



Elaine Ayumi CHIBA

CORTE EM GRAFOS E SEGMENTAÇÃO DE IMAGENS
UTILIZANDO UM ALGORITMO AGLOMERATIVO DE
AGRUPAMENTO HIERÁRQUICO

LIMEIRA
2014



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE TECNOLOGIA/ FT

ELAINE AYUMI CHIBA

CORTE EM GRAFOS E SEGMENTAÇÃO DE IMAGENS
UTILIZANDO UM ALGORITMO AGLOMERATIVO DE
AGRUPAMENTO HIERÁRQUICO

Dissertação apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestra em Tecnologia, na área de Tecnologia e Inovação, na linha de pesquisa Engenharia da Informação.

Orientador: Prof. Dr. Marco Antonio Garcia de CARVALHO

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA DISSERTAÇÃO DEFENDIDA PELA ALUNA
ELAINE AYUMI CHIBA, E ORIENTADA PELO
PROF. DR. MARCO ANTONIO GARCIA DE CARVALHO.

LIMEIRA
2014

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Vanessa Evelyn Costa - CRB 8/8295

C431c Chiba, Elaine Ayumi, 1988-
Corte em grafos e segmentação de imagens utilizando um algoritmo
aglomerativo de agrupamento hierárquico / Elaine Ayumi Chiba. – Limeira, SP :
[s.n.], 2014.

Orientador: Marco Antonio Garcia de Carvalho.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de
Tecnologia.

1. Teoria dos grafos. 2. Segmentação de imagens. 3. Corte de grafos. I.
Carvalho, Marco Antonio Garcia de. II. Universidade Estadual de Campinas.
Faculdade de Tecnologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Graph cut and image segmentation using an hierarchical
agglomerative clustering algorithm

Palavras-chave em inglês:

Graph theory

Image segmentation

Cut graphs

Área de concentração: Tecnologia e Inovação

Titulação: Mestra em Tecnologia

Banca examinadora:

Marco Antonio Garcia de Carvalho [Orientador]

João Paulo Papa

Hélio Pedrini

Data de defesa: 21-02-2014

Programa de Pós-Graduação: Tecnologia

DISSERTAÇÃO DE MESTRADO EM TECNOLOGIA
ÁREA DE CONCENTRAÇÃO: TECNOLOGIA E INOVAÇÃO

Corte em grafos e segmentação de imagens utilizando um algoritmo aglomerativo de agrupamento hierárquico

Elaine Ayumi Chiba

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:

Marco Antônio C. de Carvalho

Prof. Dr. Marco Antonio Garcia de Carvalho
FT/UNICAMP
Presidente

João Paulo Papa

Prof. Dr. João Paulo Papa
UNESP

Hélio Pedrini

Prof. Dr. Hélio Pedrini
IC/UNICAMP

RESUMO

Representar os elementos de uma imagem em forma de grafos torna a estrutura organizada permitindo formular problemas de forma flexível e ser computacionalmente mais eficiente. Existem muitas técnicas da teoria de grafos sendo utilizadas em processamento digital de imagens. Em particular, o particionamento em grafos ou corte em grafos tem sido estudada por diversos autores como uma ferramenta de segmentação de imagens. Particionamento de um grafo refere-se à sua divisão em vários subgrafos tais que cada um deles representa um objeto de interesse na imagem. Neste trabalho, propomos um algoritmo de agrupamento hierárquico aglomerativo dos nós do grafo com base nas métricas de corte e corte médio. As segmentações foram avaliadas usando o *benchmark* da *Berkeley* BSDS500 que compara e classifica as segmentações em relação à outras técnicas existentes na literatura. Os resultados obtidos são promissores e nos permite concluir de que a combinação das métricas de corte e corte médio possibilitou melhores segmentações.

Palavras-chave: Grafos; Particionamento em grafos; Corte; Corte médio; Segmentação de Imagens e Agrupamento hierárquico

ABSTRACT

Representing the elements of an image in graphs makes the structure organized allowing to formulate problems in a flexible manner and can be more computationally efficient. There are many techniques of graph theory that are used in digital image processing. In particular, the graph partitioning or graph cut has been studied by several authors as a tool for image segmentation. Partitioning a graph refers to its division into several subgraphs such that each of them represents a meaningful object of interest in the image. In this work we propose a algorithm based on hierarchical agglomerative clustering of the graph nodes driven by the cut and mean cut criteria. The segmentation results were evaluated using the *benchmark* of *Berkeley* BSDS500 that compares and classifies the results in relation to other existing techniques in the literature. The results obtained are promising and allows us to conclude that the combination of the cut and mean cut criteria possible best segmentations.

Keywords: Graph; Graph partitioning; Cut; Mean cut; Image segmentation e Hierarchical clustering

Sumário

Resumo	vi
Agradecimentos	ix
Lista de figuras	xi
Lista de tabelas	xii
Lista de algoritmos	xiii
1 Introdução	1
1.1 Contribuições	4
1.2 Organização do texto	4
2 Fundamentação teórica e trabalhos relacionados	5
2.1 Modelo de imagens	5
2.1.1 Natureza da luz e das cores	7
2.1.2 Modelos de cores	8
2.2 Segmentação de imagens	11
2.3 Teoria dos grafos	12
2.4 Segmentação de imagens por meio de corte em grafos	13
2.4.1 Modelo de grafo de similaridade grade de pixels	14
2.4.2 Corte mínimo	14
2.4.3 Corte normalizado	14

2.4.4	Corte médio	17
2.5	<i>Benchmark</i> de segmentação de imagens da Berkeley	18
2.6	Considerações finais	19
3	Um algoritmo aglomerativo guiado pelos valores de corte e de corte médio	20
3.1	Visão geral sobre o funcionamento do algoritmo	20
3.2	Heurística	22
3.3	O algoritmo	23
3.4	Considerações finais	24
4	Experimentos e resultados	25
4.1	Teste com limiares t não adaptativos	25
4.2	Testes comparativos	28
4.2.1	Experimento 1: Diferença de cor	29
4.2.2	Experimento 2: Força da borda	30
4.2.3	Comparação dos modelos de similaridade	31
4.3	Considerações finais	32
5	Conclusão	34
5.1	Publicações	35
5.2	Trabalhos futuros	35

Agradecimentos

Agradeço primeiramente ao meu orientador Prof. Dr. Marco A. G. de Carvalho por me acompanhar nesta importante etapa da minha formação e por ter oferecido a oportunidade de trabalhar na área de processamento e análise de imagens. Agradeço pelo seu empenho e preocupações no decorrer da pesquisa.

Ao André Luís da Costa, meu companheiro e amigo, agradeço por estar sempre ao meu lado, mesmo nos momentos mais difíceis. Obrigada por me incentivar a buscar sempre mais conhecimento e a me ensinar a ter perseverança em alcançar os objetivos.

Agradeço à minha família, por me apoiarem nesta empreitada. Obrigada aos meus pais Hugo Chiba e Célia Seiko Kadoguchi Chiba, e à minha irmã, Hellen Tiemi Chiba, que sempre me incentivaram a concretizar este sonho e de terem depositado toda a confiança em mim sem ao menos entender sobre o que eu estudava.

À CAPES, agradeço por terem financiado e apoiarem a realização desta pesquisa. Por fim, agradeço a todos que me ajudaram e incentivaram a concluir o mestrado.

Lista de Figuras

2.1	Representação de uma imagem monocromática	6
2.2	Espectro eletromagnético	7
2.3	Espectro luminoso	8
2.4	Modelo RGB	9
3.1	Apresentação do processo de segmentação de imagens utilizando o algoritmo hierárquico aglomerativo proposto.	21
3.2	Segmentação produzida pela regra de corte médio e pela combinação das regras de corte médio e de corte. (b) segmentação utilizando a regra de corte médio até 30 regiões. (c) segmentação utilizando a regra de corte médio até 1836 regiões. (d) segmentação utilizando a combinação das regras de corte médio e de corte com $t = 1836$ com até 30 regiões.	23
4.1	Experimento para encontrar o limiar t apropriado para o algoritmo proposto.	26
4.2	Resultado da segmentação para uma pequena amostra de imagens. As imagens (a)-(d), (i)-(l), (q)-(u) são imagens originais do conjunto de teste e de validação do banco de imagens BSDS500. As imagens (e)-(h), (m)-(p) e (v)-(z) são respectivamente os resultados da segmentação usando o algoritmo proposto com $t = 1836$	27
4.3	Segmentação produzida pelos três algoritmos utilizados para comparação	30
4.4	Segmentação da imagem 201080 do conjunto de teste utilizando o grafo de similaridade por distância de cor	32
4.5	Segmentação da imagem 201080 do conjunto de teste utilizando o grafo de similaridade por força da borda	32

4.6	Segmentação da imagem 120003 do conjunto de teste para os dois algoritmos aglo- merativos	33
-----	--------------------------------------------------------------------------------------------------------	----

Lista de Tabelas

4.1	Pontuação geral do conjunto de teste e de validação e a pontuação individual para a pequena amostra de imagens.	28
4.2	Pontuação geral dos modelos nas imagens dos conjuntos de validação e de teste do <i>bechmark</i> BSDS500 utilizando o grafo de similaridade baseado na diferença de cor	29
4.3	Pontuação geral dos modelos nas imagens dos conjuntos de validação e de teste do <i>bechmark</i> BSDS500 utilizando o grafo de similaridade baseado na força da borda .	31
4.4	Pontuação da imagem 201080 do conjunto de teste do BSDS500 utilizando o grafos de similaridade baseado na diferença de cor e na força da borda	33

Lista de Algoritmos

- 3.1 Algoritmo para agrupamento de nós guiado pelas métricas de corte e corte médio . . . 24

Capítulo 1

Introdução

A segmentação de imagens é uma tarefa importante na área de visão computacional e em processamento digital de imagens. Em termos gerais, a segmentação visa subdividir uma imagem em suas partes e objetos constituintes para posterior análise (PEDRINI E SCHWARTZ, 2008). Desta forma, na etapa de segmentação é feita a extração e a identificação de objetos ou regiões de interesse contidos na imagem, os quais poderão ser descritos por meio de propriedades geométricas e topológicas que auxiliarão no processo de interpretação (GONZALEZ E WOODS, 2010; PEDRINI E SCHWARTZ, 2008). Há diversas técnicas de segmentação de imagens que vão desde as mais básicas, e. g., rotulação de uma imagem binária produzida por limiarização global (PEDRINI E SCHWARTZ, 2008), até algoritmos e procedimentos mais elaborados, e. g., segmentação por corte em grafos (SHI E MALIK, 2000; WENBING, 2008).

Embora as técnicas mais simples de segmentação possam resolver muitos problemas, há também muitos outros casos em que é necessária a utilização de técnicas mais sofisticadas. Há cerca de duas décadas, as técnicas de corte em grafos têm sido estudadas para se realizar a segmentação de imagens. Em uma descrição superficial do processo, primeiro é gerado um grafo de similaridade a partir da imagem como é descrito em maiores detalhes na Seção 2.4. Em seguida, tendo o grafo, que deve ser conexo, o objetivo é encontrar o melhor corte ¹ que particiona o grafo em dois subgrafos conexos. Por fim, a segmentação da imagem é definida em função do particionamento realizado no grafo.

Wu e Leahy (1993) propuseram que o melhor corte poderia ser obtido pela otimização de uma função objetivo que leva em consideração as arestas do corte. Desta forma, esta primeira proposta

¹Particionar um grafo conexo G em dois subgrafos A e $B \subset G$, também conexos, tal que $A \cup B \cup C_{AB} = G$, é um processo que remove um conjunto de arestas $C_{AB} = \{\{u, v\} | u \in A \text{ e } v \in B\}$ do grafo G . A este conjunto damos o nome corte.

de segmentação por corte em grafo busca minimizar o somatório dos pesos das arestas do corte para particionar o grafo original em dois subgrafos. Porém, como observado pelos próprios autores, esta função objetivo possui a desvantagem de favorecer ao desequilíbrio dos números de nós dos subgrafos, uma vez que os cortes com menor quantidade de arestas serão favorecidos. Posteriormente, novas funções objetivo foram propostas, sendo que neste trabalho foram exploradas o *corte normalizado* proposto por Shi e Malik (2000) e o *corte médio* proposto por Wang e Siskind (2001). As funções objetivo propostas visam solucionar um problema fundamental de agrupamento, que é particionar o conjunto de nós em conjuntos disjuntos V_1, V_2, \dots, V_m , onde, por alguma medida, a similaridade entre os nós de um mesmo conjunto V_i seja maior do que a similaridade entre os nós de conjuntos diferentes V_i, V_j (JAIN E DUBES, 1988; SHI E MALIK, 2000; SCHAEFFER, 2007).

O problema de corte ou particionamento do grafo vai além de apenas definir o critério a ser otimizado. Há também a questão de se definir um algoritmo eficiente que seja capaz de realizar a otimização. Neste sentido, nem sempre é viável otimizar a função objetivo, mas sim aproximar a solução ótima. É o caso, por exemplo, do corte normalizado, em que os autores propuseram uma aproximação da solução ótima baseada em um sistema de autovalores e autovetores. Maiores detalhes sobre o algoritmo proposto por Shi e Malik (2000) podem ser encontrados na Seção 2.4.3. Outro exemplo é a otimização do corte médio. Apesar de o algoritmo proposto por Wang e Siskind (2001) conseguir otimizar a função objetivo, sua aplicação limita-se a grafos planares.

Em um trabalho recente, Costa (2013) propôs uma nova abordagem para realizar a aproximação do valor ótimo de corte normalizado em um grafo de similaridade. Este algoritmo utiliza os valores de corte normalizado para guiar o agrupamento hierárquico dos nós do grafo. Algumas vantagens do algoritmo proposto por Costa (2013), denominado aglomerativo, em relação ao algoritmo proposto por Shi e Malik (2000), denominado espectral, são:

- o algoritmo aglomerativo garante que as partições serão conexas, uma vez que os agrupamentos são feitos de forma binária utilizando os dois nós conectados por uma aresta. Já o algoritmo espectral não garante que as partições sejam conexas simplesmente porque o processo de discretização dos autovetores não leva em consideração a vizinhança dos nós;
- testes práticos mostraram que em um cenário de bipartição de grafos com aproximadamente 150000 nós gerados a partir de imagens, o algoritmo aglomerativo foi capaz de encontrar cortes cujos valores de corte normalizado são em média 40 vezes menores que os obtidos pelo algoritmo espectral;
- com relação ao desempenho computacional, existe uma relação inversa de impacto em função do número de regiões desejadas. Enquanto que no caso do algoritmo espectral o custo

aumenta conforme aumenta a quantidade de regiões desejadas, no caso do algoritmo aglomerativo o custo diminui, pois serão necessárias menos operações de agrupamento.

O interessante da abordagem de segmentação por corte em grafos é que a segmentação passa a ser resolvida pela otimização de uma função objetivo no contexto de particionamento de grafos. Mas o quanto essa função objetivo é compatível com os predicados que seriam ideais para definir a segmentação desejada? Embora não seja o objetivo deste trabalho responder esta pergunta, ela está diretamente relacionada com os estudos e desenvolvimentos realizados.

Desta forma, o algoritmo aqui proposto não busca otimizar a função objetivo de corte, mas ao invés, utiliza seus valores para guiar um processo de agrupamento dos nós, que é feito de forma hierárquica, visando produzir uma segmentação adequada. Mais ainda, é proposta uma abordagem que utiliza duas funções objetivo, o *corte* e o *corte médio*, em diferentes etapas do processo de agrupamento. O algoritmo foi inspirado no trabalho de Costa (2013), possuindo grande semelhança estrutural com o algoritmo de agrupamento hierárquico desenvolvido para minimizar a função objetivo de corte normalizado. Contudo, o algoritmo proposto é muito diferente no que diz respeito à heurística e às métricas utilizadas para escolher a aresta cujos nós são agrupados em cada iteração.

A segmentação por corte em grafos é um assunto de interesse e que vem sendo estudado pelo grupo de pesquisa que atua no Laboratório de Computação Visual (IMAGElab) da Faculdade de Tecnologia da UNICAMP. Desta forma, foi possível aproveitar conhecimentos e recursos acumulados em projetos anteriores como os mestrados de Ferreira (2011) e de Costa (2013). No início a intenção deste projeto era verificar se outras funções objetivo de corte poderiam ser otimizadas pelo mesmo algoritmo proposto por Costa (2013). Porém, estudos e experimentos preliminares mostraram que o algoritmo da forma como foi proposto não foi capaz de produzir bons resultados. Veio então a ideia de modificar o algoritmo e adaptá-lo para o objetivo de produzir uma segmentação adequada ao invés de otimizar uma função objetivo. Assim, as segmentações passaram a representar melhor os objetos da imagem.

As qualidades das segmentações foram avaliadas por meio do *benchmark* BSDS500 da Berkeley (ARBELÁEZ *et al.*, 2011), o qual utiliza segmentações feitas por humanos como referência. As imagens do BSDS500 possuem tema natural com cenários rurais e urbanos. Os experimentos são conduzidos para o algoritmo proposto e também para os algoritmos espectral e aglomerativo que utilizam o corte normalizado como função objetivo. Os resultados são então avaliados e comparados segundo as métricas do *benchmark*.

1.1 Contribuições

A principal contribuição deste trabalho é de que a metodologia proposta se trata de uma metodologia alternativa e que não segue um padrão de soluções utilizadas para corte em grafos. Então, ao invés de otimizar uma função objetivo de corte, buscamos utilizar as funções sem o objetivo de otimizá-las. Constatamos que a utilização de diferentes funções objetivo para guiar o processo de agrupamento apresentavam características que poderiam ser combinadas para gerar segmentações mais equilibradas com relação ao tamanho das regiões. Com isso, o algoritmo proposto fundamenta-se no agrupamento hierárquico dos nós do grafo, respeitando-se um conjunto de heurísticas.

Outras contribuições são a determinação de um limiar não adaptativo que é descrito com maiores detalhes na Seção 4.1, a obtenção de bons resultados de segmentações utilizando uma vizinhança relativamente pequena, a vizinhança-8 aplicado num grafo denso, e também os resultados obtidos pelas métricas do *benchmark* para o algoritmo proposto apresentaram-se satisfatórios em detrimento aos outros algoritmos.

1.2 Organização do texto

Esta dissertação é constituída por 5 capítulos. O Capítulo 2 introduz os principais conceitos teóricos e apresenta-se a visão geral dos principais trabalhos relacionados. Já o Capítulo 3 descreve o novo algoritmo proposto, em que são mostradas a estrutura e a sua heurística. No Capítulo 4 apresentam-se os experimentos realizados e os respectivos resultados de segmentação obtidos pelo algoritmo proposto. Por fim, no Capítulo 5 são realizadas algumas considerações finais e discutidas as dificuldades encontradas. Também são apresentadas algumas sugestões para trabalhos futuros.

Capítulo 2

Fundamentação teórica e trabalhos relacionados

Neste capítulo são apresentados os trabalhos relacionados e os conceitos teóricos que foram necessários para o desenvolvimento deste trabalho de mestrado. A literatura em grafos é extensa e para obter maiores informações sobre o assunto, duas boas referências são dadas em Wilson e Watkins (1990) e Gibbons (1999).

2.1 Modelo de imagens

Uma imagem em níveis de cinza, pode ser definida como uma função bidimensional de intensidade de luz $f(x,y)$, onde x e y denotam as coordenadas espaciais e o valor de f em qualquer ponto (x,y) é proporcional ao brilho da imagem naquele ponto (GONZALEZ E WOODS, 2010; PEDRINI E SCHWARTZ, 2008). Para obtermos uma imagem digital devemos submetê-la a um processo de digitalização, que consiste em dois passos, a amostragem e quantização. A amostragem consiste em discretizar o domínio de definição nas direções x e y , gerando uma matriz $M \times N$. A quantização consiste em escolher um número inteiro L de níveis de cinza possíveis para cada ponto da imagem. Desta forma, uma imagem digital pode ser considerada como sendo uma matriz de amostras em que os índices das linhas e colunas denotam um ponto na imagem e o valor correspondente ao elemento da matriz identifica o nível de cinza (PEDRINI E SCHWARTZ, 2008).

Os elementos dessa matriz são denominados *pixels* (uma abreviação do termo em inglês *picture element*). A Figura 2.1 apresenta uma imagem digital e a convenção para os eixos x e y , representada por uma matriz de números inteiros (Figura 2.1(b)) que correspondem aos níveis

de cinza dos *pixels* da Figura 2.1(a), cuja dimensão é de M *pixels* na horizontal por N *pixels* na vertical.

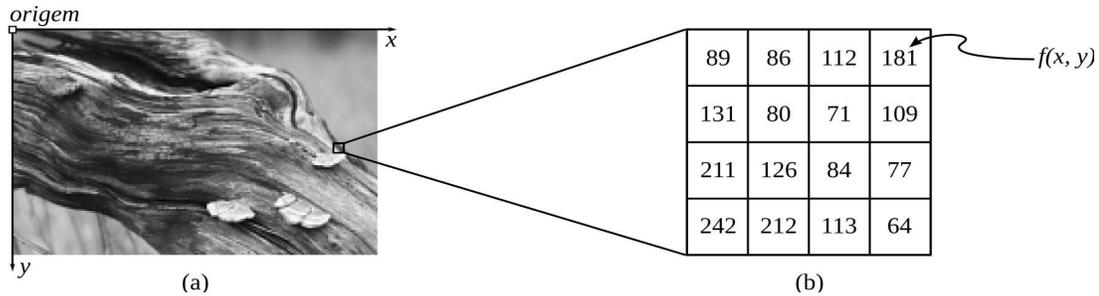


Figura 2.1: Representação de uma imagem monocromática. (a) Convenção dos eixos para representação de imagens digitais; (b) Detalhe da imagem digital: matriz de valores inteiros que representam os tons de cinza; Adaptado de (GONZALEZ E WOODS, 2010).

Uma imagem em tons de cinza é mais comumente representada por uma matriz de números inteiros de 8 bits e, por convenção, $L_{min} = 0$ representa o preto e o $L_{max} = 255$ representa o branco (GONZALEZ E WOODS, 2010; PEDRINI E SCHWARTZ, 2008).

Uma imagem colorida é normalmente representada por múltiplas bandas, onde a cor em um ponto (x,y) pode ser definido por três grandezas (PEDRINI E SCHWARTZ, 2008):

1. Luminância: brilho, cujo a intensidade percebida de um objeto que emite luz
2. Matiz: comprimento de onda dominante (cor)
3. Saturação: grau de pureza do matiz, isto é, quanto de branco está misturado ao mesmo.

Quando o matiz e a saturação são tomados juntos, são chamados de crominância. Assim, a cor é caracterizada pelo brilho e pela crominância.

Nas imagens representadas por múltiplas bandas temos uma função $f(x,y)$ para cada banda, ou seja, é necessária uma sequência de imagens monocromáticas para formar uma imagem colorida, dependendo da aplicação. É o caso de imagens coloridas padrão RGB na qual possui 3 bandas (STONE, 2001). O padrão RGB é formada pela informação de cores primárias aditivas, como o vermelho (R, *red*), verde (G, *green*) e azul (B, *Blue*) . Na subseção 2.2.2 são apresentadas informações adicionais sobre modelos de cores.

2.1.1 Natureza da luz e das cores

O olho humano é uma estrutura extremamente complexa. Com ele conseguimos focalizar os objetos, controlar a quantidade de luz que entra e produzir uma imagem nítida de um objeto. Sob esse aspecto o olho humano pode ser comparado a uma câmera fotográfica, pois ambos possuem uma abertura para a passagem da luz, a lente (cristalino) e um anteparo (retina), onde a imagem é recebida e registrada. A retina é composta por células sensíveis a luz, os cones e os bastonetes. Essas células transformam a luz luminosa das imagens em impulsos elétricos que são transmitidos ao cérebro pelo nervo óptico. As informações de cores presentes nas imagens são propriedades muito importantes para os seres humanos e também para as máquinas que processam essas características. A possibilidade de identificar um objeto e interpretar um ambiente pode ser simplificada ao utilizar a informação de cor (STONE, 2001; KODAK, 2000).

A compreensão da natureza da luz é essencial no estudo das cores. A luz branca, como a luz do dia, por exemplo, é composta por todas as cores do arco-íris. O resultado de quando ela é dividida nessas cores, como por meio de um prisma, é denominado faixa visível do espectro, que corresponde, a uma pequena região do espectro eletromagnético com comprimentos de onda é capazes de sensibilizar o olho humano. A cor é uma manifestação perceptual da luz que, por sua vez, é um sinal eletromagnético (PEDRINI E SCHWARTZ, 2008; KODAK, 2000). A Figura 2.2 ilustra o espectro total de radiações eletromagnéticas, incluindo as ondas de rádio, microondas, os raios infravermelhos, os raios X, os raios gama e, em destaque, a faixa visível do espectro.

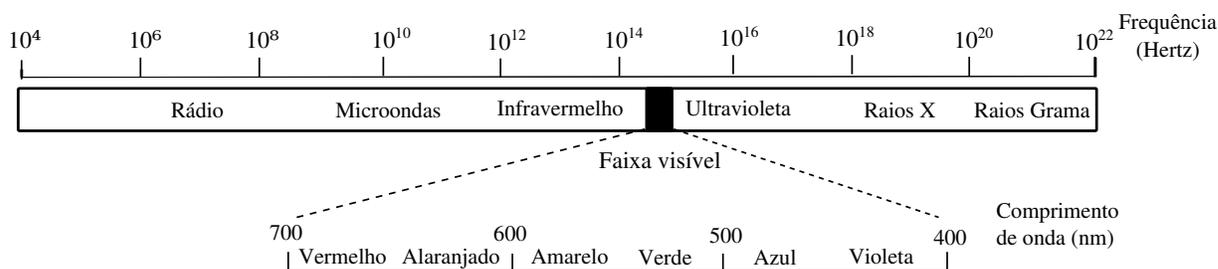


Figura 2.2: Espectro eletromagnético; adaptado de (PEDRINI E SCHWARTZ, 2008).

Um raio luminoso é constituído de partículas, chamadas fótons, cuja frequência e velocidade determinam seus comprimentos de onda, que são medidos em nanômetros (nm). Para cada cor existente há um comprimento de onda correspondente. Porém, para realizar a reprodução das cores é suficiente, na maioria das vezes, considerar a luz branca como uma mistura de apenas três componentes luminosos: as luzes vermelha, verde e azul (STONE, 2001; KODAK, 2000). A Figura 2.3 apresenta-se o espectro luminoso.

As cores vermelho, verde e azul são denominadas cores primárias aditivas, pois, a partir da

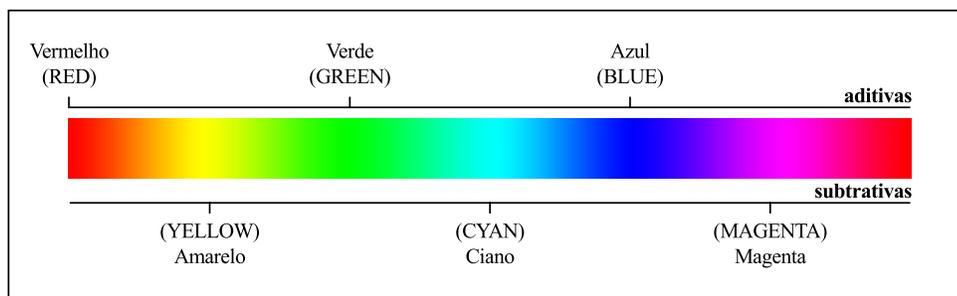


Figura 2.3: Espectro luminoso.

combinação entre as cores primárias formamos a luz branca. Quando combinamos as cores primárias entre si, obtemos os seguintes resultados: vermelho e verde temos a cor amarelo, vermelho e azul temos a cor magenta, azul e verde obtemos a cor ciano. Essas cores resultantes são cores secundárias da luz, mas podem ser chamadas como cores primárias subtrativas, pois, ao misturarmos as cores amarelo, cinza e magenta teremos a cor preta, ou seja, ausência de luz (PEDRINI E SCHWARTZ, 2008).

2.1.2 Modelos de cores

A Comissão Internacional de Iluminação (CIE, *Commission Internationale de l'Éclairage*) é o órgão responsável pela padronização na área de fotometria e colorimetria. A CIE estabeleceu vários padrões de cores. A seguir é apresentado alguns modelos utilizados em sistemas de informações gráficas (WU, 2009; PEDRINI E SCHWARTZ, 2008).

Modelo RGB

A CIE estabeleceu em 1931 de forma a padronizar o comprimento de onda das cores primárias: azul = $435.8nm$, verde = $546.1nm$, vermelho = $700nm$. Dado que $1\text{ nm} = 10^{-9}m$. O modelo RGB utiliza as cores primárias aditivas para representá-las em dispositivos como monitores e câmeras de vídeo. Este modelo se baseia no sistema de coordenadas cartesianas, onde as três cores primárias estão localizadas em três vértices de um cubo e as cores secundárias em outros três vértices deste mesmo cubo. O vértice junto à origem é o preto e o mais afastado da origem corresponde a cor branca, como pode ser visto na Figura 2.4. Por convenção os valores RGB estão normalizados entre 0 e 1 (SACHANDA, 2007; PEDRINI E SCHWARTZ, 2008).

A função de reconstrução das cores espectrais a partir das três cores primárias mostra que há uma faixa de cores espectrais que só podem ser reconstruídas com valores negativos de vermelho,

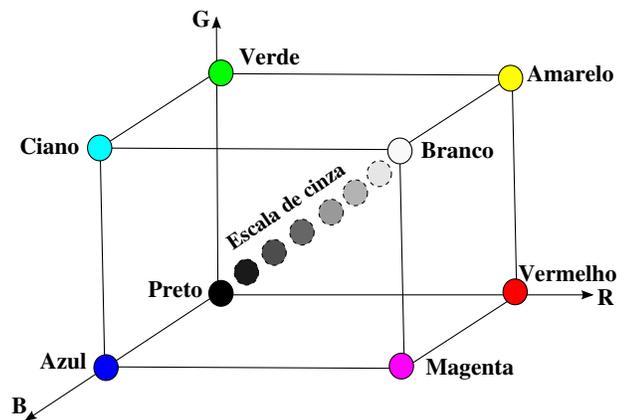


Figura 2.4: Modelo RGB; adaptado de (PEDRINI E SCHWARTZ, 2008).

ou seja, não é possível reproduzir todas as cores do espectro visível pela combinação ponderada de luzes vermelha, verde e azul. Isso foi uma das razões para a CIE estabelecer o padrão XYZ (WU, 2009).

Modelo XYZ

O modelo XYZ, foi estabelecido pela CIE para que qualquer cor espectral possa ser definida como um modelo de representação padrão de cores no qual qualquer cor espectral possa ser definida como uma combinação não negativa das cores primárias, que foi adotado cores primárias denominadas X, Y e Z. Embora não correspondam à cores espectrais, as suas componentes de cor são positivas sendo possível realizar o cálculo de obtenção dos valores de grandezas no modelo XYZ a partir de grandezas do sistema RGB. O padrão de cor XYZ pode ser transformado para o padrão RGB, através da seguinte transformação (WU, 2009; PEDRINI E SCHWARTZ, 2008):

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.3647 & -0.89658 & -0.468083 \\ -0.51515 & 1.42641 & -0.088746 \\ -0.005203 & -0.01441 & 1.00921 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.1)$$

A transformação inversa é dada por

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.010 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.2)$$

Os percentuais de cores primárias presentes em uma cor recebem o nome de coeficientes

tricromáticos, dados por

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad z = \frac{Z}{X+Y+Z} \quad (2.3)$$

A Equação 2.3 apresenta a quantidade de luz das cores primárias, normalizadas entre 0 e 1. A soma dos três coeficientes tricromáticos é $x + y + z = 1$. Desta forma o espaço de cores tridimensional é reduzido para um plano bidimensional. Assim, qualquer combinação aditiva entre duas cores sempre será localizada na reta que conecta estas duas cores no plano XY.

Modelo L*a*b*

O modelo de cor L*a*b foi desenvolvido pela CIE em 1976, a fim de simular a percepção do olho humano que possui resposta não linear às frequências do espectro visível. Este modelo é o que vem sendo mais utilizado convencionalmente para descrever todas as cores visíveis para o olho humano. O modelo L*a*b* é derivado do modelo XYZ. O componente L* corresponde à luminância, enquanto os componentes a* e b* correspondem à crominância vermelha/azul e amarela/azul, dados por (WU, 2009; PEDRINI E SCHWARTZ, 2008)

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, & \text{se } \frac{Y}{Y_n} > 0.008856 \\ 903.3\left(\frac{Y}{Y_n}\right), & \text{se } \frac{Y}{Y_n} \leq 0.008856 \end{cases} \quad (2.4)$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (2.5)$$

em que $f(t) = t^{\frac{1}{3}}$ se $t > 0.008856$ ou, caso contrário, $f(t) = 7.787t + 16/116$.

Calibração D65

Em 1931, a Comissão Internacional de Iluminação (CIE), criou padrões de iluminações denominadas iluminação A, B e C. Cada padrão de iluminação simula algum tipo de luz. A iluminação A representa uma luz doméstica proveniente de filamento de tungstênio. Iluminação B e C representam luz solar direta e com sombra, respectivamente. Mas, em 1964 a CIE passa a recomendar um novo padrão, a iluminação D. Esse padrão é definido na parte ultravioleta do espectro luminoso. Uma fase da luz do dia foi escolhida como a mais representativa e hoje é conhecido como o padrão de iluminação D65 da CIE. Esse padrão têm a intenção de representar a luz média do dia e possui uma temperatura de cor correlacionada a aproximadamente 6500K (SACHANDA, 2007).

2.2 Segmentação de imagens

O processamento digital de imagens possui um papel importante no cenário de desenvolvimento tecnológico atual. A necessidade de desenvolver sistemas autônomos que reproduzam a capacidade visual humana e que sejam capazes de reagir a estímulos visuais de forma adequada ainda é um grande desafio. Extrair informações que descrevem os objetos de interesse na imagem é um dos principais objetivos da visão computacional (GONZALEZ E WOODS, 2010; PEDRINI E SCHWARTZ, 2008).

A análise de imagens é um estudo da visão computacional que visa utilizar as informações baseadas nas formas, texturas, intensidades de níveis de cinza ou cores dos objetos presentes nas imagens. O objetivo da segmentação é o de facilitar esta análise, separando a imagem em regiões que sejam mais significativas e que possibilitem a utilização de algumas métricas que seriam inutilizáveis de outra forma.

Como resultado da segmentação de imagens, temos um conjunto de regiões ou um conjunto de contornos extraídos da imagem, onde os *pixels* em uma mesma região são considerados similares com referência a alguma característica ou propriedade tais como cor, intensidade, textura ou continuidade.

Neste trabalho, todos os algoritmos apresentados realizam a segmentação de uma região espacial R correspondente a toda região da imagem, em $n \in \mathbb{N}_{>1}$ subregiões $\{R_1, R_2, \dots, R_n\}$, de tal forma que:

1. $\cup_{i=1}^n R_i = R$, ou seja, cada pixel deve estar em ao menos uma subregião de forma que a união de todas as subregiões resulte na região correspondente à toda a imagem;
2. R_i é um conjunto conexo, ou seja, existem uma relação de adjacência entre os elementos de R_i de forma que para todos os pares possíveis de elementos possam estar conectadas por um caminho formado por uma sequência de elementos adjacentes. $\forall i$;
3. $R_i \cap R_j = \emptyset$, cada pixel deve pertencer a uma única subregião, $\forall i$ e $\forall j/i \neq j$;
4. $P(R_i) = VERDADEIRO$, $\forall i$;
5. $P(R_i \cup R_j) = FALSO$ para $i \neq j$ e tal que R_i e R_j sejam adjacentes.

Onde $P(R_i)$ são propriedades que os *pixels* de uma região devem cumprir, por exemplo, possuir a mesma intensidade (PEDRINI E SCHWARTZ, 2008).

Algumas aplicações da visão computacional que envolvem a segmentação de imagens se encontram, por exemplo, na identificação de pistas de pouso clandestinas através de análise de imagens de satélite, rastreamento de lançamento de mísseis, detecção de falhas num sistema de automação e robótica, detecção de patologias médicas (por exemplo, imagens geradas por raios X, tomografia computadorizada, ressonância e ultra-sonografia), identificação de impressões digitais e faces permitindo a distinção entre os indivíduos, reconhecimento de placas de automóveis, restauração de artefatos antigos, recuperação de imagens e vídeo por conteúdo, entre outros (PEDRINI E SCHWARTZ, 2008) (GONZALEZ E WOODS, 2010).

2.3 Teoria dos grafos

Um grafo G é uma estrutura de dados definida por $G = (V, E)$, sendo representada por um conjunto de nós $u \in V$, e por um conjunto de arestas $\{u, v\} \in E$, em que $u, v \in V$. Um subgrafo de um grafo G é o grafo H tal que $V(H) \subset V(G)$ e $E(H) \subset E(G)$ (WILSON E WATKINS, 1990). Dois nós u e v são vizinhos, denotado por $u \sim v$, se eles estão conectados por uma aresta, ou seja, $\exists \{u, v\} \in E$. Se todos os nós de um grafo forem mutualmente adjacentes, o grafo é dito completo. Arestas associadas a um mesmo nó constituem um laço ou *loop*.

Quando é necessário ser estabelecida uma orientação para as arestas, o grafo é denominado direcional ou dígrafo. O sentido da aresta é indicado através de uma seta. Neste caso, a aresta passa a ser denominada arco. Cada arco está associado a um par ordenado de nós $(u, v) \in E$.

Num grafo ponderado, um peso está associado a cada aresta, representado da forma $w_{\{u,v\}} \in W$. Neste caso, o grafo pode ser denotado por $G = (V, E, W)$, onde W representa o conjunto de pesos. O grau de um nó $Deg(u)$, em que $u \in V$, indica o número de arestas incidentes em u . Um grafo no qual todos os seus nós possuem o mesmo grau é dito grafo regular. A força de um nó $S(u) = \sum_{\{u \sim v\} \in E/u \sim v} w_{\{u,v\}}$ é o total de pesos das arestas incidentes em u . É possível definir também um quarto conjunto $A = \{a_u : \forall u \in V\}$, em que a_u é atributo do nó U (PENG *et al.*, 2013).

Um caminho em um grafo é uma sequência de nós não repetidos $u_1, u_2, \dots, u_{n-1}, u_n$, tal que nós subsequentes sejam adjacentes $u_n \sim u_{n-1}$. Um ciclo é denominado um caminho fechado que começa e termina no mesmo nó. Um grafo é dito conexo quando existe caminho entre todos os pares de nós possíveis; caso contrário é desconexo. Árvore é um grafo conexo sem ciclos. A hierarquia em árvores se faz presente quando se designa um nó raiz, e tal que exista uma relação lógica de ancestralidade entre os nós. A distância entre cada nó e a raiz de uma árvore é denominada de nível. Nós de grau 1 em uma árvore são chamados de folhas (GIBBONS, 1999; WILSON E WATKINS, 1990), ou seja, não possuem nós filhos, possuem relação somente com o nó ancestral.

Um grafo G é dito planar quando puder ser desenhado em um plano sem que ocorra cruzamento de arestas, ou seja, duas ou mais arestas não se intersectam geometricamente exceto nos nós em que são incidentes (GOLDBARG E GOLDBARG, 2012).

A representação gráfica de um grafo pode ser útil para analisar sua estrutura visualmente como um todo. Porém, a análise só é efetiva se o grafo for suficientemente pequeno. Para processamentos e análises computacionais é usual representar grafos por meio de matrizes. Algumas das formas mais utilizadas são:

1. **Matriz de Adjacência** $A_{n \times n}$, onde n é o número de nós do grafo, construída da seguinte forma:

$$[A]_{uv} = \begin{cases} 1 & \text{se } u \sim v \\ 0 & \text{caso contrário} \end{cases} . \quad (2.6)$$

Num grafo ponderado, a matriz de pesos é construída de maneira semelhante à matriz de adjacência, tal que

$$[W]_{uv} = \begin{cases} w_{\{u,v\}} & \text{se } u \sim v \\ 0 & \text{caso contrário} \end{cases} ; \quad (2.7)$$

1. **Matriz Laplaciana**, que é utilizada no algoritmo de corte normalizado pela teoria espectral de grafos (CVETKOVIĆ *et al.*, 2010). A matriz Laplaciana é formada pela subtração da matriz de graus pela matriz de adjacência ou pela matriz de pesos. Para grafos não ponderados, ela é definida como $L_{n \times n} = D_{n \times n} - A_{n \times n}$, enquanto para grafos ponderados como $L_{n \times n} = D_{n \times n} - W_{n \times n}$. A matriz $D_{n \times n}$ é diagonal, sendo definida como

$$[D]_{uv} = \begin{cases} \text{Deg}(u) & \text{se } u = v \\ 0 & \text{caso contrário} \end{cases} . \quad (2.8)$$

2.4 Segmentação de imagens por meio de corte em grafos

Particionar (ou cortar) um grafo se refere ao processo de dividir seu conjunto de nós de forma a satisfazer algum critério (WENBING, 2008). O corte em grafos vem sendo utilizado há cerca de duas décadas para segmentar imagens, sendo a imagem representada por um grafo ponderado conhecido como grafo de similaridade (SHI E MALIK, 2000). Neste grafo, os pesos das arestas representam a similaridade entre os nós por elas conectados, e valores maiores significam maior similaridade entre os nós.

2.4.1 Modelo de grafo de similaridade grade de pixels

Um modelo de grafo de similaridade é um conjunto de regras que devem ser seguidas para gerar um grafo a partir da imagem. Um modelo pode, ser definido sobre uma pré-segmentação, por exemplo *Watershed* (MA E WAN, 2008; CARVALHO *et al.*, 2009), ou sobre outras estruturas geradas a partir da imagem, como a *Quadtree* (CARVALHO *et al.*, 2010). O modelo mais simples, aqui denominado *grade de pixels*, gera um grafo de similaridade $G = (V, E, W)$ onde cada pixel da imagem é associado a um nó $u \in V$. Dois nós u e $v \in V$ são conectados por meio de uma aresta $u, v \in E$ se a distância entre seus *pixels* correspondentes não for maior que um dado raio de conexão r (SHI E MALIK, 2000; COUR *et al.*, 2005; CARVALHO *et al.*, 2010).

Os valores dos pesos das arestas devem refletir informações de similaridade entre os *pixels* associados aos nós conectados. Alguns critérios são: intensidade, cor, textura ou contornos. Assim, *pixels* similares têm maior chance de fazerem parte do mesmo objeto. Os pesos são calculados por meio de uma função de similaridade, a qual pode ser adaptada a um problema ou a um modelo específicos. As funções de similaridades utilizadas neste trabalho são definidas no Capítulo 4.

2.4.2 Corte mínimo

A ideia do corte mínimo foi proposta por Wu e Leahy (1993). Essa técnica consiste em obter-se um subgrafo $A \subset G$, onde $A = (V_a, E_a, W_a)$ e $G = (V, E, W)$, de forma que seja mínimo o valor do corte, dado por

$$\text{cut}(A) = \sum_{u \in V_a, v \in V \setminus V_a} w_{\{u,v\}}. \quad (2.9)$$

O corte mínimo em um grafo pode ser encontrado eficientemente por meio de uma busca recursiva. Porém, Wu e Leahy (1993) notaram que o critério de corte mínimo favorece ao desbalanceamento das ordens¹ dos subgrafos gerados no particionamento. Observe que, pela Equação 2.9, há uma tendência ao aumento do valor do corte conforme cresce a quantidade de arestas removidas. Desta forma, é comum que o subgrafo A cujo valor de corte é mínimo seja muito pequeno, possivelmente formado por um único nó.

2.4.3 Corte normalizado

Shi e Malik (2000) propuseram a métrica de *Corte Normalizado* (NCut - do inglês *Normalized Cut*)

¹A *ordem* de um grafo se refere à cardinalidade do conjunto de nós, *i. e.*, é a quantidade de nós que o grafo possui. O *tamanho* do grafo se refere à cardinalidade do conjunto de arestas.

como uma forma de reduzir o problema encontrado por Wu e Leahy (1993). Assim, a métrica de corte normalizado tem como objetivo produzir particionamentos melhor balanceados. O valor de corte normalizado para biparticionar um grafo G em dois subgrafos A e B é dado por

$$\text{NCut}(A, B) = \frac{\text{cut}(A, B)}{\text{Assoc}(A, G)} + \frac{\text{cut}(A, B)}{\text{Assoc}(B, G)}, \quad (2.10)$$

onde $\text{Assoc}(A, G) = \sum_{u \in A, v \in G} w(u, v)$.

O corte ótimo de um grafo de similaridade G em dois subgrafos A e B é aquele cujo valor $\text{NCut}(A, B)$ é mínimo (SHI E MALIK, 2000). Entretanto, encontrar tal particionamento é um problema de complexidade NP-Completo.

Aproximando o valor ótimo de corte normalizado

Embora não seja conhecido um algoritmo capaz de encontrar um particionamento de um grafo G em dois subgrafos A e B tal que o $\text{NCut}(A, B)$ seja mínimo, Shi e Malik (2000) descobriram que, ao expandirem a Equação 2.10 é possível encontrar um particionamento cujo valor do corte normalizado seja próximo do mínimo. Para isso, a solução dada por Shi e Malik (2000) foi explorar as propriedades do espectro do grafo, como descritas por Fiedler (1975).

O estudo da teoria espectral de grafos (TEG) (CVETKOVIĆ *et al.*, 2010) aborda as propriedades algébricas do espectro de certas matrizes associadas a um grafo. O espectro de um grafo G é o conjunto de seus autovalores, geralmente apresentados em ordem decrescente, associados às suas respectivas multiplicidades algébricas (ABREU, 2005).

Seja A uma matriz quadrada de ordem n . Um autovalor de A é um escalar λ tal que $AX = \lambda X$, para $X \neq 0$, $X \in R^n$. Todo X que satisfaça esta equação é denominado um autovetor de A associado a um autovalor de λ (CVETKOVIĆ *et al.*, 2010).

As etapas do algoritmo de segmentação por corte normalizado fundamentado na TEG, são (SHI E MALIK, 2000):

1. Dada uma imagem, cria-se o grafo de similaridade $G = (V, E, W)$ utilizando o modelo desejado, por exemplo, o modelo de grade de *pixels*.
2. Construir a matriz de pesos W e a matriz diagonal de graus D a partir do grafo de similaridade. Com isso, obtém-se a matriz Laplaciana L , onde $L(G) = D(G) - W(G)$.
3. Resolver a expressão $(D - W)x = \alpha Dx$.

4. Utilizar os autovetores obtidos para particionar o grafo. Usualmente isto é feito por meio da discretização dos autovetores.

Resolvendo a Equação $LX = \alpha Dx$, obtemos os autovalores da matriz Laplaciana L , onde, α são os autovalores e X são os autovetores associados. Sendo assim, pela análise dos autovetores é possível determinar a partição em que cada nó do grafo de similaridade irá pertencer. Finalmente, basta fazer o mapeamento do grafo para a imagem para obter as partições (SHI E MALIK, 2000; CARVALHO E COSTA, 2012).

O algoritmo proposto por Shi e Malik (2000) pode ser aplicado em dois modos de operação:

1. *2-way*, em que cada região obtida, iniciando-se pela região R definida por toda a região da imagem, é recursivamente particionada em duas novas regiões, pela análise do autovetor associado ao segundo menor autovalor;
2. *k-way*, em que o particionamento é realizado diretamente em um dado número k de regiões previamente especificado. Nesse modo é necessário o cálculo de k autovetores a partir da matriz que representa o grafo, sendo que as regiões são determinadas pela análise de todos os k autovetores calculados.

Em todos os experimentos neste trabalho que envolveram a aplicação do algoritmo de corte normalizado baseado na TEG, o modo de operação utilizado foi o *k-way*.

Um algoritmo aglomerativo para aproximar o valor ótimo de corte normalizado

Costa (2013) propôs em seu mestrado um algoritmo aglomerativo de agrupamento hierárquico dos nós do grafo para aproximar a minimização da função objetivo de corte normalizado, ao invés de utilizar a teoria espectral. Nos experimentos realizados pelo autor o biparticionamento de um grafo pelo novo algoritmo apresenta valor de corte normalizado na média cerca de 40 vezes menor que o biparticionamento pelo algoritmo baseado na teoria espectral.

O algoritmo aglomerativo funciona de forma similar a um algoritmo de ligação (*linkage*) (GOWER E ROSS, 1969; SCHAEFFER, 2007), onde a cada iteração os nós conectados por uma aresta válida escolhida são unidos (WU, 2012). As maiores diferenças entre os dois algoritmos são:

- nos algoritmos de ligação as arestas são escolhida pela ordenação dada pelos pesos associados a elas, enquanto que no algoritmo aglomerativo de corte normalizado escolhe-se o

primeiro nó de acordo com um critério baseado em seu valor de corte normalizado ². A aresta é então definida pela escolha de um segundo nó, que deve ser adjacente ao primeiro, utilizando um outro critério também fundamentado na métrica de corte normalizado;

- quando dois nós são unidos no algoritmo proposto por Costa (2013) há uma combinação das arestas incidentes em ambos os nós de forma a manter a consistência dos valores de corte que novos agrupamentos produzirão. Nos algoritmos de ligação os pesos das arestas não se alteram.

A principal diferença do algoritmo proposto neste trabalho em relação ao algoritmo proposto por Costa (2013) está nos critérios utilizados para definir o primeiro e segundo nós em cada iteração. Maiores detalhes sobre o funcionamento dos dois algoritmos e sobre os critérios utilizados são dados no Capítulo 3.

2.4.4 Corte médio

A métrica de Corte médio (MCut – do inglês *Mean Cut*) foi proposta por Wang e Siskind (2001). Ela é considerada livre de tendência quanto ao balanceamento das partições geradas (WU, 2012). É também invariante às transformações lineares dos pesos das bordas e, desta forma, insensível ao contraste da imagem. O valor de corte médio de um subgrafo $A \subset G$, onde $A = (V_a, E_a, W_a)$ e $G = (V, E, W)$, é definido por

$$\text{meancut}(A) = \frac{\text{cut}(A)}{|\{\{u, v\} : u \in V_a, v \in V \setminus V_a\}|}. \quad (2.11)$$

Sob a óptica desta função objetivo, o melhor particionamento em um grafo de similaridade é aquele que apresenta o valor mínimo de corte médio. O desempenho da minimização é de tempo polinomial, mas a solução é limitada a grafos planares conexos (WANG E SISKIND, 2001; WU, 2012). O corte médio é uma especialização de uma definição mais geral de razão de cortes (*ratiocut*) (WANG E SISKIND, 2001; WU, 2012). Considerando que as arestas de um grafo possuem dois pesos $w_{\{u,v\}}^1$ e $w_{\{u,v\}}^2$ associados, o valor de ratiocut para um subgrafo $A \subset G$ é dado por

$$\text{ratiocut}(A) = \frac{\text{cut}(A|w^1)}{\text{cut}(A|w^2)}. \quad (2.12)$$

Nota-se que no caso do corte médio, $w_{\{u,v\}}^2 = 1, \forall \{u, v\} \in E$.

²Cada nó do grafo da iteração corrente representa um subgrafo do grafo original. Assim, é possível calcular o valor de qualquer métrica de corte para todos os nós do grafo.

2.5 *Benchmark* de segmentação de imagens da Berkeley

O *benchmark* de segmentação de imagens da Berkeley *BSDS500* (ARBELÁEZ *et al.*, 2011) é uma ferramenta desenvolvida para auxiliar na avaliação da qualidade dos resultados produzidos por um algoritmo de segmentação. Ele disponibiliza uma base de dados com 500 imagens naturais que abrangem cenários urbanos, rurais e selvagens. As imagens são coloridas e possuem tamanho padronizado de 481x321 *pixels*. Para cada uma das 500 imagens há também um conjunto de segmentações manuais realizadas por humanos, que são utilizadas como *ground truth* para medir a qualidade das segmentações produzidas automaticamente por meio de algoritmos. A base de dados é dividida em três conjuntos:

1. de *treinamento*, composto por 200 imagens;
2. de *validação*, composto por 100 imagens;
3. de *teste*, composto por 200 imagens.

Neste trabalho são utilizadas três métricas para calcular a distância entre as segmentações produzidas nos experimentos e as correspondentes do *ground truth*. São elas:

1. **Medida-F**. Essa métrica pontua a segmentação com base nas bordas das regiões. É expressa da forma

$$F = 2 \cdot \frac{P \cdot R}{P + R}, \quad (2.13)$$

onde $P = \frac{VP}{VP+FP}$ e $R = \frac{VP}{VP+FN}$ são respectivamente a *Precisão* e *Revocação*. VP é o número de verdadeiros positivos, *i. e.*, o número de *pixels* preditos como pertencentes a uma borda que de fato pertencem a uma. Analogamente, FP é o número de falsos positivos e FN é o número de falsos negativos.

2. **Cobertura de Região**. Essa métrica, assim como a seguinte, avalia a segmentação com base nas regiões, ao invés de utilizar a informação de borda. A cobertura de região é definida como

$$C(S' \rightarrow S) = \frac{1}{N} \sum_{M \in S} |M| \cdot \max_{M' \in S'} O(M, M'), \quad (2.14)$$

onde $O(M, M') = \frac{|M \cap M'|}{|M \cup M'|}$ é a sobreposição entre duas regiões M e M' ; S e S' são partições da imagem; e N é o número total de *pixels* na imagem.

3. **Probabilistic Rand Index**. Essa terceira métrica também é fundamentada na informação das

regiões. É definida como

$$PRI(S, \{G_k\}) = \frac{1}{T} \sum_{i < j} [c_{ij} p_{ij} + (1 - c_{ij})(1 - p_{ij})], \quad (2.15)$$

onde S é uma partição da imagem; $\{G_k\}$ é um conjunto de segmentações da base de comparação (*ground truth*); c_{ij} é o evento onde os *pixels* i e j possuem o mesmo rótulo e p_{ij} é a sua probabilidade; e T é o total de pares de *pixels*.

2.6 Considerações finais

Neste capítulo foram descritos os principais conceitos teóricos envolvidos neste trabalho, porém, de forma resumida e direcionada para o escopo deste trabalho. Aconselha-se recorrer às referências para uma melhor compreensão dos conceitos e da discussão envolvida em cada assunto sempre que necessário.

O assunto de corte em grafos já possui uma grande discussão e outras funções objetivo de corte já foram propostas na literatura. A omissão dessas funções neste trabalho não diminui em nada a relevância delas para a pesquisa. Uma boa revisão sobre elas é dada por Wu (2012). Mais ainda, corte em grafos é uma classe de soluções para o problema geral de particionamento de grafos que é um assunto mais amplo e intimamente ligado com o problema de agrupamento em grafos, generosamente detalhado em uma revisão feita por Schaeffer (2007). Com exceção do algoritmo espectral de corte normalizado, todos os algoritmos utilizados neste trabalho podem ser considerados de agrupamento hierárquico aglomerativo.

Capítulo 3

Um algoritmo aglomerativo guiado pelos valores de corte e de corte médio

O algoritmo proposto neste trabalho é similar ao algoritmo proposto por Costa (2013). As principais particularidades do novo algoritmo são: a heurística utilizada para determinar em cada nova iteração a aresta que define os nós que serão unidos; e a representação dos grafos em cada iteração, que possuem dois pesos associados a cada aresta, como detalhado na Seção 3.1 a seguir. Deste modo, este algoritmo também não faz uso de conceitos da teoria espectral de grafos e fundamenta-se em definir o particionamento por meio do agrupamento hierárquico dos nós do grafo dois a dois de forma iterativa. Isso garante que as partições produzidas serão conexas, uma vez que dois nós somente são agrupados se há uma aresta entre eles.

3.1 Visão geral sobre o funcionamento do algoritmo

Seja $G = (V, E, W)$ um grafo de similaridade qualquer, algumas vezes sendo referenciado no texto como grafo original; e seja $G^n = (V^n, E^n, W^n, D^n)$ um grafo utilizado na iteração $n \in \mathbb{N}^0$ que representa o grafo original, onde os conjuntos V^n, E^n , e W^n são definidos similarmente aos conjuntos V, E , e W de G , e D^n é um conjunto de graus $d_{\{u,v\}} \in D^n$ associados às arestas do grafo, *i. e.*, uma aresta de um grafo G^n pode estar representado várias arestas do grafo original, sendo que o grau de uma aresta nesta definição equivale ao número de arestas representadas. O conjunto de graus das arestas é utilizado no cálculo do corte médio para os nós $u \in V^n$, como é definido a seguir

$$\text{meancut}(u) = \frac{\text{cut}(u)}{\sum_{v \in V^n, v \neq u} d_{\{u,v\}}}, \quad (3.1)$$

onde $\text{cut}(u) = \sum_{u \in V^n, v \in V^n \setminus \{u\}} w_{\{u,v\}}$. Observe que um nó $u \in V^n$ corresponde a um subgrafo $A \subset G$ do grafo original e, desta forma, é possível calcular valores de corte para todos eles.

O algoritmo hierárquico aglomerativo proposto funciona de forma similar aos algoritmos clássicos de ligação (GOWER E ROSS, 1969; SIBSON, 1973), no sentido de que ambos visam definir a aresta $\{u, v\} \in E^n$ cujos nós são agrupados em cada iteração. No entanto, enquanto os algoritmos de ligação usam os pesos para definir tal aresta, o algoritmo proposto define a aresta indiretamente pela escolha de dois nós usando diferentes critérios.

A área tracejada no diagrama de blocos apresentado na Figura 3.1 descreve as etapas de cada iteração do algoritmo aglomerativo proposto. Pode-se notar que um primeiro nó $u \in V^n$ é definido e, somente depois, um segundo nó $v \in V^n | v \sim u$ é escolhido. O critério para a definição dos nós u e v são distintos. Uma outra característica do algoritmo proposto, diferente dos algoritmos clássicos de ligação, é que as arestas incidentes aos nós em processo de agrupamento são combinadas para manter a consistência dos valores de corte. Assim, seja $u' \in V^{n+1}$ o nó definido pela união dos nós u e $v \in V^n$, e seja $z \in V^n | z \sim u$ e $z \sim v$, $w_{\{u',z\}} = w_{\{u,z\}} + w_{\{v,z\}}$, onde $w_{\{u',z\}} \in W^{n+1}$; e $w_{\{u,z\}}$ e $w_{\{v,z\}} \in W^n$. A combinação dos graus das arestas é feita similarmente.

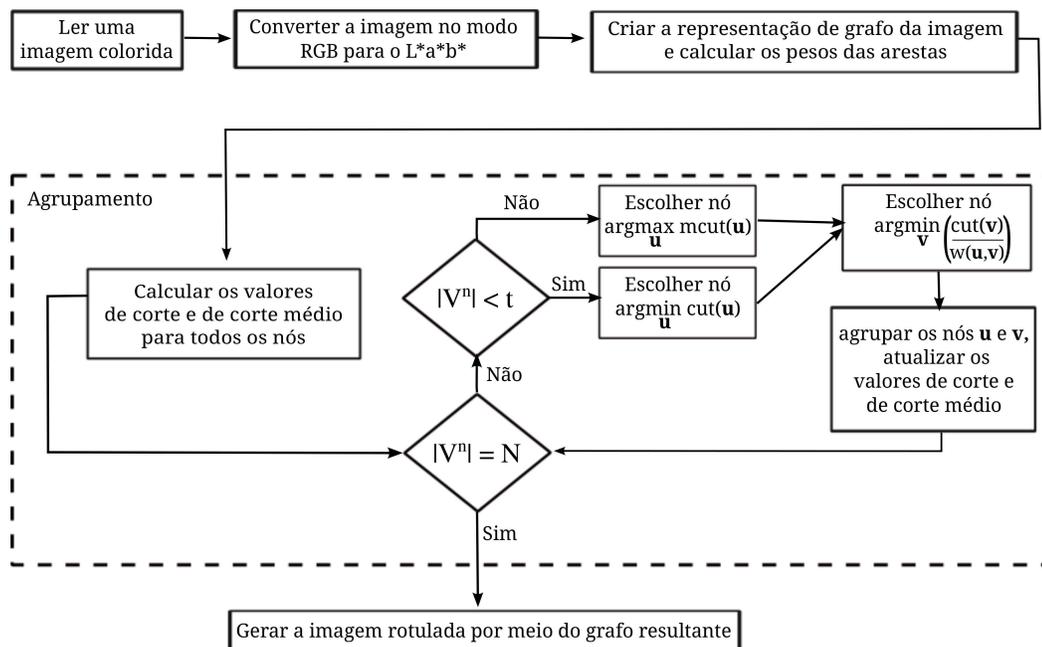


Figura 3.1: Apresentação do processo de segmentação de imagens utilizando o algoritmo hierárquico aglomerativo proposto.

3.2 Heurística

A heurística usada no algoritmo proposto é composta por duas regras: uma empregada na escolha do primeiro nó u ; e outra utilizada para determinar o segundo nó $v \sim u$. Diferente do algoritmo proposto por Costa (2013), não há uma terceira regra para restringir a formação de agrupamentos que não minimizam a função objetivo. Assim, neste trabalho o valor de corte ou de corte médio de um nó u' pode ser maior que os valores dos nós u e v que o formaram.

Uma outra particularidade do algoritmo proposto é que dois critérios são utilizados como primeira regra em diferentes etapas da execução. Os dois critérios conferem dinâmicas distintas ao processo de agrupamento, e complementares sob o ponto de vista dos tamanhos das regiões produzidas.

Primeira regra: O critério utilizado nas primeiras iterações define o nó u de forma que

$$\arg \max_u (\text{meancut}(u)). \quad (3.2)$$

O motivo para este critério é que o nó com maior valor de corte médio tenha mais chances de ter seu valor de corte médio reduzido quando agrupado com algum de seus vizinhos. Porém, observamos que este critério tende a gerar regiões com tamanhos variados, sendo que muitas delas acabam ficando muito pequenas e outras muito grandes, como pode ser observado nas Figuras 3.2(b) e 3.2(c).

Um segundo critério é utilizado para tratar este problema e, assim, gerar uma segmentação mais adequada, como, por exemplo, a segmentação exibida na Figura 3.2(d). O segundo critério passa a ser utilizado quando a ordem do grafo G^n da iteração corrente atinge um dado limiar $t \in \mathbb{N}^+$. Desta forma, em todas as iterações $n \geq t$, o primeiro nó u é definido de forma que

$$\arg \min_u (\text{cut}(u)). \quad (3.3)$$

Esse segundo critério prioriza a seleção das regiões menores, fazendo com que elas aumentem, visto que serão unidas com um de seus vizinhos. A tendência deste critério é de gerar segmentações com regiões similares em tamanho. Contudo, isso nem sempre significa que as segmentações são adequadas, como mostram os experimentos apresentados no Capítulo 4. A Figura 3.2(d) apresenta uma segmentação da imagem exibida na Figura 3.2(a) produzida pela combinação dos dois critérios descritos, com limiar $t = 1836$. Este limiar foi definido a partir de experimentos que são reportados na Seção 4.1.

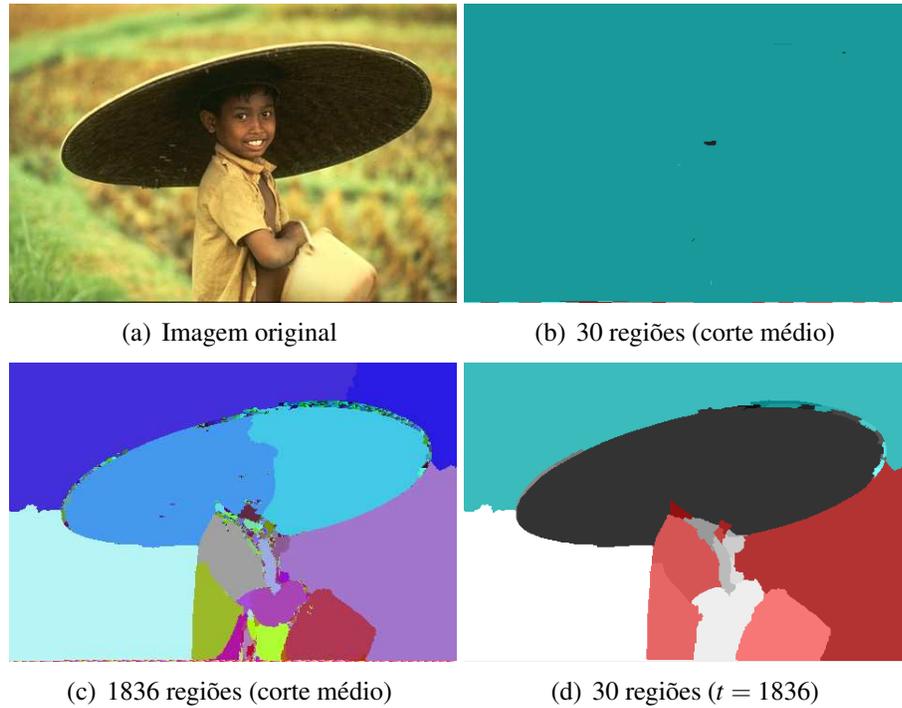


Figura 3.2: Segmentação produzida pela regra de corte médio e pela combinação das regras de corte médio e de corte. (b) segmentação utilizando a regra de corte médio até 30 regiões. (c) segmentação utilizando a regra de corte médio até 1836 regiões. (d) segmentação utilizando a combinação das regras de corte médio e de corte com $t = 1836$ com até 30 regiões.

Segunda Regra: Esta regra define o segundo nó $v \in V^n$, que deve ser adjacente ao nó u previamente escolhido. O nó v é selecionado de forma que

$$\arg \min_v \left(\frac{\text{cut}(v)}{w_{\{u,v\}}} \right). \quad (3.4)$$

Este critério seleciona o vizinho de u que possui a melhor relação entre o corte (qualidade da borda) e que seja mais similar. Observe que o critério utilizado para determinar o segundo nó é muito importante, sendo que o corte ou o valor de similaridade da aresta sozinhos não são bons critérios. O corte médio também não é um bom critério para determinar o segundo nó, ao contrário do corte normalizado que foi utilizado com sucesso por Costa (2013).

3.3 O algoritmo

O Algoritmo 1 apresenta as etapas de agrupamento dos nós de um grafo de similaridade cujo a explicação foi dada anteriormente. O critério de parada do algoritmo é o número de subgrafos

especificados por meio da variável ϕ . Observe que o limiar t especifica em quais iterações os critérios fundamentados no corte e no corte médio devem ser utilizados.

Entrada: Grafo de similaridade $G = (V, E, W)$;
 Limiar t ;
 Quantidade de regiões ϕ ;
Dados: Grafos intermediários $G^n = (V^n, E^n, W^n, D^n)$;
 Indicador da iteração i ;
 Nós intermediários u, v e u' ;
Saída: Agrupamentos de nós C ;
 $C(u) \leftarrow \{u\}, \forall u \in V$;
 $i \leftarrow 0$;
 $D^i(\{u, v\}) \leftarrow 1, \forall \{u, v\} \in E$;
 $G^i \leftarrow (V, E, W, D^i)$;
enquanto $\text{Ordem}(G^i) > \phi$ **faça**
 se $\text{Ordem}(G^i) > t$ **então**
 $u \leftarrow \arg \max_x (\text{meancut}(x)) | x \in V^i$;
 senão
 $u \leftarrow \arg \min_x (\text{cut}(x)) | x \in V^i$;
 $v \leftarrow \arg \min_x \left(\frac{\text{cut}(x)}{w_{\{u, x\}}} \right) | x \in V^i$ e $x \sim u$;
 Unir os nós u e $v \in V^i$, gerando um novo nó $u' \in V^{i+1}$ e o grafo da iteração seguinte G^{i+1} . Note que $V^{i+1} \leftarrow \{V^i \setminus \{u, v\}\} \cup \{u'\}$;
 $C(u') \leftarrow C(u) \cup C(v)$;
 $C(u) \leftarrow C(v) \leftarrow \emptyset$;
 $i \leftarrow i + 1$;

Algoritmo 3.1: Algoritmo para agrupamento de nós guiado pelas métricas de corte e corte médio.

3.4 Considerações finais

Neste capítulo foi proposto um novo algoritmo hierárquico aglomerativo guiado pelas métricas de corte e de corte médio. O processo de agrupamento dos nós é realizado dois a dois. O resultado da segmentação utilizando o algoritmo proposto pode variar dependendo da combinação entre as duas métricas, como está ilustrado na Figura 3.2.

Capítulo 4

Experimentos e resultados

Para avaliar a qualidade das segmentações geradas pelo algoritmo proposto, foi utilizado o benchmark de segmentação da *Berkeley* BSDS500. Os testes foram realizados nas imagens coloridas. As segmentações foram executadas utilizando o critério de parada de 20 regiões. Este número foi escolhido por meio da análise do número de regiões das segmentações humanas do conjunto de imagens de treinamento. Os resultados são avaliados quanto à qualidade das bordas dos segmentos, usando a Equação 2.13, e quanto à qualidade das regiões geradas, usando a Equação 2.14 e a Equação 2.15.

4.1 Teste com limiares t não adaptativos

O primeiro experimento foi realizado no conjunto de imagens de treinamento, e o objetivo era encontrar o limiar t que produzisse a melhor nota global para a medida de borda e de cobertura de região.

Inicialmente, é criada uma representação de grafos da imagem utilizando o modelo grade de *pixel* (SHI E MALIK, 2000), onde cada nó do grafo é associado a um *pixel*, e dois nós são conectados por uma aresta se os *pixels* associados estiverem dentro de um raio $r = \sqrt{2}$, que corresponde à vizinhança-8 de cada *pixel* da imagem (PEDRINI E SCHWARTZ, 2008). O peso de uma aresta é dado por

$$w_{\{u,v\}} = e^{-\frac{\Delta C}{\sigma(\Delta C)}}, \quad (4.1)$$

onde $\Delta C = \sqrt{(L_u - L_v)^2 + (a_u - a_v)^2 + (b_u - b_v)^2}$ é a distância euclidiana de cor no espaço $L^*a^*b^*$,

entre os *pixels* associados aos nós u e v .

Para localizar o limiar esperado, considerou-se que o tamanho das imagens testadas são de 481×321 *pixels*, totalizando 154401 nós no grafo inicial G^0 . A primeira busca pelo limiar t foi dividida num intervalo $[0, 154401]$ em 10 valores uniformemente distribuídos. O limiar t cuja segmentação obteve a maior nota foi escolhida para ser o pivô \hat{t} . Assim este mesmo processo é realizado recursivamente para o intervalo $[0, \hat{t}]$ até que o melhor limiar seja encontrado. A Figura 4.1 apresenta as pontuações obtidas pelo benchmark BSDS500 para diferentes valores de limiar t no conjunto de treinamento.

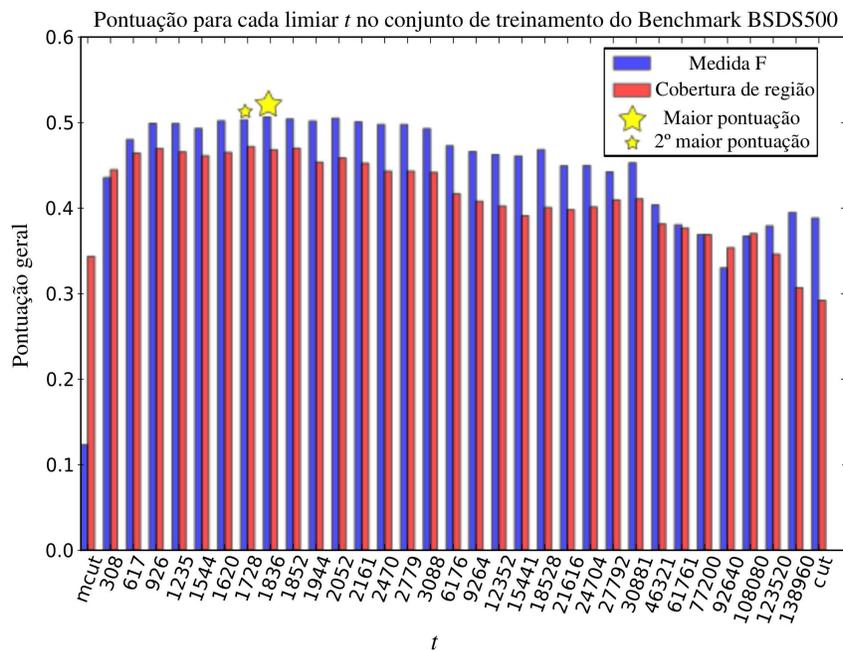


Figura 4.1: Experimento para encontrar o limiar t apropriado para o algoritmo proposto.

O limiar $t = 1836$ é o que apresentou a maior pontuação, isso representa mais do que 90% do processo de aglomeração guiado pela métrica de corte médio. Como apresentado na Figura 4.1, *mcut* refere-se ao uso de 100% da métrica corte médio e *cut* indica o uso de 100% da métrica de corte, ambos os resultados são insatisfatórios. Com isso, a combinação das duas métricas efetuadas pelo limiar t é mais adequada.

Após ter encontrado o limiar t para o conjunto de treinamento, realizou-se experimentos também para o conjunto de teste e de validação utilizando o mesmo limiar. A Figura 4.2 apresenta uma pequena amostra dos resultados das segmentações de algumas imagens em 20 regiões utilizando $t = 1836$.

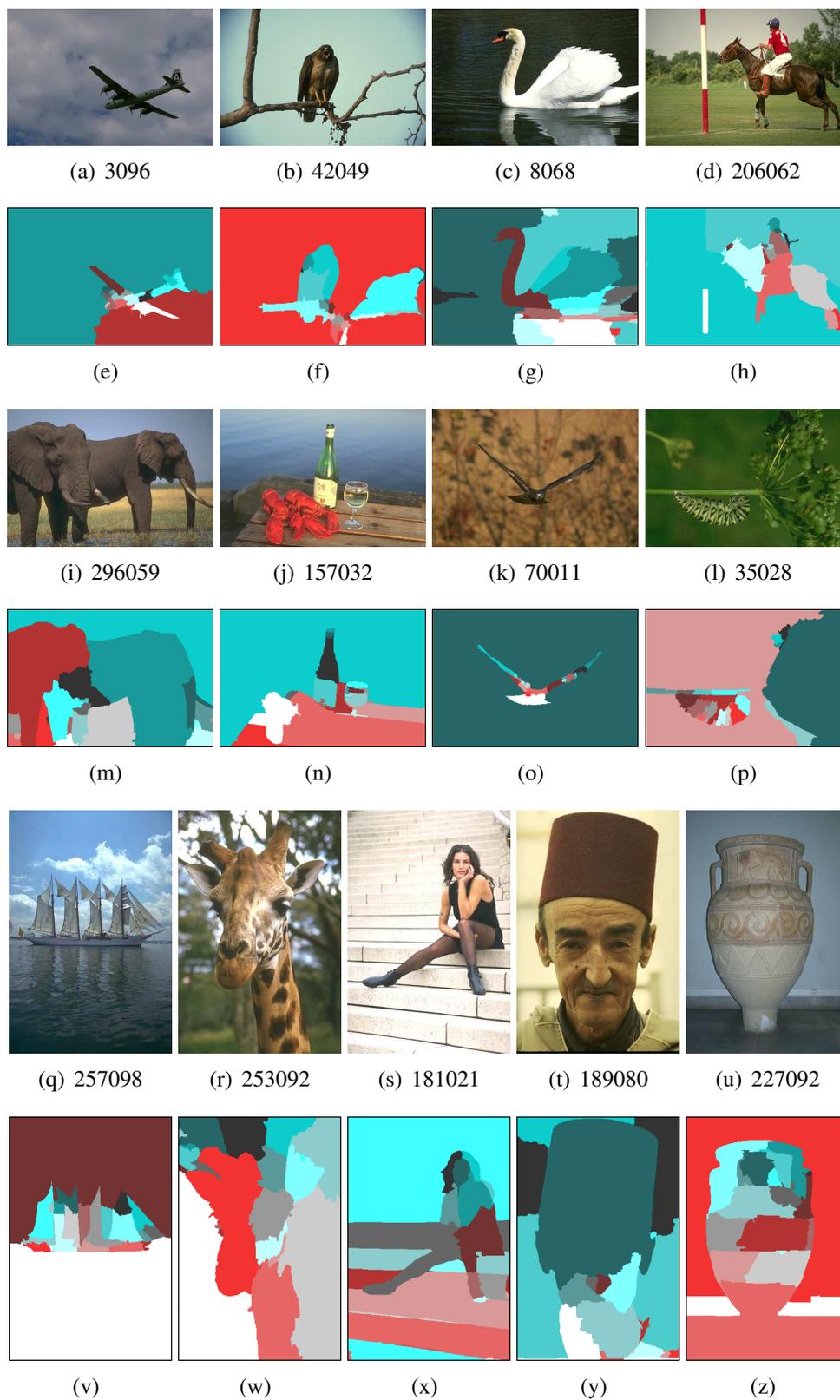


Figura 4.2: Resultado da segmentação para uma pequena amostra de imagens. As imagens (a)-(d), (i)-(l), (q)-(u) são imagens originais do conjunto de teste e de validação do banco de imagens BSDS500. As imagens (e)-(h), (m)-(p) e (v)-(z) são respectivamente os resultados da segmentação usando o algoritmo proposto com $t = 1836$.

A Tabela 4.1 apresenta a pontuação geral para os valores médios de conjunto de teste e de validação, primeira e a segunda linha da tabela respectivamente, e as pontuações individuais para as imagens apresentadas na Figura 4.2.

Tabela 4.1: Pontuação geral do conjunto de teste e de validação e a pontuação individual para a pequena amostra de imagens.

Imagens	Medida F	Cobertura de Região	PRI
Conjunto de Validação	0.4790	0.4288	0.7171
Conjunto de Teste	0.4927	0.4309	0.7231
3096	0.7133	0.7014	0.7416
42049	0.6971	0.3784	0.9120
8068	0.7276	0.5435	0.6356
206062	0.6364	0.4451	0.9109
296059	0.7459	0.6023	0.8632
157032	0.6340	0.5735	0.8799
70011	0.5605	0.8224	0.8067
35028	0.3753	0.4070	0.4936
257098	0.7037	0.6550	0.8413
227092	0.4444	0.2973	0.8663
181021	0.6786	0.4052	0.8801
189080	0.5897	0.3910	0.5860
253092	0.6987	0.4478	0.7178

As imagens 35028 e 227092 apresentaram uma pontuação baixa, como é mostrada na Tabela 4.1, entretanto as segmentações mostradas na Figura 4.2(p) e Figura 4.2(z) são visualmente satisfatórias.

4.2 Testes comparativos

Para os estudos comparativos, foram realizados as segmentações das imagens do conjunto de teste e de validação do BSDS500 em 20 regiões para os algoritmos de corte normalizado concebidos por Shi e Malik (2000) e Costa (2013), e para o algoritmo proposto neste trabalho com $t = 1836$. Em todos os casos de teste, utilizou-se o modelo de grafo por grade de *pixel*, devido a sua estrutura simples. Foram feitos dois testes utilizando diferentes métricas de similaridade: diferença de cor e força de borda.

4.2.1 Experimento 1: Diferença de cor

No modelo grade de *pixel* (SHI E MALIK, 2000) deste experimento, um dado *pixel* conecta-se aos *pixels* dentro de uma vizinhança definida por $r = \sqrt{2}$ e os pesos das arestas $w_{\{u,v\}}$ do grafo de similaridade são efetuadas pela Equação 4.1. Esta equação apresenta a métrica que calcula a diferença de cor no espaço $L^*a^*b^*$ calibrado em D65 (SACHANDA, 2007), entre os *pixels* associados aos nós u e v . Este espaço de cor foi escolhido devido à sua capacidade de simular respostas não lineares do olho humano. Também utilizou-se o exponencial da diferença de cor para efeito de normalização dos valores.

Assim, este grafo de similaridade foi dado como o grafo de entrada para o processo de agrupamento nos três algoritmos. Para o algoritmo proposto neste trabalho e no algoritmo de Costa (2013) o grafo de similaridade proporcionou boas segmentações. Mas, no caso do algoritmo de Shi e Malik (2000) suas segmentações não possuem nenhuma informação do conteúdo da imagem original. A Tabela 4.2 mostra os resultados das segmentações do conjunto de teste e de validação para os algoritmos.

Tabela 4.2: Pontuação geral dos modelos nas imagens dos conjuntos de validação e de teste do *benchmark* BSDS500 utilizando o grafo de similaridade baseado na diferença de cor. As melhores pontuações de cada métrica utilizada estão destacadas em negrito.

		Conjunto de validação			Conjunto de teste		
		Borda, medida-f	Cobertura de região	PRI	Borda, medida-f	Cobertura de região	PRI
20 regiões	Ncut espectral	0.3337	0.2116	0.6646	0.3166	0.2018	0.6887
	Ncut aglomerativo	0.4789	0.3440	0.7244	0.4862	0.3495	0.7417
	Algoritmo proposto	0.4819	0.4326	0.7201	0.4926	0.4309	0.7231

Como apresentado na Tabela 4.2, podemos observar que o algoritmo proposto obteve uma pontuação mais elevada do que a pontuação obtida pelo algoritmo de Costa (2013) e de Shi e Malik (2000). Notavelmente, a maior diferença destacou-se para os resultados da métrica de cobertura de região.

A Figura 4.3 apresenta diferentes resultados de segmentações da imagem rotulada como 70011 para os três algoritmos que estão sendo comparados neste experimento.

Em todos os resultados de segmentação do algoritmo de Shi e Malik (2000) utilizando a métrica de similaridade definida pela Equação 4.1, obtivemos resultados semelhantes ao apresentado na Figura 4.3(b). No momento do agrupamento dos autovalores e dos autovetores, o algoritmo

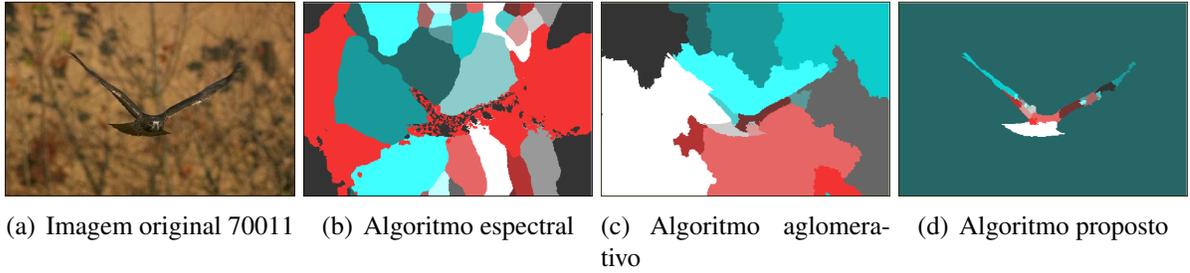


Figura 4.3: Segmentação produzida pelos três algoritmos utilizados para comparação. (a) imagem original rotulada como 70011. (b) segmentação utilizando o algoritmo baseado na teoria espectral de Shi e Malik (2000). (c) segmentação utilizando o algoritmo aglomerativo de Costa (2013). (d) segmentação utilizando o algoritmo proposto, combinado com a regra de corte médio e de corte com $t = 1836$.

espectral não leva em conta a conectividade dos nós. Cada nó é avaliado de forma isolada. Desta maneira, o grafo de similaridade testado pode não ter sido adequado à técnica de corte normalizado baseado na teoria espectral.

4.2.2 Experimento 2: Força da borda

O propósito deste experimento foi avaliar o desempenho das segmentações realizadas pelos algoritmos, na qual utilizaram a função de similaridade baseada na força da borda, definida por Cour *et al.* (2005). Dessa maneira, o peso de uma aresta $e_{\{u,v\}}$ é dado por

$$w_{\{u,v\}} = e^{-\max_{x \in \text{Line}(u,v)} \frac{\|\text{Edge}(x)\|^2}{\sigma_C}}, \quad (4.2)$$

onde $\text{Line}(u,v)$ é um segmento de reta com as extremidades nos *pixels* u e v , $\text{Edge}(x)$ é a força da borda na posição x e σ_C é um fator escalar para o contorno. Esta função é aplicada a um gradiente da imagem. Neste modelo de similaridade usa-se imagens em tons de cinza para a ponderação do grafo.

Na configuração deste experimento, o modelo *pixel grid* foi definido com uma vizinhança de $r = \sqrt{10}$ com uma taxa de amostragem de 30%. A métrica baseada na força da borda cria de forma aleatória os seus pesos. Logo, gerou-se um único grafo a partir desta métrica e esse mesmo foi aplicado aos três algoritmos para efeito de uma comparação precisa.

A Tabela 4.3 apresenta os resultados das imagens segmentadas pelos três algoritmos utilizando o grafo de similaridade baseado na força da borda. O grafo de similaridade por força da borda possui uma estrutura mais densa. Por esse motivo, os testes com o algoritmo proposto foram

realizados apenas para o conjunto de validação.

Tabela 4.3: Pontuação geral dos modelos nas imagens dos conjuntos de validação e de teste do *benchmark* BSDS500 utilizando o grafo de similaridade baseado na força da borda. As melhores pontuações de cada métrica utilizada estão destacadas em negrito.

		Conjunto de validação			Conjunto de teste		
		Borda, medida-f	Cobertura de região	PRI	Borda, medida-f	Cobertura de região	PRI
20 regiões	Ncut espectral	0.5349	0.2950	0.7214	0.5305	0.3004	0.7416
	Ncut aglomerativo	0.4598	0.3877	0.7195	0.4750	0.3941	0.7367
	Algoritmo proposto	0.4682	0.4314	0.7201	–	–	–

Um resultado interessante foi que o algoritmo proposto apresentou melhores pontuações para a medida de cobertura de região em ambos os experimentos, tanto no experimento 1 quanto neste experimento 2. Isso revela que a segmentação do algoritmo proposto obteve um grande acerto em relação à sobreposição e à quantidade de *pixels* contidos em cada partição.

4.2.3 Comparação dos modelos de similaridade

Nos experimentos 1 e 2, foram utilizados dois modelos diferentes de grafos de similaridade para serem processados pelos algoritmos. Para o algoritmo baseado na teoria espectral observamos que o grafo usado no experimento 1 não se adequou ao processo de agrupamento. Mas, para o algoritmo proposto e para o concebido por Costa (2013) o modelo foi apropriado. Assim, pelas Figuras 4.4 e 4.5 podemos observar a influência do uso dos modelos de grafos.

Nas Figuras 4.4 e 4.5 apresentam-se os resultados da segmentação de uma imagem do conjunto de teste utilizando os três algoritmos, cujo conjunto de segmentações da Figura 4.4(b) ao (d) são resultantes do uso do modelo de similaridade por diferença de cor, enquanto que da 4.5(b) ao (d) são resultados das segmentações geradas pelo uso do modelo de similaridade por meio força da borda.

A Tabela 4.4 mostra as pontuações das segmentações para a imagem 201080 produzidas pelos algoritmos. Neste teste de comparação em especial foi executado o algoritmo proposto utilizando o grafo baseado na força de borda somente para a imagem 201080, visto que não foi executada para as imagens do conjunto de teste.

De todas as imagens do conjunto de teste, a segmentação da imagem 201080 foi a que recebeu a maior pontuação no experimento 1 para o algoritmo proposto. Mas, sua pontuação diminuiu

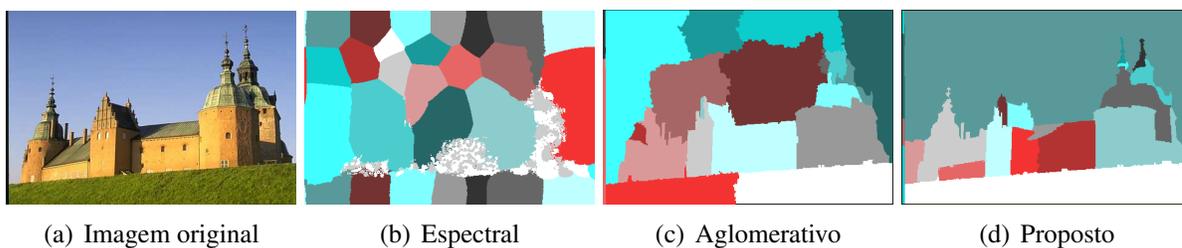


Figura 4.4: Segmentação da imagem 201080 do conjunto de teste utilizando o grafo de similaridade por distância de cor (a) imagem original. (b) segmentação utilizando o algoritmo baseado na teoria espectral de Shi e Malik (2000). (c) segmentação utilizando o algoritmo aglomerativo de Costa (2013). (d) segmentação utilizando o algoritmo proposto, combinado com a regra de corte médio e de corte com $t = 1836$.

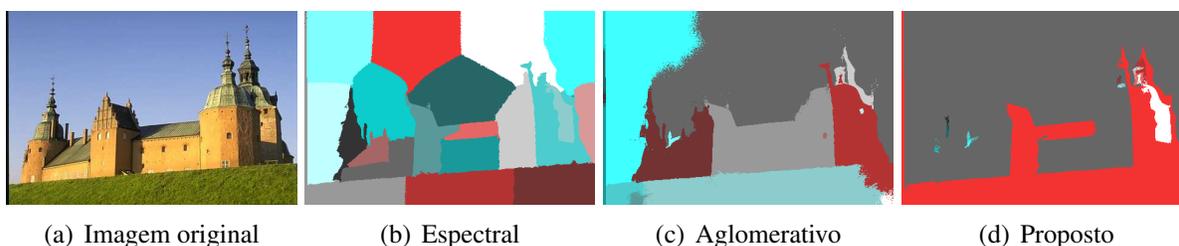


Figura 4.5: Segmentação da imagem 201080 do conjunto de teste utilizando o grafo de similaridade por força da borda (a) imagem original. (b) segmentação utilizando o algoritmo baseado na teoria espectral de Shi e Malik (2000). (c) segmentação utilizando o algoritmo aglomerativo de Costa (2013). (d) segmentação utilizando o algoritmo proposto, combinado com a regra de corte médio e de corte com $t = 1836$.

quando submetida ao modelo de grafo do experimento 2, como vista nas Figuras 4.4 e 4.5, e na Tabela 4.4. A Figura 4.6 apresenta um exemplo de resultado de segmentação obtido pelo algoritmo de Costa (2013) e o algoritmo proposto. A segmentação apresentada foi obtido via experimento 1.

Nos resultados de segmentações vistas nas Figuras 4.4, 4.5 e 4.6, observamos que no algoritmo aglomerativo de Costa (2013) formam-se pequenas regiões ao fundo da imagem. Nos resultados do algoritmo proposto, devido à combinação das métricas de corte e de corte médio há uma tendência para que essas regiões menores se unam proporcionando o destaque do fundo da imagem.

4.3 Considerações finais

Neste capítulo foram apresentados experimentos para encontrar o limiar t adequado a fim de realizar a combinação apropriada dos dois métodos, o corte e o corte médio. Feito isso realizamos testes comparativos para verificar a qualidade das segmentações geradas pelos três algoritmos: o

Tabela 4.4: Pontuação individual da imagem 201080 do conjunto de teste do BSDS500 utilizando o grafos de similaridade baseado na diferença de cor e na força da borda As melhores pontuações de cada métrica utilizada estão destacadas em negrito.

		Diferença de cor			Força da Borda		
		Borda, medida-f	Cobertura de região	PRI	Borda, medida-f	Cobertura de região	PRI
20 regiões	Ncut espectral	0.3373	0.1309	0.2906	0.6768	0.2651	0.6555
	Ncut aglomerativo	0.6677	0.3484	0.6406	0.5630	0.5739	0.5399
	Algoritmo proposto	0.8120	0.7935	0.8226	0.5905	0.5202	0.7328

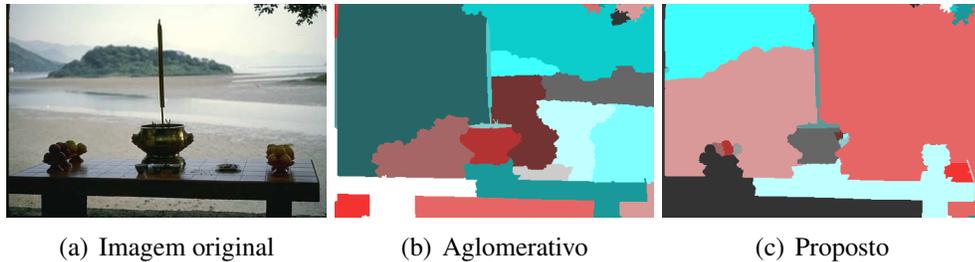


Figura 4.6: Segmentação da imagem 120003 do conjunto de teste para os algoritmos aglomerativos. (a) imagem original. (b) segmentação utilizando o algoritmo aglomerativo de Costa (2013) (c) segmentação utilizando o algoritmo proposto, combinado com a regra de corte médio e de corte com $t = 1836$.

algoritmo proposto, algoritmo de Shi e Malik (2000) e de Costa (2013). Nos testes foram considerados dois tipos de modelo de grafo de similaridade, por distância de cor e por força da borda . Em todos os experimentos apresentados neste capítulo, foram utilizados o modelo *pixel grid* devido a sua simplicidade e ampla utilização.

Obtivemos resultados significativos e expressivos para o algoritmo proposto, principalmente na métrica de cobertura de região e PRI, também fundamentada em regiões. No experimento 1 e no experimento 2 foi possível observar que suas pontuações mantiveram-se elevadas nessas métricas.

A implementação do modelo de grade de *pixel*, assim como o algoritmo de corte normalizado¹ por teoria espectral, utilizados nos experimentos, foram as disponibilizadas publicamente por Cour *et al.* (2005). O algoritmo desenvolvido por Costa (2013) foi implementado na linguagem Python (MARTELLI *et al.*, 2005), assim como o algoritmo proposto. Todos os experimentos foram executados em uma máquina de arquitetura i286, com processador de 2.2 GHz, com memória de 3.9 GB e sistema operacional Linux.

¹O algoritmo de corte normalizado pode ser encontrado em: www.cis.upenn.edu/~jshi/software

Capítulo 5

Conclusão

Neste trabalho foi apresentado um novo algoritmo aglomerativo de agrupamento hierárquico guiado pelas métricas de corte e corte médio, projetado em função da qualidade da segmentação produzida ao invés de ser na otimização de uma função objetivo de corte. A heurística do algoritmo combina os dois critérios, que possuem dinâmicas distintas no processo de agrupamento, para tornar a segmentação mais adequada do que as produzidas se os critérios são empregados separadamente.

Experimentos utilizando o conjunto de treinamento do BSDS500 foram feitos para determinar um limiar t fixo adequado para as imagens dos conjuntos de validação e de teste, os quais são utilizados para comparar o desempenho frente a outras técnicas. O melhor limiar encontrado, segundo a medida-F e a medida de cobertura de região, é $t = 1836$. Contudo, um limiar fixo pode não ser apropriado para todas as imagens.

As qualidades das segmentações do novo algoritmo são comparadas com as qualidades das segmentações produzidas pelos algoritmos de corte normalizado aglomerativo e espectral. Foram realizados dois testes com configurações distintas do modelo de grade de *pixels*. Na primeira configuração, a função de similaridade fundamenta-se na diferença da cor, e os grafos gerados são regulares, de grau 8. Na segunda configuração, a função de similaridade baseia-se na força da borda que intercepta uma linha reta entre os *pixels* associados aos nós conectados pelas arestas. Os graus médios dos grafos gerados pela segunda configuração são cerca de 10 vezes os graus dos grafos gerados pela primeira. Esta configuração é mais adequada ao algoritmo espectral, enquanto que para os algoritmos aglomerativos a primeira é mais interessante, pois: garante que as regiões serão conexas; e favorece a eficiência.

O algoritmo proposto apresentou pontuações altas nos experimentos realizados, em especial, nas medidas de cobertura de região e PRI. Isso revela que a sobreposição da segmentação produ-

zida com a segmentação humana contida no *benchmark* é melhor para a segmentação dada pelo algoritmo proposto, de modo geral.

Esta pesquisa contribuiu para ampliar os conhecimentos sobre corte em grafos, segmentação e sobre a abordagem de utilização de algoritmos aglomerativos combinados com métricas de corte. Foi verificado que o algoritmo proposto de forma geral foi capaz de produzir segmentações melhores que as produzidas pelo algoritmo desenvolvido para minimizar a função objetivo de corte normalizado. Isso revela que usos alternativos das métricas de corte em segmentação possuem potencial para apresentar resultados melhores para muitos problemas da área. Embora este estudo tenha apresentado resultados interessantes e promissores, será necessário ainda muito esforço de pesquisa para melhor compreender todos os aspectos envolvidos. Algumas sugestões para trabalhos futuros são apresentadas na Seção 5.2, a seguir. No entanto, não há dúvidas sobre ser grande o poder que as ferramentas de corte em grafos possuem.

5.1 Publicações

Um artigo relacionado à dissertação foi aprovada e está em vias de publicação na conferência VISAPP (*International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.*). Neste artigo foi abordado o método de agrupamento hierárquico aqui proposto, com sua heurística característica. Foram apresentados também resultados preliminares da busca do limiar t utilizando o conjunto de treinamento e uma comparação do algoritmo proposto com o algoritmo aglomerativo de Costa (2013).

5.2 Trabalhos futuros

Um trabalho futuro bastante pertinente seria definir um limiar adaptativo para cada tipo de imagem, podendo levar a melhores segmentações. De importância equivalente é verificar se a combinação de outros critérios leva a resultados mais interessantes, para aplicações gerais ou específicas.

Para simplificar, neste trabalho foi utilizado apenas o modelo de grade de *pixels* em duas configurações distintas. Uma com a função de similaridade baseada na distância Euclideana das cores no espaço $L^*a^*b^*$; e a outra com função de similaridade calculada a partir da interceptação de contornos. Assim, ainda é necessário verificar o desempenho dos algoritmos em uma quantidade maior de modelos e configurações de grafos de similaridade. Outros modelos de cor, como o HSV e o HSI, podem ser utilizados. De fato, a modelagem do grafo de similaridade dá muita flexibilidade

às ferramentas de corte, porém, também aumentam a complexidade de uma análise comparativa e de estudos que envolvem a solução de problemas específicos.

Referências Bibliográficas

- ABREU, N.M.M. Teoria espectral dos grafos: um híbrido entre a álgebra linear e a matemática discreta e combinatória com origens na química quântica. **TEMA Tendências em Matemática Aplicada e Computacional**, v. 6, 1–10, 2005.
- ARBELÁEZ, P.; MAIRE, M.; FOWLKES, C. e MALIK, J. Contour detection and hierarchical image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 33, n. 5, 898–916, maio 2011.
- CARVALHO, M.A.G. e COSTA, A.L. **New Frontiers in Graph Theory**, capítulo Combining hierarchical structures on graphs and Normalized Cut for image segmentation. Intech, 2012.
- CARVALHO, M.A.G.; FERREIRA, A.C.B. e COSTA, A.L. Image segmentation using quadtree-based similarity graph and normalized cut. In **15th CIARP – Iberoamerican Congress on Pattern Recognition**. São Paulo–SP, Brasil, 2010.
- CARVALHO, M.A.G.; FERREIRA, A.C.B.; PINTO, T.W. e CESAR–JR, R.M. Image segmentation using watershed and normalized cuts. In **Proc. of 22th Conference on Graphics, Patterns and Images (SIBGRAPI)**. Rio de Janeiro - Brazil, 2009.
- COSTA, A. L. **Corte normalizado em grafos : um algoritmo aglomerativo para segmentação de imagens de colônias de bactérias**. 2013. Dissertação de mestrado. Faculdade de Tecnologia - FT, Universidade Estadual de Campinas - UNICAMP, Brasil.
- COUR, T.; BÉNÉZIT, F. e SHI, J. Spectral segmentation with multiscale graph decomposition. In **Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition – CVPR’05**, v. 2, pp. 1124–1131. 2005.
- CVETKOVIĆ, D.; ROWLINSON, P. e SIMIĆ, S. **An Introduction to the Theory of Graph Spectra**. Cambridge University Press, UK, 2010.
- FERREIRA, A. C. B. **Um Estudo Comparativo de Segmentação de Imagens por Aplicações do Corte Normalizado em Grafos**. 2011. Master. University of Campinas, Campinas, São Paulo, Brazil.
- FIEDLER, M.A. Property of eigenvectors of nonnegative symmetric matrices and its applications to graph theory. **Czech Math Journal**, v. 25, n. 100, 619–633, 1975.
- GIBBONS, A. **Algorithmic Graph Theory**, v. 2. Cambridge University Press, 1999.

- GOLDBARG, M. e GOLDBARG, E. **Grafos: conceitos, algoritmos e aplicações**. Elsevier, 2012.
- GONZALEZ, R. e WOODS, R. **Processamento digital de imagens**. Addison–Wesley, 3^a ed., 2010.
- GOWER, J.C. e ROSS, G.J.S. Minimum spanning trees and single linkage cluster analysis. **Applied statistics**, pp. 54–64, 1969.
- JAIN, A. e DUBES, R. **Algorithms for Clustering Data**. Prentice Hall, 1988.
- KODAK. **Exploring the Color Image**. Eastman Kodak Company, Kodak publication H-188, United States of America, 2000.
- MA, X. e WAN, W. Texture image segmentation on improved watershed and multiway spectral clustering. In **Proceedings of International Conference on Audio, Language and Image Processing – ICALIP**, pp. 1693–1697. 2008.
- MARTELLI, A.; RAVENSCROFT, A. e ASCHER, D. **Python Cookbook**. O’Reilly Media, 2 ed., 2005.
- PEDRINI, H. e SCHWARTZ, W.R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. Thomson, 2008.
- PENG, B.; ZHANG, L. e ZHANG, D. A survey of graph theoretical approaches to image segmentation. **Pattern Recognition**, v. 46, n. 3, 1020–1038, 2013.
- SACHANDA, J. **Colorimetry: Understanding the CIE System**. Wiley, Hoboken, New Jersey, 2007.
- SCHAEFFER, S.E. Graph clustering. **Computer Science Review**, v. 1, 27–64, 2007.
- SHI, J. e MALIK, J. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 1, 888–905, 2000.
- SIBSON, R. Slink: An optimally efficient algorithm for the single-link cluster method. **The Computer Journal**, v. 16, n. 1, 1973.
- STONE, M.C. A survey of color for computer graphics. In **SIGGRAPH**, Course 4. StoneSoup Consulting, August 2001.
- WANG, S. e SISKIND, J.M. Image segmentation with minimum mean cut. In **Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on**, v. 1, pp. 517–524. 2001.
- WENBING, T. Image thresholding using graph cuts. **IEEE transactions on systems, man, and cybernetics**, v. 36, n. 11, 1181–1182, 2008.
- WILSON, R.J. e WATKINS, J.J. **Graphs, an introductory approach**. John Wiley & Sons, 1990.
- WU, J. Cluster analysis and k-means clustering: An introduction. **Advances in K-means Clustering**, pp. 1–16, 2012.

WU, S.T. Síntese de imagens: Uma introdução ao mundo de desenho e pintura dos sistemas digitais. Livro texto: Sistemas de informações gráficas, Departamento de Engenharia de Computação e Automação Industrial, Faculdade de Engenharia e de Computação, Universidade Estadual de Campinas, Campinas - SP, Brasil, Março 2009.

WU, Z. e LEAHY, R. An optimal graph theoretic approach to data clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 15, n. 11, 1101–1113, 1993.