, p. 674–693, Jul 1989. ISSN 0162-8828. Citado na página 46.

MARIN, D.; AQUINO, A.; GEGUNDEZ-ARIAS, M. E.; BRAVO, J. M. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on Medical Imaging*, v. 30, n. 1, p. 146–158, Jan 2011. ISSN 0278-0062. Citado na página 24.

MATHERON, G. *Éléments pour une théorie des milieux poreux*. [S.l.]: Masson, Paris, França, 1967. Citado na página 27.

MEYER, F. The dynamics of minima and contours. In: _____. *Mathematical Morphology and its Applications to Image and Signal Processing*. Boston, MA: Springer US, 1996. p. 329–336. ISBN 978-1-4613-0469-2. Citado na página 33.

NETTO, P. O. B. *Grafos: teoria, modelos, algoritmos*. São Paulo SP: E. Blucher, 2012. Citado na página 20.

O, N. K.; KIM, C. Unsupervised texture segmentation of natural scene images using region-based markov random field. *Journal of Signal Processing Systems*, v. 83, n. 3, p. 423–436, 2016. ISSN 1939-8115. Disponível em: <<http://dx.doi.org/10.1007/s11265-015-1030-4>>. Citado na página 15.

OLIVEIRA, M. C. S.; SILVA, N. R.; CASANOVA, D.; PINHEIRO, L. F. S.; KOLB, R. M.; BRUNO, O. M. Automatic counting of stomata in epidermis microscopic images. In: *X Workshop de Visão Computacional*. [S.l.: s.n.], 2014. Citado na página 52.

PAULUS, D.; HORNEGGER, J. *Applied Pattern Recognition: Algorithms and Implementation in C++*. Vieweg+Teubner Verlag, 2003. (Vieweg IT). ISBN 9783528355586. Disponível em: <<https://books.google.ie/books?id=b1s4N98t3ucC>>. Citado 2 vezes nas páginas 23 e 24.

- PEDRINI, H.; SCHWARTZ, W. *Análise de imagens digitais: princípios, algoritmos e aplicações*. THOMSON PIONEIRA, 2008. ISBN 9788522105953. Disponível em: <<https://books.google.com.br/books?id=13KAPgAACAAJ>>. Citado 7 vezes nas páginas 9, 15, 20, 22, 24, 29 e 42.
- PINTO, T. W.; CARVALHO, M. A. G. de; PEDRONETTE, D. C. G.; MARTINS, P. S. Image segmentation through combined methods: Watershed transform, unsupervised distance learning and normalized cut. In: *2014 Southwest Symposium on Image Analysis and Interpretation*. [S.l.: s.n.], 2014. p. 153–156. Citado na página 38.
- PORTELA, N. M.; CAVALCANTI, G. D.; REN, T. I. Semi-supervised clustering for {MR} brain image segmentation. *Expert Systems with Applications*, v. 41, n. 4, Part 1, p. 1492 – 1497, 2014. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417413006672>>. Citado na página 24.
- SHI, J.; MALIK, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 22, n. 8, p. 888–905, ago. 2000. ISSN 0162-8828. Disponível em: <<http://dx.doi.org/10.1109/34.868688>>. Citado 4 vezes nas páginas 16, 38, 39 e 40.
- WILSON, R.; WATKINS, J. *Graphs: an introductory approach : a first course in discrete mathematics*. Wiley, 1990. ISBN 9780471615545. Disponível em: <<https://books.google.com.br/books?id=cevuAAAAMAAJ>>. Citado 2 vezes nas páginas 19 e 20.
- WU, Z.; LEAHY, R. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 15, n. 11, p. 1101–1113, Nov 1993. ISSN 0162-8828. Citado na página 39.
- YU, P.; QIN, A. K.; CLAUSI, D. A. Unsupervised polarimetric sar image segmentation and classification using region growing with edge penalty. *IEEE Transactions on Geoscience and Remote Sensing*, v. 50, n. 4, p. 1302–1317, April 2012. ISSN 0196-2892. Citado na página 24.
- ZHANG, P.; LI, M.; WU, Y.; GAN, L.; LIU, M.; WANG, F.; LIU, G. Unsupervised multi-class segmentation of {SAR} images using fuzzy triplet markov fields model. *Pattern Recognition*, v. 45, n. 11, p. 4018 – 4033, 2012. ISSN 0031-3203. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031320312001999>>. Citado na página 24.

Apêndices

A Segmenting High-Quality Digital Images of Stomata Using the Wavelet Spot Detection and the Watershed Transform

Segmenting High-quality Digital Images of Stomata using the Wavelet Spot Detection and the Watershed Transform

Kauê T. N. Duarte, Marco A. G. de Carvalho and Paulo S. Martins

University of Campinas (UNICAMP), School of Technology, R. Paschoal Marmo, 1888, 13484 Limeira, Brazil
kaue.unicamp2011@gmail.com, {magic, paulo}@ft.unicamp.br

Keywords: Stomata, Wavelets, Automatic Counting, Watershed, Image Processing, Image Segmentation.

Abstract: Stomata are cells mostly found in plant leaves, stems and other organs. They are responsible for controlling the gas exchange process, i.e. the plant absorbs air and water vapor is released through transpiration. Therefore, stomata characteristics such as size and shape are important parameters to be taken into account. In this paper, we present a method (aiming at improved efficiency) to detect and count stomata based on the analysis of the multi-scale properties of the Wavelet, including a spot detection task working in the CIELab colorspace. We also segmented stomata images using the Watershed Transform, assigning each spot initially detected as a marker. Experiments with real and high-quality images were conducted and divided in two phases. In the first, the results were compared to both manual enumeration and another recent method existing in the literature, considering the same dataset. In the second, the segmented results were compared to a gold standard provided by a specialist using the F-Measure. The experimental results demonstrate that the proposed method results in better effectiveness for both stomata detection and segmentation.

1 INTRODUCTION

A stoma (also stomate; plural stomata) is a single pore in the epidermis of leaves and other organs in plants that is used to control gas exchange. The size of its opening is regulated by two surrounding guard cells (Willmer and Fricker, 1996). The word stomata means mouth in Greek, due to the fact that it is responsible for the interaction between the internal and external plant environment, and it provides the carbon dioxide and oxygen used in the photosynthesis and respiration. Given that each type of plant has correspondingly a characteristic type of stomata with its own unique size, density and distribution, the detection and analysis of stomata poses a challenge.

Recently, the digital image processing of stomata has received significant attention from researchers in the academia, thus supporting the identification of the level of environmental stress suffered by plants (Laga et al., 2014)(Jian et al., 2011)(Oliveira et al., 2014). For example, the work by (Violet-Chabrand and Brendel, 2014) addresses the evaluation of stomata density and the authors opt for processing groups of stomata instead of individual ones. In essence, there are several methods that aim at quantifying as well as segmenting the stomata present in plant epidermis.

The human visual perception of the context that

we are part of is different from how the computational processes interpret the same type of information. Human beings possess the innate capability of identifying patterns. It is not so straightforward to transfer this capacity to the computational context. However, a simple conversion of the color space to CIELab, as described by Connolly and Fleiss (Connolly and Fleiss, 1997), allows the computer to simulate human vision.

Spots are important features within an image context. They usually correspond to bright small objects such as cells that stand out locally in their neighborhood. The application of computational processes to automatically count and segment spots and cells (i.e. stomata in our case) is not a trivial task and it clearly offers the domain experts and field practitioners the needed support to the identification of anomalies. By not employing automatic routines for processing stomata (e.g. counting, segmentation, estimation of density), one must rely on manual procedures which are not always viable due to the amount of time taken (Stepka, 2013). There has been considerable work in the past years dedicated to counting cells on digital biological images (Stepka, 2013; Oliveira et al., 2014; Olivo-Marin, 2002; Lojk et al., 2014; Mallat, 1989; Venkatalakshmi and Thilagavathi, 2013). However, the issue of stomata detection still remains largely an

open gap in the literature.

The goal of this paper is to present a technique to count and segment stomata based on the Wavelet Spot Detection and the Watershed Transform. Besides, we also intend to increase the accuracy of the techniques that perform this task. Figure 1 illustrates a high-resolution microscope image (through a combined set of lens) of a *Ugni Molinae* species, with 40x magnification.

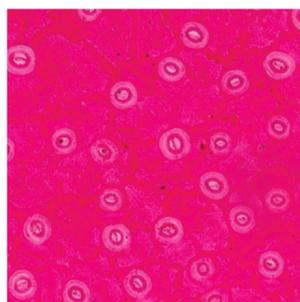


Figure 1: Stomata from *Ugni Molinae* Species.

The remainder of this paper is organized as follows: In Section 2 we review previous work. The proposed approach is described in Section 3. In Section 4 the results are presented and discussed. We summarize and present our conclusions in Section 5.

2 RELATED WORK

Counting cells plays a key role in stomata analysis, and it may include tasks such as segmentation and classification. Within this context, Venkatalakshmi and Thilagavathi (Venkatalakshmi and Thilagavathi, 2013) introduced a method to automatically count red cells (Red Blood Cell Count - RBC), which is used as a means to diagnose health issues in patients. The following steps were defined: 1) Preprocessing, which converts the color space to HSV, and then only uses the S component; 2) Segmentation based on the thresholding of the histogram and on mathematical morphology operators; 3) Combination of the morphological operators, logic and the Hough transform technique to extract the RBC from other cells and from the background; 4) Quantification of the RBC in the image and 5) Definition of the interface for the physician's analysis. The results obtained were compared to manual detection and sent to a doctor for patient analysis.

A general automated approach for detecting and segmenting blood cells of different species was proposed by Karel (Stepka, 2013). His approach can also be directly used in chamber grids. The proposed

method comprises of four main steps: 1) preprocessing, consisting of gray level transformation and noise removal by use of Gaussian filtering; 2) use of Hough Transform in order to identify linear structures; 3) finding cell samples; 4) matching between the cell samples and the other cells present in the image by means of correlation. The author provided experimental results that showed good performance in comparison to manual counting.

The method proposed by Oliveira (Oliveira et al., 2014) is used to detect stomata using mathematical morphology. The images used by the authors are from five types of plants. The method consists of the following steps: 1) application of the Gaussian filter; 2) application of the morphology operators *opening-by-reconstruction* and *opening-closing-by-reconstructions*; 3) use the regional minima and 4) elimination of segments that exceed a threshold perimeter value. The results show an average precision around 94%.

Olivo-Marin proposed a new method to detect luminous (bright) points in fluorescent images from biological immunomicroscopy experiments (Olivo-Marin, 2002). The method provides a count of the number of points using the Undecimated Wavelet Transform and Spot Detection. The approach used is to decompose the Wavelet and define, in all levels, the combination of standards, better identifying the spots across different resolutions. The advantage of the method is the satisfactory identification of spots, besides the fact that it may be applied to several types of images.

Lojk (Lojk et al., 2014) proposed a method for automatic and semi-automatic counting in images from fluorescent microscopy. It consists of the following steps: 1) increase of the contrast with the Contrast Limited Adaptive Histogram Equalization algorithm; 2) application of the Otsu Threshold to convert the image to binary format; 3) image segmentation using the Watershed Transform algorithm (allowing cell overlapping), and 4) blob detection. The accuracy of the method was estimated as above 91% in comparison to manual procedures.

Jian (Jian et al., 2011) proposed a remote sensing processing technology using classification. Their goal was to estimate the stomata density of *Populus euphratica* leaves. Their method was defined as follows: 1) Each image is acquired and preprocessing is applied in order to improve visualization. Three specimens are collected from distinct parts of the leaf; 2) Each image is classified using an object-oriented procedure and parameters such as shape, scale and compactness are manually defined; 3) The results are imported into ArcGIS in order to calculate the grid num-

ber for each stomata in all images; 4) The root mean square error is calculated resulting in an accuracy of 98% (of tested samples) in relation to the manual procedure.

Laga (Laga et al., 2014) proposed an supervised approach to stomata detection that measures both their morphologic and structural features. They defined the stomata aperture width as well as guard cell length and width. These parameters are important since they allow a better understanding of water loss associated with the opening and closing of guard cells. The goal was to count the number of stomata and measure their morphological features. The results were obtained from 24 images and were very close to the manual counting. The approach was defined as follows: 1) Image acquisition: It was considered a flag leaf of a wheat plant captured by a “Leica AS LMD” laser dissection microscope; 2) Stomata cell detection, which is based on template fitting. The templates are stored in a database and are distinguished by shape, size and orientation. In this step, the image is preprocessed in order to convert it into a representation that enhances its initial conditions, and thus facilitates stomata detection; 3) Measurement of stomata features: Once the stomata regions have been identified, the method proceeds by measuring the length and width of the stomata aperture and its related guard cells., and 4) Detection refinement: Stomata detection may result in a number of false positives. Because of that, constraints on stomata morphology were used to discard them.

Our work differs from previous works by the following features: 1) Stomata counting: we use the wavelet transform and spot detection (Olivo-Marin, 2002) in order to detect and count stomata in digital images; previous work use different approaches; 2) Spot detection: compared to (Olivo-Marin, 2002), we adapted the wavelet spot detection to work with the detection of dark points instead of bright spots; 3) High-quality images: the set of images adopted in our work were acquired from a high-resolution microscope; 4) CIELab colorspace: the use of this colorspace improves the efficiency of the method, since the a^* channel provides better information about the stomata cells, and 5) Watershed Transform: the segmentation process was used to improve the identification of stomata, enabling various future calculations such as stomata area and density.

3 PROPOSED APPROACH

In this section we present the proposed method to detect and segment stomata images. The process con-

sists of three steps: 1) acquisition and extraction; 2) cell detection, and 3) image segmentation. Each step is labeled accordingly in Figure 2 and discussed in turn in the following paragraphs.

3.1 Acquisition and Extraction

A total of 64 images were acquired by a high-precision Hamamatsu Nanozoomer XR microscope (Hamamatsu, 2016) and used in our process. Each image has between 20 and 30 GB and 135168 x 78848 pixels. The initial task was to crop the original image and to analyze each crop individually.

Each image is composed of three channels, i.e. R (red), G(green) and B(blue) (Figure 3,a-c). However, obtaining relevant information through these channels alone is not trivial. Therefore, the conversion of color space from RGB to CIELab is required. Each $L^*a^*b^*$ channel has a distinct meaning: L^* represents lightness (0 for black and 100 for white) and a^* and b^* are chromatic values. This process is carried out in two phases (Connolly and Fleiss, 1997): 1) conversion from RGB to XYZ, and 2) conversion from XYZ to CIELab. The trichromatic space was calculated as follows:

$$X = R * 0.4303 + G * 0.3416 + B * 0.1784 \quad (1)$$

$$Y = R * 0.2219 + G * 0.7068 + B * 0.0713 \quad (2)$$

$$Z = R * 0.0202 + G * 0.1296 + B * 0.9393 \quad (3)$$

The values of $L^*a^*b^*$ were calculated as follows:

$$L^* = 116f\left(\frac{Y}{Y_0}\right) - 16 \quad (4)$$

$$a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right] \quad (5)$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right] \quad (6)$$

where: X_0 , Y_0 and Z_0 are the trichromatic values for lightness, which is D50 (see Figure 3 - bottom half, as an example of separate channels $L^*a^*b^*$).

Through the analysis of each channel independently, we may decide which channel to use in the detection and counting of stomata, i.e. the proposed analysis elicits which channel may allow a better visibility of stomata. In the bottom half of Figure 3 we may clearly identify the center of stomata, which is emphasized in black. Therefore, this is the channel that is used throughout the process.

3.2 Cell Detection

Having extracted the second channel in the CIELab color space, we move on to determine the number of

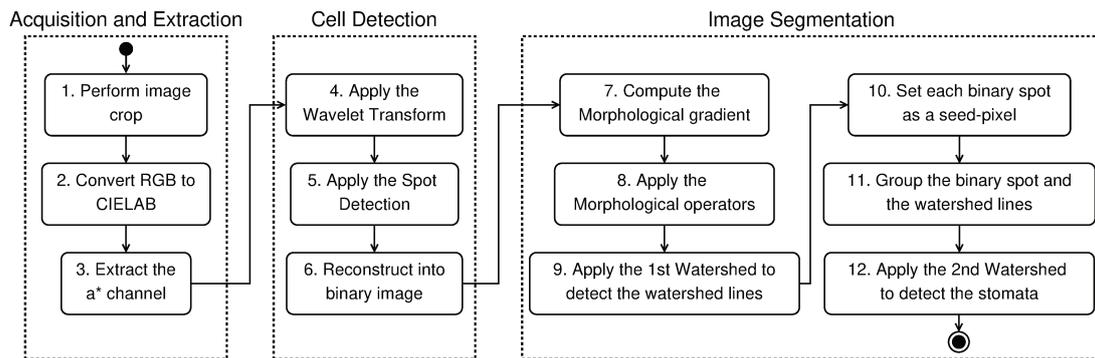


Figure 2: Activity Diagram of the Proposed Approach.

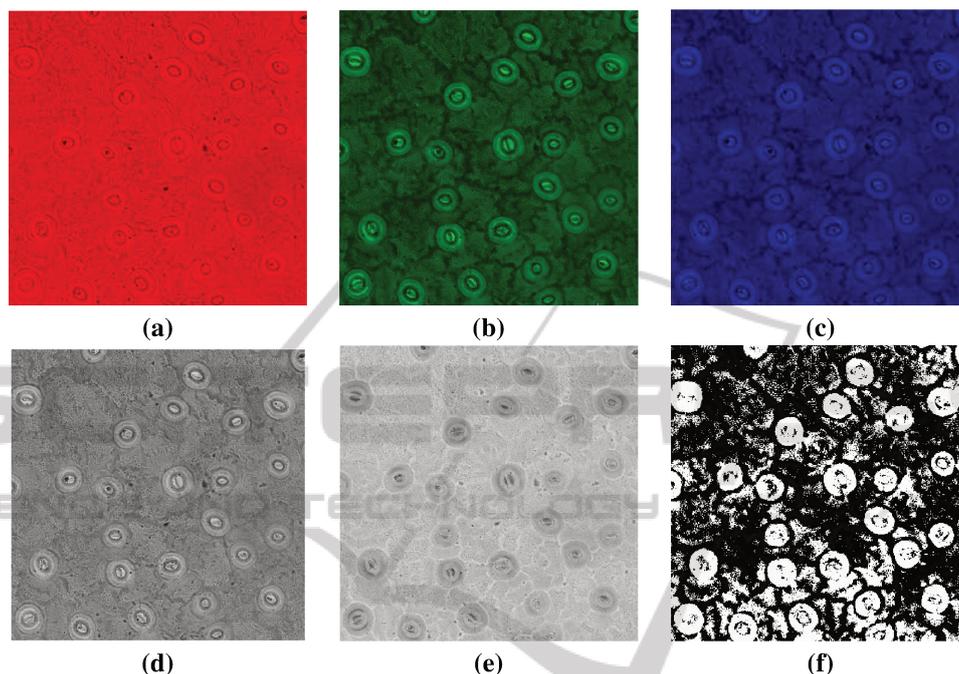


Figure 3: Stomata from *Ugni Molinae Species* in RGB to CIELab (a) RGB-channel R; (b) RGB-channel G; (c) RGB-channel B; (d) CIELab-channel L; (e) CIELab-channel a^* ; (f) CIELab-channel b^* .

stomata cells by applying the Wavelet Spot Detection algorithm (Olivo-Marin, 2002). The Wavelet Transform is based on the “à trous wavelet algorithm”, a multi-scale representation where each level i is computed as follows:

$$W_i = A_{i-1}(x, y) - A_i(x, y) \quad (7)$$

where $0 < i \leq J$, $A_0(x, y)$ is the original image, $A_i(x, y)$ is the image after the convolution of $A_{i-1}(x, y)$ with a specific and increased kernel, and J is the number of scales. The $W_i(x, y)$ wavelet coefficients were then filtered out, thus reducing the influence of noise and eliminating the non-significant coefficients. Finally, a correlation image $PJ(x, y)$ is computed by means of the direct product of the wavelet coefficients at all

J scales. The spot detection is obtained, therefore, after a thresholding of the $PJ(x, y)$ image. Contrary to existing work which detects the bright points, our work focus on detection of darker spots since they are easier to be identified in the correlation image after the conversion from RGB to CIELab.

3.3 Image Segmentation

After the spot detection analysis, we applied the well-known Watershed Transform, used to find non-uniform image contours. However, we used a slightly adapted method proposed by Meyer (Meyer, 1994). Firstly, we used the a^* channel from the converted image to extract a morphological gradient (dilate -

erode), as shown in Figure 4(c). Secondly, it is important to notice that stomata images have texture on their epidermis, what precludes the direct (or immediate) application of the Watershed. Therefore, six morphological operators were applied into the same a* channel prior to the Watershed, in this order: Open, Erode, Reconstruct, Close, Dilate and Reconstruct, using a kernel $k = 2$. These operators were applied to smooth the image.

Figure 4(d-g) illustrates the images resulting from the application of the morphological operators. Then, we applied the first Watershed producing an image with watershed lines. We override the a* channel image with both watershed lines and spots location.

Next, we used a second Watershed to find the regions using the spots as markers, as illustrated in Figure 4(h). Finally, we combined the mask obtained from the previous step with the original image to generate the result shown in Figure 4(i).

4 RESULTS

In this section we present the results divided into two groups: 1) Detection/Counting, and 2) Segmentation. The first group is a necessary (previous) step for the second one. Both counting and segmentation are applied to the same dataset containing 64 image crops (1024 x 1024 pixels each image). The same dataset was used in both studies.

For performance evaluation of the counting and segmentation approaches, we use recall, precision (Baeza-Yates and Ribeiro-Neto, 1999) and the F-Measure (Arbelaez et al., 2011), which are defined as follows:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$F = \frac{2 * Recall * Precision}{Recall + Precision} \quad (10)$$

where TP is a true positive, FN is a false negative, and FP is false positive. Note that in segmentation, a TP indicates that a pixel is identified as representing a stoma by both algorithm and gold standard; FN corresponds to the case where the algorithm finds a background pixel whereas the gold standard found a stoma pixel; A FP occurs whenever the algorithm identifies a stoma pixel, but it is in fact a background pixel set by the gold standard.

In counting, a FP occurs when a dot is found in an area where there is no stoma. A single dot found within the actual stoma area represents a TP . Finally,

a FN represents the absence of a dot in an actual stoma area.

Table 1 shows a sample with 5 randomly chosen images, retrieved from the 64 image-set that was used in this work. Its purpose is only to convey an idea of how the results were calculated, i.e. the full range of results are actually shown in the following subsection.

4.1 Detection and Counting

In the detection and counting step, we compare our approach with two others methods: 1) manual, considered a gold standard, and 2) the method by Oliveira et al (Oliveira et al., 2014). In order to test the proposed method, we discarded all stomata that were split in two across the center in the cropped image before applying it to the image crops (1024 x 1024 pixels). Those stomata that were only partially cut by the crop (but with the whole center preserved) were considered. The results are shown in Table 1, which compares the performance of the algorithm versus the manual procedure using the recall and precision measures.

In Figure 4(b) we illustrate the results in terms of the stomata identified directly on the image of the plant tissue. As we have mentioned earlier, we used 64 cropped images of the plant (*Ugni Molinae*), and the results show an improvement over the work by Oliveira (Oliveira et al., 2014). The proposed approach was compared to manual detection for all the samples, and the following results were obtained:

1. *Recall*: The average recall reached by our method was 98.24% against 95.13% obtained by Oliveira (Oliveira et al., 2014). Furthermore, recall was improved upon the latter work in 84.37 % of the images.
2. *Precision*: The average precision reached by the proposed method was 98.34% against 92.81 % from Oliveira (Oliveira et al., 2014). Moreover, compared to the latter work, the proposed method showed a more satisfactory precision in 90.62% of images.
3. *F-Measure*: The average F-Measure found by our method was 98.25% against 93.80% from Oliveira (Oliveira et al., 2014). It is also important to notice that 84.37% of images scored an improved F-measure in comparison to latter work. The graph shown in Figure 5 illustrates the comparison of the methods for each image using the F-Measure.

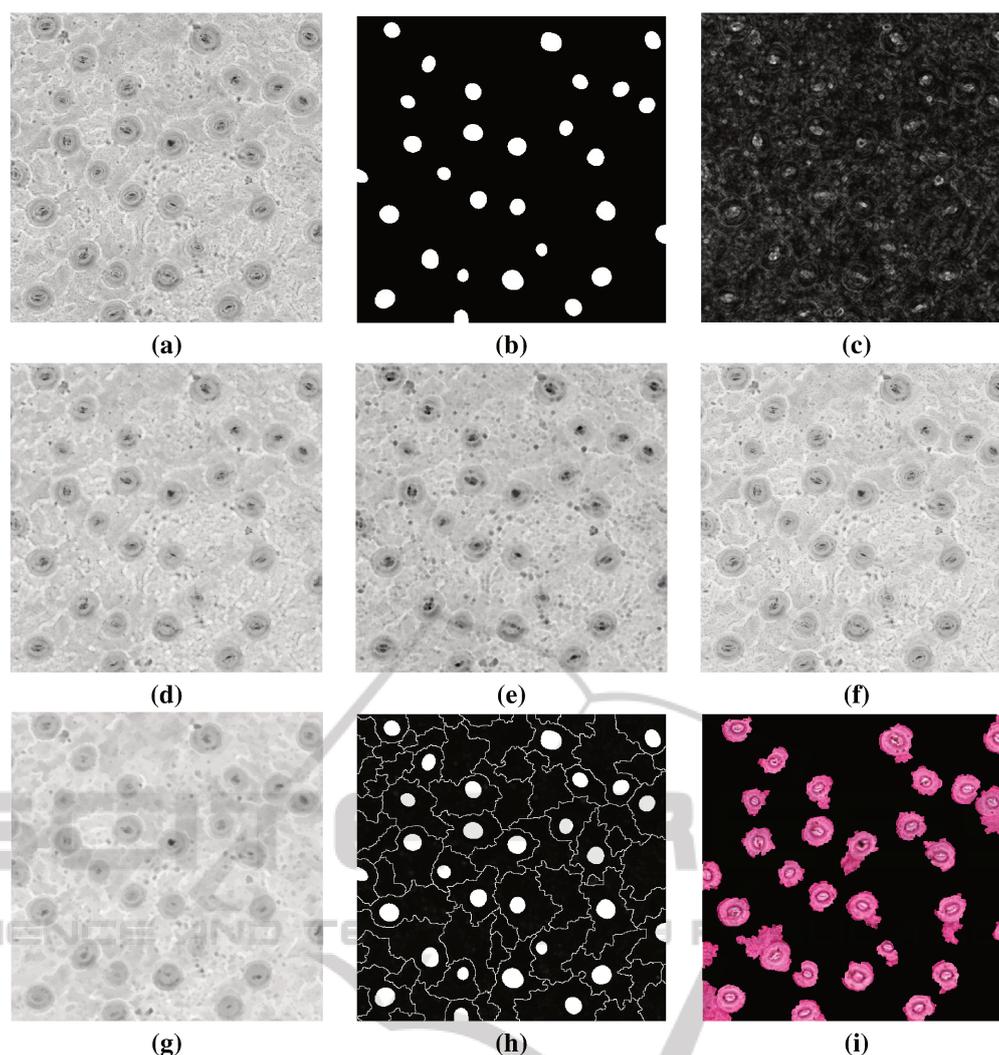


Figure 4: Stomata Segmentation Process from *UgniMolinae* Species (a) Channel a^* ; (b) Binary spots from Wavelet Spot Detection; (c) Morphological Gradient; (d) Open; (e) Erode; (f) Reconstruct; (g) Close; (h) Watershed lines and spots; (i) Stomata detected. The images (a) and (c-g) had their contrast enhanced for better visualization.

4.2 Segmentation

The segmented image was compared to the gold standard pixel by pixel. For the test, we also discarded all split stomata that occur during the crop process. The gold standard was prepared by a specialist using the Interactive Segmentation Tool (McGuinness and O'Connor, 2010), and then compared using the F-Measure.

This new approach reaches an F-Measure of 70%. Figure 6 shows a graph that represents the relationship between Recall, Precision and F-Measure for each tested image.

5 SUMMARY AND CONCLUSION

The analysis of stomata count is a crucial activity in the assessment of plants and the state-of-the-art in this field is still limited. In this work we introduced a new method for detection/counting and segmentation of stomata using the conversion of the color space, the Nondecimated Wavelet Transform for the detection of dark spots (i.e. stomata), and the Watershed Transform. We conducted experiments with high quality digital images and the results obtained indicated a satisfactory determination of the number of stomata in the tissue. The first case of the comparison analysis showed that our approach provided precision and re-

Table 1: Performance analysis of the proposed approach in comparison to the gold standard. N_h is the number of stomata detected by the specialist; N_a represents the number of stomata detected automatically; FP is the number of false positives; FN is the number of false negatives; P is the Precision; R represents Recall; F is the F-Measure.

Image #	Manual	Our Method							Oliveira (2014)						
	N_h	N_a	TP	FP	FN	R	P	F	N_a	TP	FP	FN	R	P	F
1	27	29	27	2	0	93.1	100	96.4	30	25	5	2	83.3	92.6	87.7
2	24	23	23	0	1	100	95.8	97.8	21	20	1	4	95.2	83.3	88.8
3	21	21	21	0	0	100	100	100	23	20	3	1	86.9	95.2	90.8
4	23	24	23	1	0	95.8	100	97.8	25	23	2	0	92	100	95.8
5	25	27	25	2	0	92.6	100	96.1	20	20	0	5	100	80	88.8

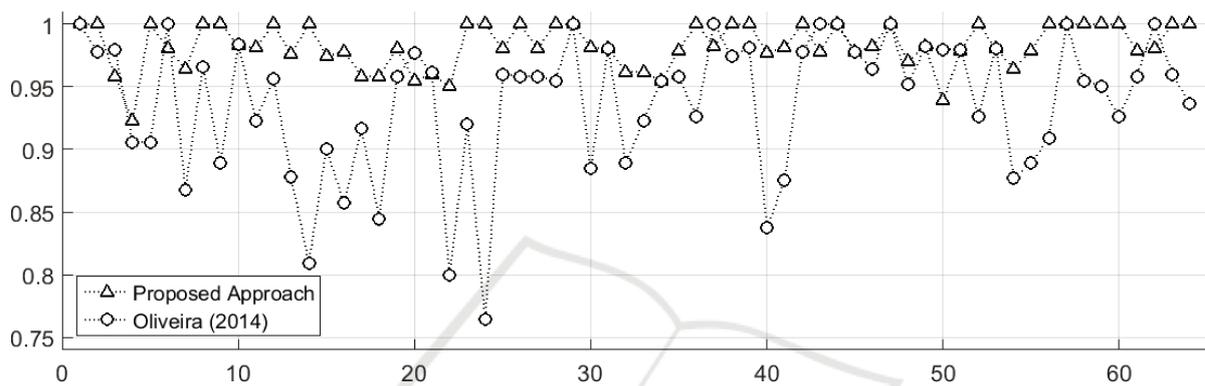


Figure 5: Comparison Analysis by F-Measure for 64 images: Detection/Counting.

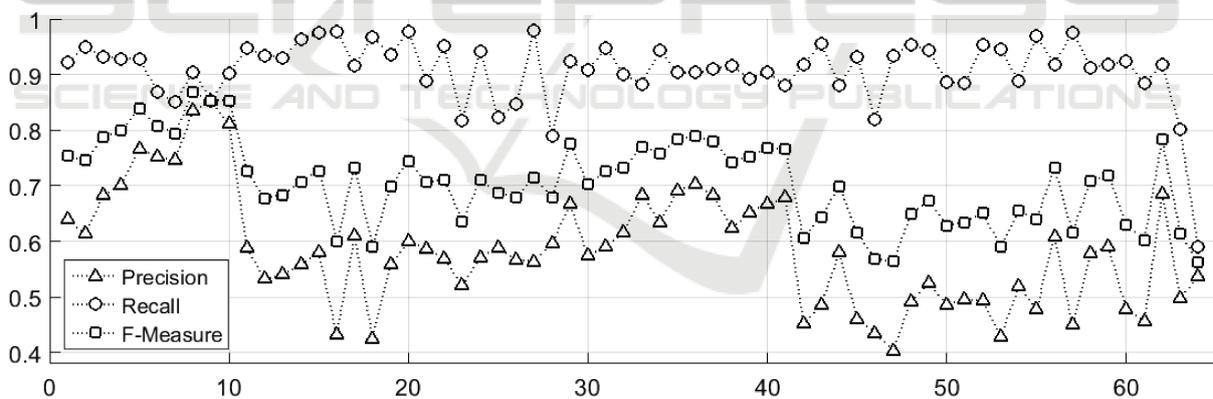


Figure 6: Performance evaluation of segmentation for 64 stomata images: Segmentation.

call values (84.37% and 90.62% respectively) larger than previous work and comparable in performance to manual, non-automatic counting. The second case indicated an F-Measure of 70 % in comparison to manual segmentation.

As future work, we will focus on a new approach to count the stomata that were split in the image-cropping process, in order to further improve precision. Deep Learning could be a future approach to be applied, depending on the improvement of our dataset, e.g. regarding a larger number of images. We also aim at image segmentation using Spectral

Graph Theory (SGT) and Normalized Cut with texture analysis and comparison with the proposed approach, as well as the computation of additional information about stomata such as area, aperture width, guard-cells width and length.

ACKNOWLEDGEMENTS

The authors would like to thank SCIANLab from the University of Chile for providing the images of vegetable tissue used in this work.

REFERENCES

- Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916.
- Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Connolly, C. and Fleiss, T. (1997). A study of efficiency and accuracy in the transformation from rgb to cielab color space. *IEEE Transactions on Image Processing*, 6(7):1046–1048.
- Hamamatsu (2016). Hamamatsu Nanozoomer XR microscope. <http://www.hamamatsu.com/eu/en/community/nanozoomer/index.html>. [Online; accessed 21-September-2016].
- Jian, S., Zhao, C., and Zhao, Y. (2011). Based on remote sensing processing technology estimating leaves stomatal density of populus euphratica. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 547–550.
- Laga, H., Shahinnia, F., and Fleury, D. (2014). Image-based plant stomata phenotyping. In *Control Automation Robotics Vision (ICARCV), 2014 13th International Conference on*, pages 217–222.
- Lojk, J., Sajn, L., Aibej, U., and Pavlin, M. (2014). Automatic cell counter for cell viability estimation. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on*, pages 239–244.
- Mallat, S. G. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.
- McGuinness, K. and O’Connor, N. E. (2010). A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition*, 43(2):434–444.
- Meyer, F. (1994). Mathematical morphology and its applications to signal processing topographic distance and watershed lines. *Signal Processing*, 38(1):113 – 125.
- Oliveira, M. C. S., Silva, N. R., Casanova, D., Pinheiro, L. F. S., Kolb, R. M., and Bruno, O. M. (2014). Automatic counting of stomata in epidermis microscopic images. In *X Workshop de Visão Computacional*.
- Olivo-Marin, J.-C. (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989 – 1996.
- Stepka, K. (2013). *Image Analysis: 18th Scandinavian Conference, SCIA 2013, Espoo, Finland, June 17-20, 2013. Proceedings*, chapter Automated Cell Counting in Bürker Chamber, pages 236–245. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Venkatalakshmi, B. and Thilagavathi, K. (2013). Automatic red blood cell counting using hough transform. In *Information Communication Technologies (ICT), 2013 IEEE Conference on*, pages 267–271.
- Violet-Chabrand, S. and Brendel, O. (2014). Automatic measurement of stomatal density from microphotographs. *Trees*, 28(6):1859–1865.
- Willmer, C. and Fricker, M. (1996). *Stomata*. Topics in plant functional biology. Springer.

B Adding GLCM Texture Analysis to a Combined Watershed Transform and Graph Cut Model for Image Segmentation

Adding GLCM Texture Analysis to a Combined Watershed Transform and Graph Cut Model for Image Segmentation

Kauê T. N. Duarte, Marco A. G. de Carvalho, and Paulo S. Martins

University of Campinas - UNICAMP, School of Technology
R. Paschoal Marmo, 1888, 13484-332, Limeira-SP, Brazil
kaue.duarte@pos.ft.unicamp.br, {magic,paulo}@ft.unicamp.br

Abstract. Texture analysis is an important step in pattern recognition, image processing and computer vision systems. This work proposes an unsupervised approach to segment digital images combining the Watershed Transform and Normalized Cut in graphs (NCut) using texture information obtained from the Gray-Level Co-occurrence Matrix (GLCM). We corroborate the enhancement of image segmentation by means of the addition of texture analysis through several experiments carried out using the BSDS500 Berkeley dataset. For example, an improvement of 7% and 12% was found in relation to the Combined Watershed+NCut and Quadtree techniques, respectively. The overall performance of the proposed approach was indicated by the F-Measure through comparisons against other important segmentation methods.

Keywords: Gray-Level Co-occurrence Matrix · Normalized Cut · Watershed · Texture Analysis · Image Segmentation

1 Introduction

The complexity of obtaining regions of interest in an image depends both on the segmentation algorithm used in this task and on its intended level of sophistication, as well as on the level of detail in the image required by the application. Traditional approaches aim either at identifying abrupt changes in gray level transitions (as these changes usually characterize the existence of edges) or at grouping pixels according to their affinity or similarity.

Graph-based techniques are among the most widely used in image segmentation, mainly due to their flexibility in representing an image as well as the ease of manipulation and adaptation. In this sense, spectral graph theory using the so-called spectral cuts method has been exploited in a number of applications of image segmentation [4, 7, 13, 19]. On the other hand, in order to accurately define the region boundaries, some methods have been tested and combined with different techniques including morphological operations, the Watershed Transform, and different graph representations, which provided satisfactory results [5, 16, 17].

In this paper, we present an approach to image segmentation based on the combination of Watershed Transform [3], texture analysis via the GLCM technique, and the Normalized Cut on graph. Specifically, we apply the hierarchical Watershed [15] as a way of simplifying the initial image and, therefore, reduce the computational effort in the spectral analysis step. A similarity graph (or tree) is built based on the concept of adjacency between pixels or regions. In essence, the texture features extracted from the GLCM are used to weight the adjacency criteria between regions before the final segmentation using NCut. Finally, we apply the proposed approach to the BSDS500 Berkeley Dataset [1], which provides a set of 500 generic images for validation and comparison analysis.

The main contributions of our work can be summarized as follows: 1) to carry out GLCM texture analysis in a dataset widely used in the literature; 2) to combine different strategies in order to implement a framework suitable to image segmentation, and 3) to perform a quantitative comparison between relevant approaches using a benchmark and F-measure.

This work is organized as follows: In Section 2 we address related work. The proposed approach is presented in Section 3. The experiments and the evaluation of the method are shown in Section 4. Finally, our conclusions are discussed in Section 5.

2 Related Work

Malik *et al.* [13] proposed an image segmentation algorithm aiming at processing the occurrence of texture and brightness. The authors base their arguments upon the fact that generic images are composed of textured and untextured regions, and the need to identify both contours and textures. The contour is treated through the *Intervening Contour* framework, whereas texture is analyzed using *textons*. The segmentation process is carried out by a feature extraction step followed by the application of the Normalized Cut. In the first step, the contour information in the image is obtained through the *Orientation Energy (OE)* from elongated quadrature filter pairs, and texture is calculated by means of a histogram of *texton* densities [11]. Similarity is evaluated using the χ^2 test function. In the second step, the Normalized Cut procedure is applied to the similarity graph obtained in the previous step, thus resulting in the segmented image. The dataset used was the same as the one used in our work.

In addition to Malik *et al.* [13], Sun *et al.* [19] and Liu and Zeng [12] have used the concepts of *textons*, graph and the Normalized Cut to segment generic and mammography images, respectively. Although *textons* are a powerful abstraction to handle texture information and have been widely used, they lie outside the focus of our work.

Hill *et al.* [10] proposed an image segmentation algorithm using the Watershed Transform and texture gradients. The process is carried out through the Watershed followed by the Non-decimated Complex Wavelet Packet Transform (NDXWPT). The Watershed is not directly applied (within this context) to an

original image, but to the gradient image initially extracted. Upon completion of the initial segmentation, the NDXWPT is applied to extract the features. This is a more effective alternative to the Gabor filters, since it requires less computational power. The image segmentation results vary according to the size set for the minimum size of regions.

Callaghan and Bull [16] proposed a method to general images capable of processing textured and non-textured objects. A combination of techniques was applied to improve segmentation results, and the overall approach consisted of two steps: 1) refinement of the texture-gradient Watershed algorithm [10], and 2) construction of a graph representation using primitive regions, which is required by the spectral partitioning algorithm. The combination of techniques led to an improvement of the results in comparison to the work by Hill *et al.* [10].

Our work combines elements of previous developments from Pinto *et al.* [17] and Zhu *et al.* [6]: 1) The work by Pinto *et al.* [17] presented a method to segment generic images, and the same dataset was used in our work to validate the approach. The method employs the Hierarchical Watershed to reduce the problem of over-segmentation (caused by the Watershed) and the computational time. A similarity graph is created from the regions and each region corresponds to a node and the similarity of the regions is mapped onto the edges. The affinity among regions is generated by a similarity function, and the final regions are then created by the Normalized Cut. 2) Zhu *et al.* [6] introduced an image segmentation method using the Gray-Level Co-occurrence Matrix (GLCM) [9] and Normalized Cut in a graph, including a data set composed of images from ultrasonography. The image is initially associated to a graph, where each pixel corresponds to a node and each similarity among pixels is related to an edge. The co-occurrence matrix is calculated using the GLCM algorithm from windows of size 3. The authors utilize homogeneity, contrast and energy as texture descriptors. Similarity is determined through a change in the Gaussian function to take into account the texture features. The Normalized Cut is applied as a last step to partition the graph and generate the final segmentation. The method was validated by comparing the results with manual segmentation produced by an specialist.

The technique described in our paper adds to the work of Pinto *et al.* [17] texture feature extraction using the GLCM along with an extensive analysis of the results within the context of the Watershed Transform and the Normalized Cut. Such combination of techniques, which includes the Watershed, has not been implemented by Zhu *et al.* [6] and it leads to a better identification of regions, as discussed in Section 4.

3 Proposed Approach

Our approach consists of four main steps: 1) Initial Segmentation; 2) Similarity Graph Computation; 3) Feature Extraction, and 4) Final Segmentation (Fig. 1). In the first step, hierarchical Watershed Transform reduces the image to a specified number of regions. A similarity graph is computed in the second step,

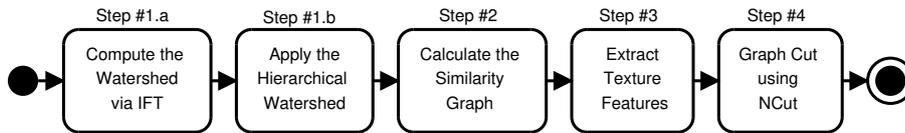


Fig. 1. Segmentation Model

based on the adjacency between regions obtained from previous step. Next, the texture features are computed and associated to the edges of the similarity graph. The spectral decomposition is then carried out using a graph cut, and the result is mapped back onto the original image. The details about each step are presented as follows.

3.1 Initial Segmentation

The initial segmentation in this work is based on the Watershed Transform [3]. The Watershed associates an image to a topographic map, where the brightness of each point represents its height. Through a process called immersion, the topography is flooded, starting through seed pixels (i.e. regional or local minima) until the regions are defined. Consequently, the process of flooding the watersheds until they merge generates the so-called watershed lines. The regions are defined by these lines.

The Watershed Transform may be straightforwardly applied to the gray levels of an image. Nevertheless, the data obtained in this way may present noise or incorrect information. Therefore, the gradient of the image (by means of the morphologic gradient) is a common preprocessing step applied to counteract these undesirable effects. The Watershed Transform may be implemented through several means, and the one being used in this work is the Image Foresting Transform (IFT) [8].

The Hierarchical Watershed [15] is then applied to limit the number of regions in the image, thus reducing the problem of over-segmentation as well as the computational effort. This technique creates a set of labeled hierarchical clusters. This labeled image is an image where each pixel is labeled with the value that uniquely identifies the region. In our case, each region is labeled with whole numbers which range from 1 up to the number of regions set by the Hierarchical Watershed algorithm.

3.2 Similarity Graph

After the generation of regions from the Watershed Transform, we can build a graph using the adjacency relationship. We assume $G = (V, E, W)$ as a undirected graph where V is a set of nodes, E is a set of edges and W a set of weights, given two adjacent nodes i, j and an edge $e(i, j) \in E$. In our approach, the set V represents the regions found, i.e. each graph node is associated to a Watershed region. The neighborhood between regions defines the E set. Lastly, the

edge weights W are defined according to some similarity criteria among regions (nodes), such as region gray level (RG) average, and calculated by Equation (1):

$$W_{ij}^{RG} = \exp\left(\frac{-\|x_i - x_j\|}{\delta^2}\right) \quad (1)$$

where x_i and x_j are, respectively, the similarity criteria between regions i and j .

Other region features could be included and combined altogether as edges weight in the similarity graph. In our case, we use texture descriptors (TX) given by Equation 2:

$$W_{ij}^{TX} = \exp\left(-\left(\sum_{d=1}^T \frac{\|\tilde{D}_i - \tilde{D}_j\|}{\tilde{D}_i + \tilde{D}_j}\right)\right) \quad (2)$$

where T represents the number of descriptors, and \tilde{D}_i, \tilde{D}_j represent one particular descriptor applied to regions i and j , respectively. Both texture components and the combination method between edge weights can be obtained from Equation (3):

$$W_{ij} = W_{ij}^{RG} \times W_{ij}^{TX} \quad (3)$$

We now turn our attention to feature extraction.

3.3 Feature Extraction

This step is concerned with the extraction of the texture feature using the Gray-Level Co-occurrence Matrix method (GLCM) [9]. The Co-occurrence Matrix stores the frequency of a pixel of value m to a pixel value n . This matrix has a dimension $N \times N$, where N is the maximum value of the pixel values (gray scale values, or colors) in the image. Each value of the index of the matrix corresponds to the gray scale value of the image. Therefore, the size of the image under analysis is not considered to be a discriminating factor. The occurrence of transitions between pixels in the Co-occurrence Matrix are generated from the definition of some parameters: 1) θ angle; and 2) distance d , i.e. each matrix accounts for one angle and one distance - and the number of matrices generated is defined by the number of angles \times number of distances. The angles correspond to the orientation from m to n that the matrix stores the occurrence. In Haralick [9], four angles are used for each distance d . The distance d is related to the interval (i.e. difference between indexes) from pixel m to pixel n .

An essential concept in calculating texture features is the definition of the appropriate size (in # pixels) of the (texture) window, necessary to the analysis of the pixel neighborhood for a given criteria. In our case, as we work with image regions, each region is considered a window which generates a set of irregular windows. In this work, for each and all regions r , the GLCM is calculated as follows: if the pixel is not labeled as r , then it is not considered in

the co-occurrence analysis. In general, each region has its own GLCM, and if an occurrence from m to n exceeds the region's boundary, then the occurrence of this pixel is disregarded.

In Haralick [9], fourteen types of texture descriptors were proposed, which address several aspects of texture (e.g. correlation, entropy, contrast) using the co-occurrence matrix. However, according to Baraldi and Parmiggiani [2], only a few descriptors are of major relevance. In our work, we use three descriptors:

- *Contrast (C)*: reflects the sharpness of images and the depth of texture grooves. Deeper texture grooves are associated with high contrast and better visual sharpness; on the contrary, low contrast leads to shallow grooves and blurred images. Higher number of pixels with high difference in gray scale (e.g., contrast profile) is associated with higher values of contrast;
- *Energy (E)*: The energy measures the local uniformity of the gray levels. It is derived from the Angular Second Moment (ASM). If pixels are very similar, the ASM value is large;
- *Homogeneity (H)*: it measures how close the distribution of elements in the GLCM is to the diagonal of GLCM. The contrast typically decreases as homogeneity increases.

These descriptors are defined as follows:

$$C = \sum_i \sum_j (i,j)^2 P(i,j), \quad E = \sum_i \sum_j (P(i,j))^2, \quad H = \sum_i \sum_j \frac{P(i,j)}{1+|i-j|} \quad (4)$$

where i,j represent the row and column position, respectively, P represents the gray occurrences, and the descriptors are summed for all four angles ($0^\circ, 45^\circ, 90^\circ, 135^\circ$).

3.4 Final Segmentation

In this last step, the graph cut technique was used in order to divide a graph G in two subgraphs S_1 and S_2 . Specifically, we apply the Normalized Cut approach (NCut) [18] in order to partition the similarity graph and therefore, to generate the regions that compose the resulting image. It takes advantage of the algebraic properties of the Laplacian Matrix to separate nodes according to their dissimilarity levels. In this approach, the NCut equation is given by:

$$NCut(S_1, S_2) = \frac{Cut(S_1, S_2)}{SumCon(S_1, V)} + \frac{Cut(S_1, S_2)}{SumCon(S_2, V)} \quad (5)$$

where $Cut(S_1, S_2)$ is the degree of dissimilarity between S_1 and S_2 , $SumCon(S_1, V)$ is the total weight of the edges connecting nodes from a subgraph S_1 to all nodes in the original graph V , and $SumCon(S_2, V)$ is similarly defined for a subgraph S_2 .

The optimal NCut is the one that minimizes Equation (5). This equation was extended leading to a well-known equation in linear algebra called the Rayleigh

Quotient. In this case, it can be minimized using spectral graph properties of the graph’s Laplacian Matrix [18]. The number of final regions are arbitrarily defined, but taking into account the desired level of detail that needs to be obtained in the segmentation.

4 Experiments and Results

In this section, we first introduce the dataset, the metrics, methods and the notation used for evaluation, and then we present a set of experiments that evaluate the efficiency and effectiveness of the proposed approach.

Dataset. All experiments were built upon the Berkeley *BSDS500 Dataset*, which is composed of 500 general-purpose images, where 200 are dedicated to training, 100 are used for validation and 200 are test images [18]. As the proposed approach does not require any training, the training subset of the dataset was used in the testing phase. Fig. 2 shows a collection of three sample images.

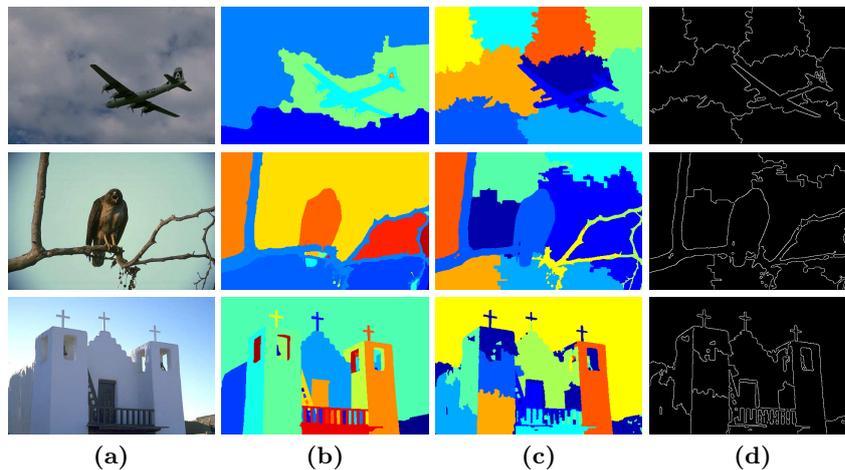


Fig. 2. Image segmentation with the proposed approach: (a) Original Image; (b) Ground-truth (3); (c) segmented image; (d) Border of (c)

Evaluation. The results were evaluated using the F-measure metric, which is calculated in relation to the ground truth provided by the Berkeley Segmentation Dataset and Benchmark [14]. The F-measure is the harmonic average between Recall and Precision values [1]. Through the F-Measure, the evaluation ceases to be merely visual and turns quantitative - as the values defined represent the quality of segmentation regarding the ground-truth set by the benchmark.

Methods used for Comparison. The following methods were used in this work to compare and assess the proposed approach: 1) *Pixel Grid*: This approach applies the Normalized Cut method directly to pixels to delimit the

final regions [18]. Each pixel is initially defined as a node and each edge represents the similarity between them; 2) *Multiscale* spectral image segmentation, which is also based on the Pixel Grid [7] similarity graph. However, this model generates a graph for different scales before applying the Normalized Cut; 3) *Quadtree*: In Carvalho *et al.* [5], it is crucial to adopt an algorithm that uses hierarchical structures in the generation of the similarity graph. Therefore, Carvalho *et al.* [5] proposed an algorithm with two major steps: a) the application of the Quadtree to represent the image in several scales before generation of the similarity graph, where each region corresponds to a node and the edges their similarity; and b) use of the Normalized Cut to create the final regions; 4) *Component tree*: Carvalho *et al.* [4] propose an algorithm that resorts to Component Trees to group nodes (thus creating a hierarchical graph) before the generation of the similarity graph, which is then ensued by the Normalized Cut; 5) *Combined Watershed+NCut*: This is the previous work by Pinto *et al.* [17] upon which we introduce texture information using the GLCM technique.

Basic notation. In order to evaluate the quality of the segmentation results using texture information, we carried out a systematic test case. Therefore, a number of parameters that directly affect the results were permuted, which generated 192 test cases or variations which were applied to the 500 images of the BSDS500. The results were stored in a set of folders that may be referred to using the $A_B_C_D_E_F$ notation, where: A = # regions in the Hierarchical Watershed; B = the GLCM distance; C = use of symmetry of direction; D = the value of δ in Equation 1; E = the final number of regions in the algorithm, and F = the GLCM descriptors used. The configuration of $A_B_C_D_E_F$ values used in the experiments were defined as follows: A{500,1000}, B{1,2}, C{Yes,No}, D{2,4,8}, E{10,30}, F{H,C,E,H+C+E}, and their combination led to 192 variants.

The next two subsections introduce a comparative analysis between the proposed approach and other methods described in the literature, including discussions about previous work by Pinto *et al.* [17].

4.1 Comparative Analysis with previous work by Pinto *et al.* [17] - Without GLCM feature

We denote previous work by Pinto *et al.* [17] as Watershed+NCut. Although we have defined two groups for the number of (final) regions (Ncut = {10,30}), for the sake of comparison in the following analysis we will emphasize Ncut = 10, which was the number of groups used by Pinto *et al.* [17]. The following comparisons were organized around the BSDS500 dataset, i.e., Training, Validation and Test sets. Table 1 illustrates a sample of the results, discussed as follows: 1) *Training set*: In the training phase, the Watershed+NCut was ranked in the 33th position. All the first 32 variations had the number of regions set to 1000. The addition of processing of texture lead to a more accurate segmentation in 97.5% of the cases. The best instance was (1000_1_Y_4_10_E) and it represented an improvement of 62% of the images tested in comparison to the work of Pinto [17]. However, it is conceivable that other variations have a better image-to-image

comparison under this criterion, e.g. the third best case - (1000_2_Y_4_10_H) - which represents 64% of the best segmented images regarding the work by Pinto *et al.* [17]. 2) *Validation set*: The Watershed+NCut was ranked 43. Notice that from 33 to 42 the number of regions was 500 and δ remained unchanged with value 8. The remaining (lower) ranks had $\delta = 2$. 3) *Test set*: In this set the Watershed+NCut was ranked 52. In 100% of the test images the use of a variant of the descriptor produced better results. When comparing the best case (1000_2_Y_4_10_H) with the Watershed+Ncut, the improvement was observed in 63% of the images.

Table 1. Comparative analysis between variants of the proposed approach with and without texture information

Phase	Method/Variation	F-measure	Ranking
Training	1000_1_Y_4_10_E	0,5056	1
	1000_1_Y_4_10_H	0,5055	2
	1000_2_Y_4_10_H	0,5054	3
	Watershed+Ncut	0,4882	33
Validation	1000_1_Y_8_10_H	0,5000	1
	1000_1_N_8_10_H	0,5000	2
	1000_1_Y_4_10_H	0,4988	3
	Watershed+Ncut	0,4852	43
Test	1000_2_Y_4_10_H	0,5110	1
	1000_2_N_4_10_H	0,5110	2
	1000_1_Y_4_10_H	0,5088	3
	Watershed+Ncut	0,4777	52

Fig. 3 illustrates the results obtained from the Test set. The Precision increases when the number of final regions is 10. However, for 30 regions, it is the Recall that reaches larger values. Notice that the image set under consideration is a generic one, as they vary in type (i.e. buildings, people, landscapes): if we adopt instead a specific dataset where all the images are of the same type (e.g. all buildings), this requires reconfiguration or specialization of the GLCM parameters for the particular types of images, and the F-measure performance in this case is (arguably) expected to improve.

4.2 Comparison with different methods

We focus our analysis in the comparison to the different approaches mentioned earlier in this section. Table 2 summarizes the comparison to other alternative methods. In all cases, the number of regions was set to 10. Only the validation and test sets were considered, since the proposed method had already been trained regarding its parameters.

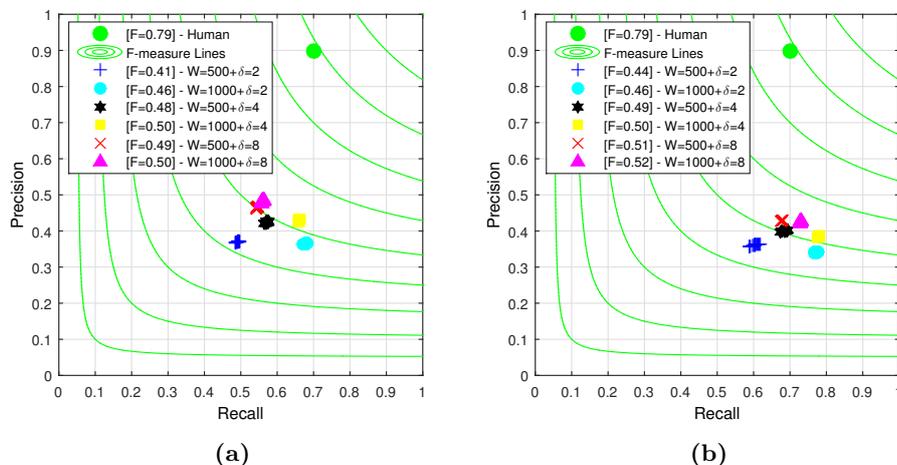


Fig. 3. Results evaluated by F-measure i: (a) With 10 regions; (b) With 30 regions

In the validation set, our approach outperformed the other methods and showed an improvement of 3% in relation to Watershed+NCut, which was ranked 44th. In the test set, our method (and variations) occupied the first 17 best positions, and the best result represented a 7% improvement over Watershed+NCut (ranked 54th). These results indicate that the inclusion of texture descriptors using the GLCM does improve the segmentation results, although it should not be taken as a final indicator of efficiency, but as an indication of the improvement over the original method (Watershed+NCut) that had no processing of texture.

4.3 Computational Effort

The experiments were carried out on a 2.50GHz i5-2450M, 4GB RAM. The approach was implemented through MATLAB r2015a. Two ideas were used in order to reduce processing time: 1) The Watershed, which is the most time-consuming task in the segmentation chain, was processed just once for the 500 images and the labeled matrices were stored on MATLAB structure (mat files) for a specific number of regions. All subsequent experiments started off from these structures instead of re-executing the Watershed. Thus, the runtime for each image was reduced to approximately 50%; 2) Since images are independent from each other, it was possible to use parallelism in the generation of results. The processing was split into four threads (on a 4-core processor, 1 thread per core), which resulted in a 75% reduction in segmentation time (i.e. processing of the dataset). Overall, the percent runtime for 500 images was reduced to 87.5%, taking around 6 hours instead of 46.

Table 2. Comparison Analysis (F-Measure) among alternative methods. Best results in bold face.

Method	Validation	Test
Component Tree	0,4997	0,5053
Multiscale	0,4756	0,4805
Quadtree	0,4476	0,4566
Pixel Grid	0,4610	0,4538
Watershed+Ncut	0,4852	0,4777
1000_2_Y_4_10_H	0,4983	0,5110
1000_2_N_4_10_H	0,4983	0,5110
1000_1_Y_8_10_H	0,5000	0,5064
1000_1_N_8_10_H	0,5000	0,5063

5 Summary and Conclusion

In this paper, we have presented an approach to texture processing in image segmentation based on a combination of the Watershed Transform, Gray-Level Co-occurrence Matrix (GLCM) and Graph Cut.

We built a similarity graph with nodes that correspond to the regions generated by the Watershed, and its edge weights were set by GLCM features. The Normalized Cut was applied to partition the graph and spectral analysis was used to optimize the cut criteria.

We showed that the use of GLCM in this approach does improve the results of image segmentation in comparison to a previous work by Pinto *et al* where texture was not used, in particular with respect to the F-measure. We also showed that the proposed combination has a better performance, using the same metric, when compared to the other four approaches.

The dataset employed was composed of generic images and the method was parameterized for such context. It is known that a given class of images may require specific parametrization to reach better segmentation results. The proposed method is amenable to further configuration in order to segment specific types of images such as buildings, biological, and others. In addition, although not the focus of this work, we were able to identify the processing bottlenecks and the computational effort was then substantially reduced in comparison to the earlier work by Pinto *et al*, what may in future work pave the way to the implementation of the model in embedded platforms.

In future work, we would like to incorporate other types of texture descriptors. Also, we wish to apply this approach to a specific application related to the medical field.

Acknowledgements

The authors thank FAPESP (São Paulo Research Agency) grant 2013/00575-3.

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE Trans. Pat. Anal. Mach. Int.* 33(5), 898–916 (May 2011)
2. Baraldi, A., Parmiggiani, F.: An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing* 33(2), 293–304 (Mar 1995)
3. Beucher, S., et al.: The watershed transformation applied to image segmentation. *Scanning Microscopy-Supplement* pp. 299–299 (1992)
4. Carvalho, M.A.G., d. Costa, A.L., Ferreira, A.C.B., Junior, R.M.C.: Image segmentation using component tree and normalized cut. In: 2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images. pp. 317–322 (Aug 2010)
5. Carvalho, M.A.G., Ferreira, A.C.B., Costa, A.L.: Image Segmentation Using Quadtree-Based Similarity Graph and Normalized Cut, pp. 329–337. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
6. Chang-ming, Z., Guo-chang, G., Hai-bo, L., Jing, S., Hualong, Y.: Segmentation of ultrasound image based on texture feature and graph cut. In: *Computer Science and Software Engineering International Conference*. vol. 1, pp. 795–798 (Dec 2008)
7. Cour, T., Benezit, F., Shi, J.: Spectral segmentation with multiscale graph decomposition. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2, pp. 1124–1131 vol. 2 (June 2005)
8. Falcão, A.X., Stolfi, J., de Alencar Lotufo, R.: The image foresting transform: Theory, algorithms, and applications. *IEEE Trans. Pattern Anal. Mach. Intel.* 26(1), 19–29 (Jan 2004)
9. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* (6), 610–621 (1973)
10. Hill, P.R., Canagarajah, C.N., Bull, D.R.: Image segmentation using a texture gradient based watershed transform. *IEEE Transactions on Image Processing* 12(12), 1618–1633 (Dec 2003)
11. Julesz, B.: Textons, the elements of texture perception, and their interactions. *Nature* 290 5802, 91–7 (1981)
12. Liu, L., Zeng, F.: Digital mammogram segmentation based on normalized cuts. In: *Global High Tech Congress on Electronics - IEEE*. pp. 127–130 (Nov 2012)
13. Malik, J., Belongie, S., Leung, T., Shi, J.: *Contour and Texture Analysis for Image Segmentation*, pp. 139–172. Springer US, Boston, MA (2000)
14. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proc. 8th Int'l Conf. Computer Vision*. pp. 416–423 (2001)
15. Meyer, F.: *The Dynamics of Minima and Contours*, pp. 329–336. Springer US, Boston, MA (1996)
16. O'Callaghan, R.J., Bull, D.R.: Combined morphological-spectral unsupervised image segmentation. *IEEE Transactions on Image Processing* 14(1), 49–62 (Jan 2005)
17. Pinto, T.W., de Carvalho, M.A.G., Pedronette, D.C.G., Martins, P.S.: Image segmentation through combined methods: Watershed transform, unsupervised distance learning and normalized cut. In: 2014 Southwest Symposium on Image Analysis and Interpretation. pp. 153–156 (April 2014)
18. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8), 888–905 (Aug 2000)
19. Sun, F., He, J.P.: A normalized cuts based image segmentation method. In: 2009 Second International Conference on Information and Computing Science. vol. 2, pp. 333–336 (May 2009)

Anexos

A Carta de Permissão para Inclusão do Apêndice A no Trabalho



Dear Kauê Tartarotti Nepomuceno Duarte,

I hereby authorize the inclusion of your paper, with title: “Segmenting High-quality Digital Images of Stomata using the Wavelet Spot Detection and the Watershed Transform” and published in the proceedings of VISAPP 2017, in your Msc thesis, as long as all references to where the paper has been published are explicit referred, namely: Proceedings Title, Editors, ISBN and the corresponding link to our Digital Library.

Setúbal, 5 of June, 2017

Best regards,

SCITEPRESS
Science and Technology Pub., Lda
Rua dos Lusíadas, 118 - 3º Frt.
1300-376 LISBOA
NIF: 509 223 192
Vitor Pedrosa