

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**GUILHERME GARCIA SCHU PEIXOTO**

**SEGMENTAÇÃO DE IMAGENS COLORIDAS POR  
ÁRVORES BAYESIANAS ADAPTATIVAS**

Porto Alegre  
2017

**GUILHERME GARCIA SCHU PEIXOTO**

**SEGMENTAÇÃO DE IMAGENS COLORIDAS POR  
ÁRVORES BAYESIANAS ADAPTATIVAS**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.  
Área de concentração: Engenharia da Computação

**ORIENTADOR: Prof. Dr. Jacob Scharcanski**

Porto Alegre  
2017

**GUILHERME GARCIA SCHU PEIXOTO**

**SEGMENTAÇÃO DE IMAGENS COLORIDAS POR  
ÁRVORES BAYESIANAS ADAPTATIVAS**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: \_\_\_\_\_

Prof. Dr. Jacob Scharcanski, UFRGS

Doutor pela University of Waterloo – Waterloo, Canadá

**Banca Examinadora:**

Prof. Dr. Altamiro Amadeu Susin, PPGEE-UFRGS

Doutor pelo Institut National Polytechnique de Grenoble – Grenoble, França

Prof. Dr. Cláudio Rosito Jung, PPGC-UFRGS

Doutor pela Universidade Federal do Rio Grande do Sul – Porto Alegre, Brasil

Prof. Dr. Valner João Brusamarello, PPGEE-UFRGS

Doutor pela Universidade Federal de Santa Catarina – Florianópolis, Brasil

Coordenador do PPGEE: \_\_\_\_\_

Prof. Dr. Valner João Brusamarello

Porto Alegre, julho de 2017.

## DEDICATÓRIA

Dedico este trabalho aos meus pais, aos meus bons amigos e minha namorada. Em um mundo complexo e multi-perspectivo tal como o que vivemos hoje, seria ingenuidade minha acreditar que apenas a razão pura, aquela que reside nos confins da nossa imaginação e que se externaliza por meio de símbolos e números convencionados, traria alguma espécie de conforto para a minha vida pessoal e profissional. Mas sim, acredito ser, a superposição das constantes e contínuas reflexões da minha mãe, das experiências diversas relatadas pelo Pastor, os inúmeros abraços e carinhos da Mari, junto a investigação árdua e incansável necessária para entender o mundo em que vivemos, que de alguma forma me traz um pouco de felicidade duradoura. A presença destas pessoas na minha vida complementam as muitas horas em silêncio de estudo, e sem estas, provavelmente, a minha ciência da realidade tenderia ao delírio, ou, a ilusão.

## **AGRADECIMENTOS**

Gostaria de agradecer ao meu orientador, Professor Jacob, pela imensa paciência, dedicação e pelas profundas interpretações dos diversos conceitos que envolveram este trabalho.

À banca examinadora, pelas relevantes críticas e ponderações apontadas.

Ao Programa de Pós-Graduação em Engenharia Elétrica, PPGEE, pela oportunidade de realização de trabalhos em minha área de pesquisa.

Aos colegas de laboratório Lucas, Muhammad, John e Eric, pelas colaborações e auxílio nos diversos desafios associados ao desenvolvimento deste estudo.

Aos grandes amigos Eliezer e Yuri, pelas conversas inspiradoras, frustrações compartilhadas e perspectivas de futuro dentro da academia.

À CAPES pela provisão da minha bolsa de mestrado.

E gostaria de agradecer ao contribuinte brasileiro, que de forma indireta, por intermédio da CAPES e do CNPq, financia uma grande parcela da pesquisa feita no Brasil. Nenhum projeto de nação contemporâneo deve subestimar a importância da Ciência, da pesquisa e da tecnologia para o desenvolvimento de um país.

## RESUMO

A segmentação de imagens consiste em uma tarefa de fundamental importância para diferentes aplicações em visão computacional, tais como por exemplo, o reconhecimento e o rastreamento de objetos, a segmentação de tumores/lesões em aplicações médicas, podendo também servir de auxílio em sistemas de reconhecimento facial. Embora exista uma extensa literatura abordando o problema de segmentação de imagens, tal tópico ainda continua em aberto para pesquisa. Particularmente, a tarefa de segmentar imagens coloridas é desafiadora devido as diversas inomogeneidades de cor, texturas e formas presentes nas feições descritivas das imagens. Este trabalho apresenta um novo método de *clustering* para abordar o problema da segmentação de imagens coloridas. Nós desenvolvemos uma abordagem Bayesiana para procura de máximos de densidade em uma distribuição discreta de dados, e representamos os dados de forma hierárquica originando *clusters* adaptativos a cada nível da hierarquia. Nós aplicamos o método de *clustering* proposto no problema de segmentação de imagens coloridas, aproveitando sua estrutura hierárquica, baseada em propriedades de árvores direcionadas, para representar hierarquicamente uma imagem colorida. Os experimentos realizados revelaram que o método de *clustering* proposto, aplicado ao problema de segmentação de imagens coloridas, obteve para a medida de performance *Probabilistic Rand Index (PRI)* o valor de 0.8148 e para a medida *Global Consistency Error (GCE)* o valor 0.1701, superando um total de vinte e um métodos previamente propostos na literatura para o banco de dados BSD300. Comparações visuais confirmaram a competitividade da nossa abordagem em relação aos demais métodos testados. Estes resultados enfatizam a potencialidade do nosso método de *clustering* para abordar outras aplicações no domínio de Visão Computacional e Reconhecimento de Padrões.

**Palavras-chave:** Clustering, Segmentação de imagens coloridas, Árvores direcionadas, Teoria de decisão Bayesiana.

## ABSTRACT

Image segmentation is an essential task for several computer vision applications, such as object recognition, tracking and image retrieval. Although extensively studied in the literature, the problem of image segmentation remains an open topic of research. Particularly, the task of segmenting color images is challenging due to the inhomogeneities in the color regions encountered in natural scenes, often caused by the shapes of surfaces and their interactions with the illumination sources (e.g. causing shading and highlights). This work presents a novel non-supervised classification method. We develop a Bayesian framework for seeking modes on the underlying discrete distribution of data and we represent data hierarchically originating adaptive clusters at each level of hierarchy. We apply the proposal clustering technique for tackling the problem of color image segmentation, taking advantage of its hierarchical structure based on hierarchy properties of directed trees for representing fine to coarse levels of details in an image. The experiments herein conducted revealed that the proposed clustering method applied to the color image segmentation problem, achieved for the Probabilistic Rand Index (*PRI*) performance measure the value of 0.8148 and for the Global Consistency Error (*GCE*) the value of 0.1701, outperforming twenty-three methods previously proposed in the literature for the BSD300 dataset. Visual comparison confirmed the competitiveness of our approach towards state-of-art methods publicly available in the literature. These results emphasize the great potential of our proposed clustering technique for tackling other applications in computer vision and pattern recognition.

**Keywords:** Clustering, color image segmentation, directed trees, Bayesian decision theory.

# SUMÁRIO

<b>LISTA DE ILUSTRAÇÕES</b> . . . . .	9
<b>LISTA DE TABELAS</b> . . . . .	13
<b>LISTA DE ABREVIATURAS</b> . . . . .	14
<b>LISTA DE SÍMBOLOS</b> . . . . .	15
<b>1 INTRODUÇÃO</b> . . . . .	19
<b>2 CONCEITOS BÁSICOS EM RECONHECIMENTO DE PADRÕES</b> . . . . .	25
<b>2.1 Representação de Dados de uma Imagem</b> . . . . .	25
<b>2.2 Classificação Supervisionada e Não-Supervisionada</b> . . . . .	28
2.2.1 Classificação Supervisionada . . . . .	29
2.2.2 Classificação Não-Supervisionada . . . . .	29
<b>3 REVISÃO DA LITERATURA</b> . . . . .	31
<b>3.1 Métodos Baseados em Contornos</b> . . . . .	32
3.1.1 Completamento de Bordas . . . . .	32
3.1.2 Exploração da Informação de Escala . . . . .	33
<b>3.2 Métodos Baseados em Regiões</b> . . . . .	34
3.2.1 Limiarização de Histogramas . . . . .	34
3.2.2 <i>Clustering</i> . . . . .	35
<b>3.3 Trabalhos Relacionados</b> . . . . .	38
3.3.1 <i>Mean Shift</i> . . . . .	38
3.3.2 <i>Superpixels</i> . . . . .	40
<b>4 FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	42
<b>4.1 A Regra de Decisão Bayesiana</b> . . . . .	42
<b>4.2 Máxima Probabilidade a <i>posteriori</i></b> . . . . .	46
<b>4.3 Distribuições Gaussianas</b> . . . . .	47
4.3.1 Distribuição Gaussiana Unidimensional . . . . .	47
4.3.2 Distribuição Gaussiana Multivariada . . . . .	48
4.3.3 Mistura de Gaussianas Multivariadas . . . . .	50
<b>4.4 Conceitos Relevantes de Teoria de Grafos</b> . . . . .	51
4.4.1 Árvores Direcionadas . . . . .	52

<b>5</b>	<b>METODOLOGIA</b>	54
<b>5.1</b>	<b>Formulação do Problema e Resumo do Algoritmo</b>	54
5.1.1	Resumo da Abordagem Proposta	55
<b>5.2</b>	<b>Método Proposto</b>	57
5.2.1	Árvores Bayesianas Adaptativas	57
5.2.2	Representação Hierárquica de Imagens Coloridas	63
<b>6</b>	<b>RESULTADOS EXPERIMENTAIS</b>	67
<b>6.1</b>	<b>Medidas de Performance</b>	68
6.1.1	<i>Probabilistic Rand Index (PRI)</i>	68
6.1.2	<i>Variation of Information (VOI)</i>	69
6.1.3	<i>Global Consistency Error (GCE)</i>	69
6.1.4	<i>Boundary Displacement Error (BDE)</i>	70
<b>6.2</b>	<b>Configuração Experimental e Efeitos na Segmentação</b>	71
<b>6.3</b>	<b>Ajuste de Parâmetros e Análise de Variâncias</b>	75
<b>6.4</b>	<b>Resultados Quantitativos</b>	80
<b>6.5</b>	<b>Comparação com Outros Métodos</b>	82
<b>7</b>	<b>DISCUSSÃO</b>	87
<b>8</b>	<b>CONCLUSÕES</b>	93
<b>8.1</b>	<b>Trabalhos Futuros</b>	94
	<b>REFERÊNCIAS</b>	97

## LISTA DE ILUSTRAÇÕES

Figura 1:	Conjunto de imagens exemplo (MARTIN et al., 2001). . . . .	20
Figura 2:	Exemplo de segmentação de uma imagem colorida. Em (a) apresenta-se a imagem original e em (b) apresenta-se a respectiva segmentação delimitada por contornos. Na imagem original o número de segmentos (partições) é igual ao número de <i>pixels</i> da imagem. Na imagem segmentada o número de segmentos é igual a nove (MARTIN et al., 2001). . . . .	21
Figura 3:	Exemplos de segmentação manual (MARTIN et al., 2001). . . . .	22
Figura 4:	Representação do espaço de feições (b) contendo todos os <i>pixels</i> de uma imagem colorida (a) (MARTIN et al., 2001). . . . .	26
Figura 5:	Bandeira do Japão utilizada como exemplo de um simples problema de classificação de <i>pixels</i> . . . . .	28
Figura 6:	Diagrama de Venn representando dois possíveis eventos $A$ e $B$ em um universo de eventos $U$ . . . . .	43
Figura 7:	Diagrama de Venn representando a intersecção de dois possíveis eventos $A$ e $B$ em um universo de eventos $U$ . . . . .	44
Figura 8:	Representação gráfica de distribuições Gaussianas para uma variável contínua e unidimensional: (a) representação de uma distribuição Gaussiana com parâmetros $\mu = 0$ e $\sigma^2 = 4$ ; (b) representação de uma distribuição Gaussiana com parâmetros $\mu = 2$ e $\sigma^2 = 4$ . . . . .	48
Figura 9:	Representação gráfica de distribuições Gaussianas: (a) Representação tridimensional de uma distribuição Gaussiana com parâmetros $\mu = [0, 0]$ e $\Sigma = 5 \times \mathbb{I}$ , onde $\mathbb{I}$ é a matriz identidade de tamanho $2 \times 2$ ; (b) Representação da distribuição Gaussiana mostrada em (a) numa perspectiva bidimensional; (c) Representação tridimensional de uma distribuição Gaussiana com parâmetros $\mu = [0, 0]$ e $\Sigma = 10 \times \mathbb{I}$ , onde $\mathbb{I}$ é a matriz identidade de tamanho $2 \times 2$ ; (d) Representação da distribuição Gaussiana mostrada em (c) numa perspectiva bidimensional. . . . .	49
Figura 10:	Representação gráfica de uma mistura de três ( $J = 3$ ) distribuições Gaussianas multivariadas: (a) Representação bidimensional da combinação linear de três Gaussianas com parâmetros: $\mu_1 = [1, 2]$ ; $\mu_2 = [-1, -2]$ ; $\mu_3 = [-5, -4]$ ; e $\Sigma_1 = \Sigma_2 = \Sigma_3 = 2 \times \mathbb{I}$ , onde $\mathbb{I}$ é a matriz identidade de tamanho $2 \times 2$ ; (b) Representação da mistura de Gaussianas mostrada em (a) numa perspectiva tridimensional. . . . .	51
Figura 11:	Representação de um grafo não-direcionado acíclico com 8 vértices (em azul) e 14 arestas (em preto). . . . .	52

- Figura 12: Representação de duas árvores direcionadas. A primeira árvore possui o vértice raiz  $v_{r=1}$  e a segunda árvore possui o vértice raiz  $v_{r=7}$ . . . . . 53
- Figura 13: Visão geral do algoritmo de segmentação de imagem proposto. . . . . 56
- Figura 14: Ilustração da interseção expressa pelo conjunto  $\{\eta_{\Phi}^j \cap \mathcal{E}^j\}$  no espaço de cores RGB: (a) a imagem colorida  $I$  é decomposta em regiões coloridas homogêneas, obtidas pelo algoritmo *SLIC* (contornos das regiões em vermelho), tal que  $\bigcup_{j=1}^J R_j = I$ . Utilizando um grafo de regiões adjacentes  $\mathcal{G}$ , a vizinhança espacial de uma região arbitrária  $R_j$  é representada (nodos azuis indicam as regiões adjacentes a região  $R_j$  e as arestas que as conectam estão marcadas em verde); (b) cada região da imagem  $I$  esta associada ao valor médio das cores RGB ( $\mathbf{x}_j$ ) dos *pixels* daquela região no espaço de feições. Vetores de cores em destaque azul indicam os valores das características médias dos vizinhos espaciais da região  $R_j$  em  $\mathcal{G}$ , e uma esfera vermelha de raio  $\Phi$  delimita a similaridade entre  $\mathbf{x}_j$  e seus vizinhos no espaço de feições; (c)  $\mathbf{x}_j$  possui vizinhos no espaço de feições definidos pelo conjunto  $\eta_{\Phi}^j$ ; (d) os vetores característicos associados a vizinhança espacial da região  $R_j$  são dados pelo conjunto  $\mathcal{E}^j$ ; (e) o conjunto  $\{\eta_{\Phi}^j \cap \mathcal{E}^j\}$  contém os vetores característicos candidatos que potencialmente podem se tornar nodos raiz no espaço de feições. . . . . 60
- Figura 15: Ilustração do procedimento para unir nodos de árvores direcionadas em um espaço de feições abstrato de três dimensões: a) árvores direcionadas originais com raízes  $\mathbf{x}_{1,r}$ ,  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$ ; b) *clusters* obtidos por meio da atribuição:  $\mathbf{x}_{1,r} \rightarrow \mathbf{x}_{1,r}$ ;  $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8 \rightarrow \mathbf{x}_{5,r}$ ;  $\mathbf{x}_{10}, \mathbf{x}_{11}, \mathbf{x}_{12} \rightarrow \mathbf{x}_{9,r}$ ; c) as probabilidades *a priori*  $\Pi_1, \Pi_5$  e  $\Pi_9$  das raízes  $\mathbf{x}_{1,r}, \mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$  são atualizadas para  $\Pi_1^*, \Pi_5^*$  e  $\Pi_9^*$ , respectivamente. 63
- Figura 16: Ilustração da representação hierárquica (com  $L = 4$  decomposições) da região do céu no domínio espacial da imagem.  $I^0$  é a imagem original.  $I^1$  é a imagem obtida pelo método de inicialização *SLIC*.  $I^2$  é originada unindo as regiões de  $I^1$ , e  $I^3$  é a  $L$ -ésima decomposição. Ponto pretos indicam nodos raízes no domínio espacial para cada nível de hierarquia (i.e. regiões  $R_{z,r}^{s+1}$  em  $I^{s+1}$ ) associadas ao vetores característicos  $\mathbf{x}_{z,r}^s$  que são nodos raízes no espaço de feições) as linhas verdes representam os caminhos diretos entre os nodos. Note que regiões da imagem  $I^{s+1}$  são ancestrais de regiões na imagem  $I^s$ . Contornos das regiões são marcados em vermelho. . . . . 66
- Figura 17: Exemplo de resultado utilizando três tipo de técnicas de inicialização diferentes. Em (a) apresenta-se a imagem original. Em (b) apresenta-se a segmentação obtida mediante a aplicação do algoritmo *SLIC* de *superpixels*. Em (c) apresenta-se a segmentação obtida mediante a aplicação do algoritmo de *superpixels* proposto por Malik. Em (d) apresenta-se a segmentação obtida mediante a aplicação do método Mean Shift. . . . . 72

Figura 18:	Efeito visual da inicialização no método <i>ABT</i> : o algoritmo proposto por Mori, a abordagem <i>Mean Shift</i> e o método de inicialização <i>SLIC</i> são apresentados em (a), (b) e (c), respectivamente. A segmentação final obtida pelo método <i>ABT</i> para o algoritmo proposto por Mori, a abordagem <i>Mean Shift</i> e o método <i>SLIC</i> são apresentadas em (d), (e) e (f), respectivamente (contornos das regiões em vermelho). . . . .	74
Figura 19:	Efeito visual do aumento do parâmetro $\Phi$ . Foram configurados os parâmetros: <i>Mean Shift</i> como o método de inicialização; a média dos canais de cores $L^*a^*b^*$ como o vetor característico; e a norma Euclideana como a medida de similaridade: (a) Segmentação com $\Phi = 0.1$ , (b) Segmentação com $\Phi = 0.5$ , (c) Segmentação com $\Phi = 1$ (contornos das regiões em vermelho). . . . .	74
Figura 20:	Valores médios de <i>PRI</i> computados para o banco de dados BSD300 para todas as possíveis configurações testadas em função do parâmetro limiar $\Phi \in \{0.1, \dots, 1.2\}$ . BLMS: Bhattacharyya- $L^*a^*b^*$ - <i>Mean Shift</i> ; ELMS: Euclidean- $L^*a^*b^*$ - <i>Mean Shift</i> ; BRMS: Bhattacharyya- $R^*G^*B^*$ - <i>Mean Shift</i> ; ERMS: Euclidean- $R^*G^*B^*$ - <i>Mean Shift</i> ; BHMS: Bhattacharyya-RGB Histogram- <i>Mean Shift</i> ; EHMS: Euclidean-RGB Histogram- <i>Mean Shift</i> ; BLM: Bhattacharyya- $L^*a^*b^*$ -Mori; ELM: Euclidean- $L^*a^*b^*$ -Mori; BRM: Bhattacharyya- $R^*G^*B^*$ -Mori; ERM: Euclidean- $R^*G^*B^*$ -Mori; BHM: Bhattacharyya- $R^*G^*B^*$ Histogram-Mori; EHM: Euclidean- $R^*G^*B^*$ Histogram-Mori; BLS: Bhattacharyya- $L^*a^*b^*$ - <i>SLIC</i> ; ELS: Euclidean- $L^*a^*b^*$ - <i>SLIC</i> ; BRS: Bhattacharyya- $R^*G^*B^*$ - <i>SLIC</i> ; ERS: Euclidean- $R^*G^*B^*$ - <i>SLIC</i> ; BHS: Bhattacharyya- $R^*G^*B^*$ Histogram- <i>SLIC</i> ; EHS: Euclidean- $R^*G^*B^*$ Histogram- <i>SLIC</i> . . . . .	79
Figura 21:	Comportamento da média da medida <i>PRI</i> , para o banco de dados BSD300, em função de $\Phi \in \{0.1, \dots, 1.2\}$ . Os parâmetros do método <i>Mean Shift</i> foram testados para os intervalos $hs \in [7, 9]$ e $hr \in [10, 15, 20]$ . Os demais parâmetros foram fixados como: a média da cor $L^*a^*b^*$ como vetor característico $x_j$ e como medida de distância a norma Euclideana. . . . .	79
Figura 22:	Valores médios de <i>PRI</i> em função do limiar $\Phi \in \{0.1, \dots, 1.2\}$ para a configuração: <i>Mean Shift</i> como o método de inicialização, a média dos canais de cor $L^*a^*b^*$ como o vetor característico e a norma Euclideana como a medida de distância entre pontos. . . . .	80
Figura 23:	Valores médios de <i>VOI</i> em função do limiar $\Phi \in \{0.1, \dots, 1.2\}$ para a configuração: <i>Mean Shift</i> como o método de inicialização, a média dos canais de cor $L^*a^*b^*$ como o vetor característico e a norma Euclideana como a medida de distância entre pontos. . . . .	81
Figura 24:	Valores médios de <i>GCE</i> em função do limiar $\Phi \in \{0.1, \dots, 1.2\}$ para a configuração: <i>Mean Shift</i> como o método de inicialização, a média dos canais de cor $L^*a^*b^*$ como o vetor característico e a norma Euclideana como a medida de distância entre pontos. . . . .	81
Figura 25:	Valores médios de <i>BDE</i> em função do limiar $\Phi \in \{0.1, \dots, 1.2\}$ para a configuração: <i>Mean Shift</i> como o método de inicialização, a média dos canais de cor $L^*a^*b^*$ como o vetor característico e a norma Euclideana como a medida de distância entre pontos. . . . .	81

Figura 26:	Resultados de segmentação obtidos pelo método <i>ABT</i> para diferentes imagens do banco de dados BSD300 (contornos das regiões em vermelho): em (a), são mostradas as segmentações obtidas para imagens que apresentam construções diversas tais como casas, prédios e igrejas; em (b), são mostradas segmentações obtidas para imagens que apresentam algumas paisagens naturais, tais como lagos, montanhas e pedras; em (c), são mostradas segmentações obtidas para imagens que contém animais, tais como bois, um coala, e cavalos. . . . .	83
Figura 27:	a) Comparação qualitativa do método proposto (e) em relação ao métodos: (a) SFSBM (MIGNOTTE, 2014a); (b) CTM (YANG et al., 2008); (c) TBES (MOBAHI et al., 2011); (d) <i>Mean Shift</i> (COMANICIU; MEER, 2002). . . . .	85
Figura 28:	Ilustração da segmentação de faces de três sujeitos diferentes do banco de dados FEI (THOMAZ; GIRALDI, 2010) (as curvas vermelhas indicam os contornos das regiões segmentadas). . . . .	95

## LISTA DE TABELAS

Tabela 1:	Níveis dos Fatores Controláveis. . . . .	76
Tabela 2:	Tabela ANOVA para o projeto de experimentos realizado. . . . .	77
Tabela 3:	Comparativo dos resultados obtidos com o método <i>ABT</i> , para o banco de dados BSD300, sob a configuração: <i>Mean Shift</i> como o método de inicialização, a média dos canais de cor $L^*a^*b^*$ como o vetor característico e a norma Euclideana como a medida de distância entre pontos em função do limiar $\Phi \in \{0.1, \dots, 1.2\}$ variando. . . . .	82
Tabela 4:	Valores médios das medidas de performance <i>PRI</i> , <i>VOI</i> , <i>GCE</i> e <i>BDE</i> , para o banco de dados BSD300. . . . .	84

## LISTA DE ABREVIATURAS

<b>ABT</b>	Árvores Bayesianas Adaptativas (do inglês, <i>Adaptive Bayesian Trees</i> )
<b>ANOVA</b>	Análise de variâncias
<b>BDE</b>	<i>Boundary Displacement Error</i>
<b>BSD300</b>	banco de dados de segmentação Berkeley (do inglês, <i>Berkeley Segmentation Dataset</i> )
<b>FCM</b>	<i>Fuzzy C-means</i>
<b>GCE</b>	<i>Global Consistency Error</i>
<b>GMM</b>	Modelo de Misturas de Gaussianas (do inglês, <i>Gaussian Mixture Model</i> )
<b>KNN</b>	K-vizinhos mais próximos (do inglês, <i>K-Nearest Neighbors</i> )
<b>LRE</b>	<i>Local Consistency Error</i>
<b>MAP</b>	estimativa da máxima probabilidade a <i>posteriori</i> (do inglês, <i>maximum a posteriori probability</i> )
<b>MRF</b>	Campos Aleatórios de Markov (do inglês, <i>Markov Random Fields</i> )
<b>pixels</b>	elementos da imagem (do inglês, <i>picture elements</i> )
<b>PRI</b>	<i>Probabilistic Rand Index</i>
<b>RGB</b>	vermelho, verde e azul (do inglês, <i>Red, Green and Blue</i> )
<b>SLIC</b>	<i>Simple Linear Iterative Clustering</i>
<b>SVM</b>	Máquinas de Vetores de Suporte (do inglês, <i>Support Vector Machines</i> )
<b>VOI</b>	<i>Variation of Information</i>

## LISTA DE SÍMBOLOS

$I$	imagem digital
$(m, n)$	coordenadas de localização de um <i>pixel</i> na imagem $I$
$\mathbf{x}_i$	vetor característico $i$ no <i>pixel</i> $(m, n)$ da imagem $I$
$M$	número de linhas na grade da imagem $I$
$N$	número de colunas na grade da imagem $I$
$D$	dimensionalidade do vetor característico
$R_j$	região $j$ de uma imagem $I$
$X$	conjunto de vetores característicos
$\mathcal{F}(\cdot)$	classificador
$\Omega$	conjunto de rótulos
$\omega$	rótulos de $\Omega$
$Q$	número total de rótulos em $\Omega$
$K$	número total de <i>clusters</i> encontrados pelo algoritmo <i>K-means</i>
$\bar{r}$	indicador binário
$\bar{V}$	medida de distorção
$\bar{\mu}$	centróide do cluster originado pelo algoritmo <i>K-means</i>
$\ \cdot\ $	distância Euclideana
$\hat{f}(\cdot)$	estimador de densidade
$\hat{k}(\cdot)$	função perfil de kernel
$\hat{\mathbf{x}}$	centro do kernel (janela)
$\hat{h}$	largura de banda
$x$	variável contínua
$\hat{c}_{\hat{k}, D}$	constante de normalização da função perfil de kernel
$\nabla$	operador gradiente
$\hat{g}(\cdot)$	derivada da função perfil de kernel
$\mathbf{m}_{\hat{h}, \hat{g}}$	vetor <i>mean shift</i>

$L$	componente de luminância do espaço de cores $L^*a^*b^*$
$a$	componente cromática do espaço de cores $L^*a^*b^*$
$b$	componente cromática do espaço de cores $L^*a^*b^*$
$A$	evento aleatório contido em $U$
$B$	evento aleatório contido em $U$
$U$	universo de eventos
$\text{card}(\cdot)$	função cardinalidade
$p(\cdot   \cdot)$	probabilidade condicional
$p(\cdot)$	probabilidade marginal
$\Lambda$	conjunto de possíveis ações
$\alpha$	ações de $\Lambda$
$\lambda(\cdot   \cdot)$	função custo
$R(\cdot   \cdot)$	risco condicional
$y$	variável aleatória contínua de valor real arbitrário
$\mu$	média de uma Gaussiana unidimensional
$\sigma^2$	variância de uma Gaussiana unidimensional
$\sigma$	desvio-padrão de uma Gaussiana unidimensional
$\mathcal{N}(\cdot; \cdot)$	distribuição Gaussiana
$\mathbf{y}$	vetor aleatório contínuo de valor real arbitrário
$\boldsymbol{\mu}$	média de uma Gaussiana multivariada
$\mathbb{I}$	matriz identidade
$\Sigma$	matriz de covariância
$ \cdot $	determinante
$G(\cdot)$	mistura de Gaussianas multivariadas
$\Pi$	fator de ponderação de uma componente de uma mistura de Gaussianas multivariadas
$V$	conjunto de vértices/nodos de um grafo
$v$	vértices/nodos de $V$
$E$	conjunto de arestas de um grafo
$e$	arestas de $E$
$\hat{P}(v)$	pai de um vértice/nodo $v$ de uma árvore direcionada
$v_r$	vértice/nodo raiz de uma árvore direcionada
$\mathcal{I}$	parâmetro de inicialização
$\mathcal{D}$	parâmetro associado a medida de distância entre pontos de dados

$\mathcal{FV}$	parâmetro associado ao procedimento de extração de feições
$\Phi$	parâmetro que controla o limiar de similaridade
$\eta_{\Phi}^j$	conjunto de vizinhos de $\mathbf{x}_j$ no espaço de feições
$\eta_{\Phi}$	conjunto de todos os $\eta_{\Phi}^j$
$\mathcal{G}$	grafo de regiões adjacentes
$\bar{V}$	conjunto de vértices/nodos de $\mathcal{G}$
$\bar{v}$	vértice/nodo pertencente a $\bar{V}$
$\bar{E}$	conjunto de arestas de $\mathcal{G}$
$\bar{e}$	aresta pertencente a $\bar{E}$
$\mathcal{E}^j$	conjunto de vetores característicos associados aos vizinhos espaciais da região $R_j$ delimitados por $\mathcal{G}$
$\mathcal{E}$	conjunto de todos os $\mathcal{E}^j$
$\mathbf{x}_r$	nodo raiz de uma árvore direcionada
$\mathbf{x}_{j,r}$	vetor característico $\mathbf{x}_j$ correspondente ao nodo raiz $\mathbf{x}_r$
$\hat{\mathbf{x}}_{k,r}$	vetor característico $\mathbf{x}_k$ estimado correspondente ao nodo raiz $\mathbf{x}_r$
$\eta_{\Phi}^{j,s}$	conjunto de vizinhos de $\mathbf{x}_j^s$ no espaço de feições no nível de hierarquia $s$ .
$s$	nível de hierarquia
$I^s$	imagem $I$ no nível de hierarquia $s$
$L$	número total de níveis de hierarquia $s$
$R_j^s$	região $R_j$ no nível de hierarquia $s$
$\mathbf{x}_j^s$	vetor característico associado a região $R_j^s$ no nível de hierarquia $s$
$X^s$	conjunto de vetores característicos $\mathbf{x}_j^s$ no nível de hierarquia $s$
$\mathbf{x}_{j,r}^s$	vetor característico correspondente ao nodo raiz $\mathbf{x}_r^s$ no espaço de feições no nível de hierarquia $s$
$R_j^s$	região correspondente ao nodo raiz no domínio espacial da imagem associado ao vetor característico $\mathbf{x}_{j,r}^s$
$S_{teste}$	segmentação teste obtida por um método/algoritmo
$S_k$	segmentação manual considerada o padrão ouro para uma base de dados
$PRI(.,.)$	função <i>Probabilistic Rand Index</i>
$\hat{T}(\cdot)$	operador binário
$\bar{c}$	indicador de probabilidade entre <i>pixels</i> de $S_{teste}$
$\bar{p}$	indicador de probabilidade entre <i>pixels</i> de $S_k$
$VOI(.,.)$	<i>Variation of Information</i>
$H(\cdot)$	entropia
$MI(.,.)$	variação mútua

$C(S, x_i)$	conjunto dos <i>pixels</i> que formam o segmento $S$ que contém $x_i$
$LRE(., ., .)$	função <i>Local Refinement Error</i>
$GCE(., .)$	função <i>Global Consistency Error</i>
$B_{teste}$	conjunto de <i>pixels</i> localizados nas bordas da segmentação $S_{teste}$
$B_k$	conjunto de <i>pixels</i> localizados nas bordas da segmentação $S_k$
$D_{B_2}^{B_1}$	distribuição de distâncias partindo de $B_2$ para $B_1$
$\mu_{D_{B_2}^{B_1}}$	média da distribuição $D_{B_2}^{B_1}$
$\hat{d}_E(., .)$	distância Euclideana
$\hat{d}_B(., .)$	distância Battacharyya
$BDE(., .)$	função <i>Boundary Displacement Error</i>

# 1 INTRODUÇÃO

De forma geral, uma imagem digital pode ser representada por uma função discreta  $I(m, n) = x_i$ , cujas coordenadas  $(m, n)$  identificam um ponto na imagem denominado de *pixel* (GONZALEZ; WOODS, 2002). Uma imagem  $I$  possui  $M \times N$  *pixels*, de modo que  $m = 1, 2, \dots, M$  e  $n = 1, 2, \dots, N$ . Nesta representação, o valor de  $x_i \in \mathbb{R}^D$  no *pixel* identificado pelas coordenadas  $(m, n)$  corresponde a uma característica da imagem  $I$  cujo significado físico depende do contexto em que essa imagem é adquirida. Por exemplo: em uma imagem de ressonância magnética,  $x_i$  corresponde ao valor do sinal  $i$  associado à magnetização de um determinado volume; em uma imagem de raios-X digital,  $x_i$  é proporcional à atenuação dos fótons gerados por uma fonte emissora (GONZALEZ; WOODS, 2002).

Para o sistema visual humano, uma imagem não é apenas um conjunto arbitrário de *pixels*, mas sim, um arranjo significativo de regiões e objetos. A Figura 1 mostra um conjunto de imagens. Apesar da grande variabilidade de informações contidas nestas imagens, de forma geral, pode-se argumentar que seres humanos não tem problemas em interpretá-las: em (a), identifica-se um avião que voa sob um céu nublado, em (b), identificam-se três ursos na beira de um córrego e em (c), identifica-se um homem que trabalha. Mesmo que algum aspecto interpretativo deste conjunto de imagens tenha sido omitido, é razoável supor que seres humanos possam concordar acerca dos diferentes objetos, animais, cenários contidos nestas imagens.

No início do século 20, a qualidade de agrupamento do sistema visual humano, foi extensivamente estudada por psicólogos seguidores da escola *Gestalt* (MALIK et al., 2001).

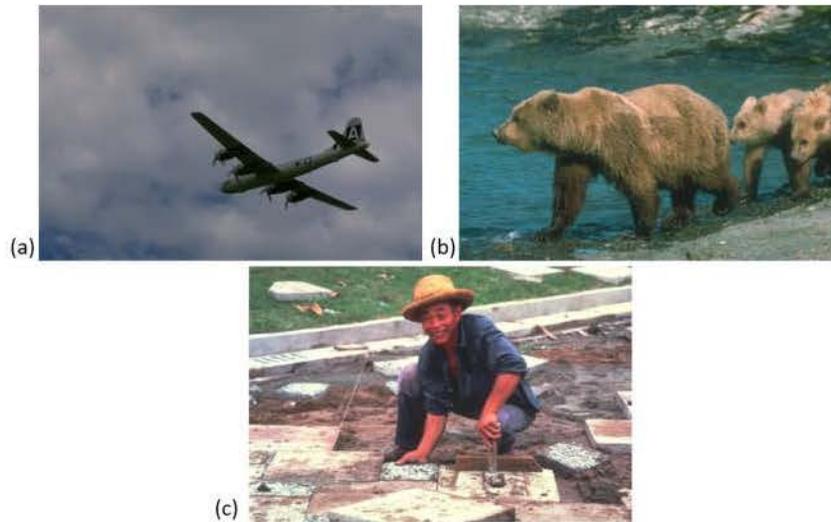


Figura 1: Conjunto de imagens exemplo (MARTIN et al., 2001).

Eles identificaram diversos fatores relacionados a percepção visual humana de agrupamento: similaridade, proximidade, continuidade, simetria, paralelismo, a necessidade por fechamento, e familiaridade. Em âmbito de visão computacional, tais fatores vem sendo usados como guia para o estudo de muitos algoritmos de agrupamento e, em especial, para a investigação do problema de segmentação de imagens (MALIK et al., 2001).

A segmentação de imagens consiste em uma etapa fundamental para aplicações em visão computacional (PAL; PAL, 1993). Por exemplo: um carro autônomo deverá segmentar as imagens adquiridas por suas câmeras para saber se é possível fazer um determinado retorno em uma via. Em medicina, para realizar a quantificação de volume de tumores, tipicamente faz-se uso de algoritmos de segmentação aplicados às modalidades de Ressonância Magnética e Tomografia Computadorizada. Em sensoriamento remoto, é possível estimar, via segmentação seguida da operação de subtração, se houve ou não alteração em uma determinada região rural de um ano para outro. Posto que a segmentação representa uma tarefa inicial em um sistema visual, o resultado das etapas consecutivas de processamento dessas aplicações são altamente dependentes da qualidade inicial da segmentação (PAL; PAL, 1993).

A segmentação consiste no processo de particionamento de uma imagem colorida  $I$  em regiões que não se interseccionam, tal que cada região é homogênea e a intersecção

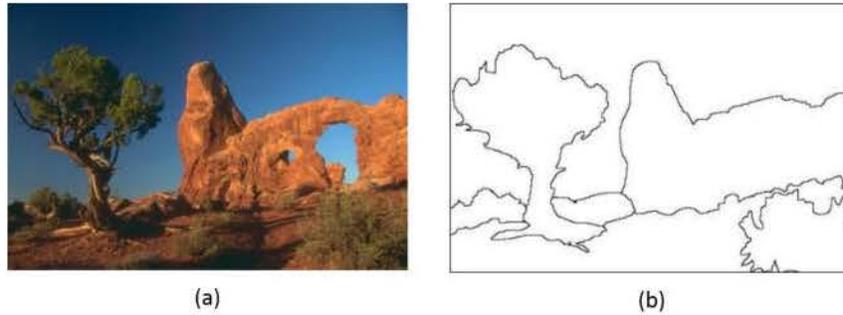


Figura 2: Exemplo de segmentação de uma imagem colorida. Em (a) apresenta-se a imagem original e em (b) apresenta-se a respectiva segmentação delimitada por contornos. Na imagem original o número de segmentos (partições) é igual ao número de *pixels* da imagem. Na imagem segmentada o número de segmentos é igual a nove (MARTIN et al., 2001).

de duas regiões distintas deve ser vazia. A Figura 2 mostra um exemplo de segmentação para uma imagem colorida  $I$ . Nota-se que na imagem original (Figura 2 (a)) o número de segmentos (partições) é igual ao número de *pixels* da imagem, enquanto que o número de partições da imagem segmentada (Figura 2 (b)) é igual a nove.

Formalmente, pode-se definir o problema de segmentação de imagens como um processo de particionamento da imagem  $I$  em um conjunto de regiões  $\{R_1, R_2, \dots, R_J\}$ , tal que:

$$\bigcup_{j=1}^{J \leq M \times N} R_j = I \text{ com } R_{j'} \cap R_j = \emptyset, j' \neq j. \quad (1)$$

De forma geral, como já exposto em parágrafos anteriores, embora seja razoável supor que seres humanos consigam concordar acerca dos diferentes objetos em uma cena, a segmentação de uma imagem tende a ser uma tarefa subjetiva (UNNIKRIHSHNAN; PANTOFARU; HEBERT, 2007). A Figura 3 apresenta um exemplo retirado do banco de dados BSD300 (MARTIN et al., 2001). Neste banco de dados, diferentes sujeitos realizaram a tarefa de segmentar imagens coloridas de forma manual. Na Figura 3 observa-se que para a mesma imagem original (a) do avião, há cinco diferentes segmentações manuais, respectivamente (b), (c), (d), (e) e (f). A tarefa da segmentação é considerada, portanto, como um problema mal definido, uma vez que há muitas possíveis soluções válidas para uma mesma imagem original (UNNIKRIHSHNAN; PANTOFARU; HEBERT, 2007). Nota-se, no entanto, que embora cada segmentação tenda a ser única (com mais ou menos detalhes

em destaque), o avião (objeto saliente) esta presente em todas as imagens segmentadas manualmente. Tal observação permite que algoritmos de segmentação de imagens possam ser avaliados de forma qualitativa, e também quantitativa, como argumentaremos nos próximos Capítulos.

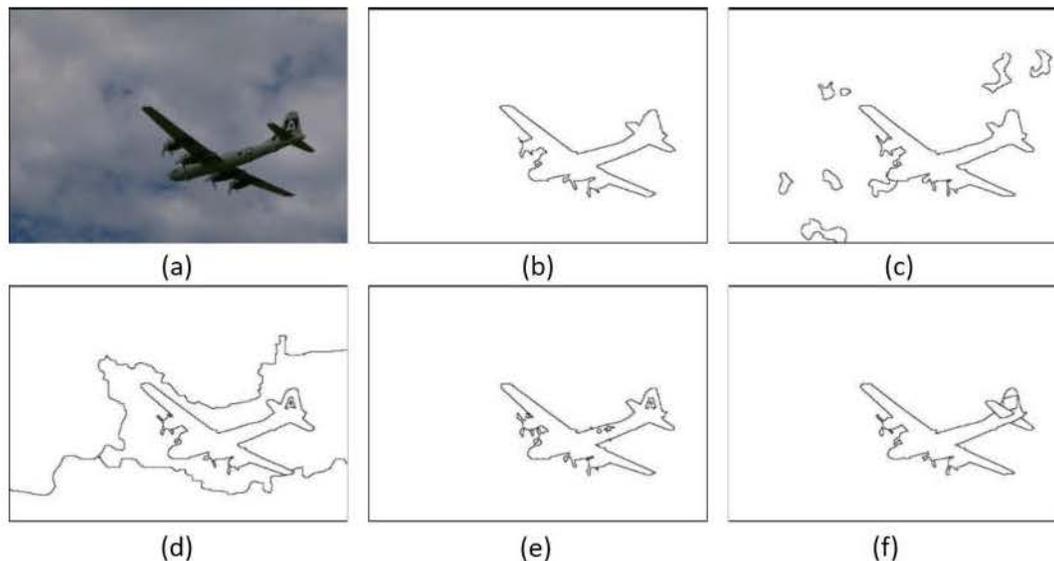


Figura 3: Exemplos de segmentação manual (MARTIN et al., 2001).

Neste trabalho, nós apresentamos um novo algoritmo de *clustering* e nós o utilizamos para resolver o problema de segmentação de imagens coloridas. Nós modificamos o método baseado em teoria de grafos proposto por (KOONTZ; NARENDRA; FUKUNAGA, 1976), de procura de máximos de densidade de uma distribuição de dados discreta, desenvolvendo uma formulação Bayesiana para modelar e detectar os máximos de densidade de uma distribuição de características (e.g. cores) de uma imagem colorida. Uma representação hierárquica de uma distribuição de características de uma imagem é obtida baseada na hierarquia de árvores direcionadas. A seguir, nodos da estrutura hierárquica de árvores direcionadas são unidos iterativamente, formando agrupamentos (do inglês, *clusters*) de características de uma imagem e as probabilidades *a priori* associadas a estes agrupamentos são atualizadas iterativamente até a convergência do algoritmo proposto.

Nossa proposta hierárquica baseada em árvores direcionadas representa um espaço de feições (espaço de características) com diferentes níveis de detalhes em cada nível hierárquico. As características representadas em níveis com maiores detalhes (i.e. nos ní-

veis mais baixos da hierarquia) estão associados a características representadas em níveis com menores detalhes (i.e. nos níveis mais altos da hierarquia), originando um aninhamento de nodos do tipo raiz de árvores direcionadas em diferentes níveis de hierarquia. Nós exploramos configurações adequadas para utilizar o método proposto no problema de segmentação de imagens coloridas, incluindo extração de características, uma etapa de inicialização e medidas de similaridade.

As principais contribuições deste trabalho são enumeradas a seguir:

1. Uma nova abordagem Bayesiana para procura de máximos de densidade de uma distribuição de dados é proposta, na qual a probabilidade *a posteriori* associada a cada amostra dos dados é computada utilizando uma mistura de Gaussianas multivariadas;
2. Com base no esquema proposto, vetores de dados distintos serão associados a diferentes *clusters* descritos por misturas de Gaussianas com diferentes números de componentes Gaussianas multivariadas. Sendo assim, nosso método não limita o número de componentes Gaussianas em uma mistura representando um *cluster*, tal como em métodos típicos de estimação de misturas de Gaussianas;
3. Nós propomos também um procedimento iterativo para construção de *clusters*, originando uma estrutura hierárquica para representação de características de uma imagem;
4. Aplicando a técnica de *clustering* desenvolvida para resolver o problema de segmentação de imagens coloridas, uma abordagem não supervisionada para segmentação de imagens é proposta, com grande potencial para superar diversas técnicas do estado-da-arte em termos de medidas de desempenho tipicamente utilizadas dentro da área.

Esta dissertação organiza-se nos Capítulos a seguir. No Capítulo 2, nós apresentamos uma breve introdução aos fundamentos de reconhecimento de padrões, definindo alguns conceitos, tais como o vetor característico e a definição do que é um algoritmo de *clustering*. No Capítulo 3, nós revisamos a literatura disponível usada para resolver o problema

de segmentação de imagens, incluindo métodos historicamente relevantes e métodos recentemente publicados. No Capítulo 4, nós apresentamos a fundamentação teórica utilizada para construção da abordagem proposta. No Capítulo 5, nós apresentamos o método proposto das Árvores Bayesianas Adaptativas aplicado ao problema de segmentação de imagens coloridas. No Capítulo 6, nós avaliamos o método proposto utilizando imagens coloridas do banco público de dados BSD300 (MARTIN et al., 2001), realizando uma variedade de experimentos diferentes. Nos Capítulos finais 7 e 8, nós discutimos os nossos resultados e apresentamos as nossas conclusões.

## 2 CONCEITOS BÁSICOS EM RECONHECIMENTO DE PADRÕES

Neste Capítulo iremos convencionar termos recorrentes que serão de fundamental importância para o entendimento dos demais Capítulos desta dissertação. A notação matemática aqui definida persistirá por todo o documento de modo a evitar confusões e equívocos de interpretação.

### 2.1 Representação de Dados de uma Imagem

Dada uma imagem  $I$  com  $M \times N$  *pixels*, e sabendo que  $I(m, n) = \mathbf{x}_i$  é um vetor característico (e.g. cor)  $i$  no *pixel*  $(m, n)$ , podemos escrever uma coleção de  $M \times N$  vetores característicos  $\mathbf{x}_i$ , como o conjunto  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{M \times N} \mid \mathbf{x}_i \in \mathbb{R}^D\}$ , com  $m = 1, 2, \dots, M$ ,  $n = 1, 2, \dots, N$  e  $i = 1, 2, \dots, M \times N$ . Esta notação indica que há um vetor característico  $\mathbf{x}_i$  associado a cada *pixel* localizado nas coordenadas  $(m, n)$  na imagem  $I$ .

O vetor característico  $\mathbf{x}_i$  corresponde a um conjunto de atributos/feições (do inglês, *features*) que descreve um padrão para um *pixel* (amostra) observado<sup>1</sup>. Cada componente do vetor  $\mathbf{x}_i$  representa uma característica discriminante diferente para um dado *pixel*. Uma vez construído o vetor  $\mathbf{x}_i$ , o *pixel* observado poderá ser representado como um ponto (vetor) em um espaço vetorial de  $D$  dimensões. Tal espaço é denominado espaço de feições (do inglês, *feature space*).

<sup>1</sup>Nesta dissertação os padrões utilizados correspondem a vetores característicos, mas de forma geral não há restrição para a definição de padrões, podendo corresponder a séries temporais, grafos, valores ordinários e etc (FRIEDMAN; HASTIE; TIBSHIRANI, 2001).

A Figura 4 mostra um exemplo representando um espaço de feições de forma gráfica. Para cada *pixel* da imagem colorida em (a), podemos construir um vetor característico  $\mathbf{x}_i = [R_i^*, G_i^*, B_i^*]$ , onde  $R_i^*, G_i^*, B_i^*$  correspondem as componentes de cor: vermelho ( $R^*$  do inglês, *Red*), verde ( $G^*$  do inglês, *Green*) e azul ( $B^*$  do inglês, *Blue*) associados ao  $i$ -ésimo vetor característico da imagem  $I$  no *pixel*  $(m, n)$ . Neste caso, o conjunto de vetores característicos  $X$  possuirá a informação dos canais de cor RGB de todos os  $M \times N$  vetores que forem observados na imagem  $I$ . Uma vez que este espaço de cores é um espaço de 3 dimensões, pode-se representá-lo graficamente tal como mostrado no ítem (b) da Figura 4. Como se observa, os vetores característicos  $\mathbf{x}_i$  associados a cada *pixel* correspondem a pontos com coordenadas  $[R_i^*, G_i^*, B_i^*]$  neste espaço de feições.

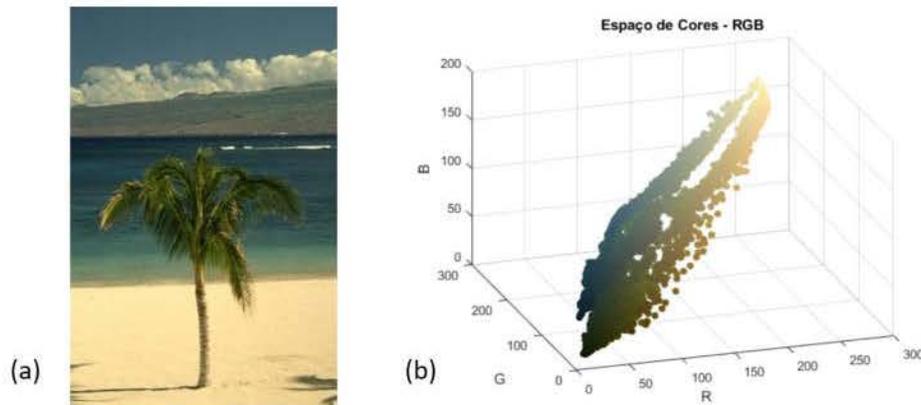


Figura 4: Representação do espaço de feições (b) contendo todos os *pixels* de uma imagem colorida (a) (MARTIN et al., 2001).

A tarefa de reconhecer padrões de forma artificial, por meio da computação digital, refere-se a ação de adivinhar ou prever a natureza desconhecida de um padrão observado (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). A natureza desconhecida de um padrão observado é denominado de classe<sup>3</sup>. De forma geral, em reconhecimento de padrões busca-se construir uma função  $\mathcal{F} : X \rightarrow \Omega$  que mapeia o conjunto de vetores característicos  $X = \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathbb{R}^D\}$ , com  $i = 1, \dots, M \times N$ , para o conjunto de classes/rótulos (do inglês, *labels*)  $\Omega = \{\omega_1, \omega_2, \dots, \omega_Q\}$ . A função mapeadora  $\mathcal{F}$  é denominada de clas-

<sup>2</sup>Denotamos as componentes de cor  $R^*$ ,  $G^*$  e  $B^*$  do espaço de cores RGB com um \* de modo a discriminar uma região  $R$  da componente de cor  $R^*$ .

<sup>3</sup>Para tarefas de classificação não-supervisionadas, a natureza de um padrão observado corresponde a um agrupamento (do inglês, *cluster*).

sificador (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). Dado um vetor característico  $\mathbf{x}_i$ ,

$$\forall \mathbf{x}_i \in X, \quad \exists \omega_k \in \Omega \quad \text{tal que} \quad \mathcal{F}(\mathbf{x}_i) = \omega_k.$$

De modo a ilustrar os conceitos acima introduzidos, propõem-se o seguinte exemplo: solicita-se que um sistema de classificação automático encontre os rótulos dos *pixels* da Figura 5 desconsiderando as bordas da imagem.

A tarefa do classificador neste exemplo, consiste em dividir o espaço de feições em regiões que correspondam a classe dos *pixels* que se encontram no círculo (que previamente sabemos ser identificado pelo rótulo  $\omega_2$ ), e no retângulo (que previamente sabemos ser identificado pelo rótulo  $\omega_1$ ). Assumindo que não há ruído algum na imagem da Figura 5, uma solução imediata para este problema pode ser construída extraindo as características de cor RGB dos *pixels* da Figura 5. Definindo um vetor característico associado a cada *pixel* da Figura 5 como  $\mathbf{x}_i = [R_i^*, G_i^*, B_i^*]$ , e sabendo que a cor branca neste espaço de cores ocorre quando  $[R^* = 255, G^* = 255, B^* = 255]$ , pode-se encontrar o rótulo correspondente a cada padrão  $\mathbf{x}_i$ , com  $i = 1, 2, \dots, M \times N$ , por meio da expressão:

$$\mathcal{F}(\mathbf{x}_i) = \begin{cases} \omega_1, & \text{se } \mathbf{x}_i = [255, 255, 255], \\ \omega_2, & \text{caso contrário.} \end{cases}$$

Logo vê-se que o modo como um classificador  $\mathcal{F}$  é construído tende a depender fortemente do problema específico que esta sendo abordado. No exemplo da Figura 5, a construção do classificador, com base no conhecimento prévio sobre o espaço de cores RGB, revela-se essencial para definir um critério de classificação adequado. Da mesma forma, especialistas podem ser consultados sobre: quais são as características (*features*) mais discriminantes para se classificar tumores como malignos ou benignos para fins de diagnóstico médico (BERNART; SCHARCANSKI; BAMPI, 2016); para fins de biometria (SOLDERA; BEHAINE; SCHARCANSKI, 2015) e para fins de representação de textura (LIU et al., 2015).

Mesmo com o auxílio de especialistas apontando conjuntos de características discri-

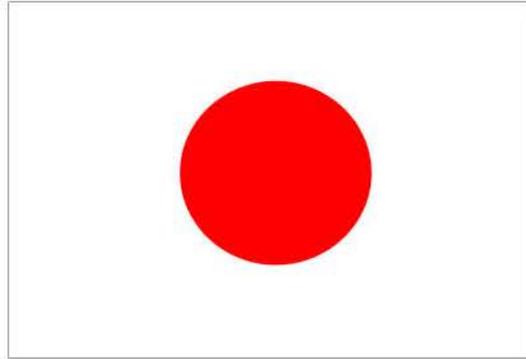


Figura 5: Bandeira do Japão utilizada como exemplo de um simples problema de classificação de *pixels*.

minantes, é importante ressaltar que erros de classificação são comuns na área de reconhecimento de padrões (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). No exemplo da Figura 5, foi assumido que a imagem não estava corrompida com ruído algum para a construção da função  $\mathcal{F}$ . Entretanto, os problemas de reconhecimento de padrões e visão computacional tipicamente lidam com dados reais que estão sujeitos a variabilidades nos sistemas de medição e extração de características, tais como variação de iluminação e sombreamento durante a aquisição de uma imagem digital por uma câmera. Sendo assim, o classificador  $\mathcal{F}$  proposto para resolver o exemplo da Figura 5, dificilmente teria o mesmo sucesso na classificação de *pixels* de imagens similares não sintéticas, originando diversos erros de classificação dos *pixels*.

Do ponto de vista estatístico o melhor classificador possível é aquele que minimiza a probabilidade do erro de classificação (FRIEDMAN; HASTIE; TIBSHIRANI, 2001). A decisão que minimiza a probabilidade do erro de classificação esta diretamente associada à regra de decisão Bayesiana que será discutido no Capítulo 4 desta dissertação.

## 2.2 Classificação Supervisionada e Não-Supervisionada

A tarefa do reconhecimento de padrões consiste em atribuir rótulos a dados observados. Isto pode ser realizado de duas formas: fazendo uso de um conjunto de dados previamente disponíveis, denominado de conjunto de dados de treinamento (classificação supervisionada); e levando em conta apenas as relações entre os dados de um conjunto de

dados teste (classificação não-supervisionada).

### 2.2.1 Classificação Supervisionada

O reconhecimento de padrões supervisionado é constituído de duas etapas: uma etapa de treinamento e uma etapa de teste. Na etapa de treinamento, o conjunto de rótulos verdadeiros  $\Omega$  de um conjunto de dados  $X$  é conhecido. Conhecendo o conjunto de rótulos verdadeiros associados a  $X$ , a função mapeadora  $\mathcal{F}$  é aprendida, isto é, constrói-se um classificador utilizando os dados de treinamento na esperança de que ele tenda a funcionar posteriormente com dados de um conjunto teste. Na etapa de teste, uma vez apresentados novos dados com rótulos de natureza não conhecida, faz-se uso da função  $\mathcal{F}$  aprendida na etapa de treino para prever o conjunto de rótulos associados aos dados de teste, buscando minimizar os erros de classificação.

Na literatura corrente, existem diversos métodos de classificação supervisionada. Alguns exemplos são: máquinas de vetores de suporte (*SVM*, do inglês *Support Vector Machines*) (CORTES; VAPNIK, 1995); K-vizinhos mais próximos (*KNN*, do inglês *K-Nearest Neighbor* (COVER; HART, 1967)), Adaboost (FREUND; SCHAPIRE, 1995). Uma discussão acerca da performance de classificadores supervisionados pode ser encontrada em (AMANCIO et al., 2014).

### 2.2.2 Classificação Não-Supervisionada

Em reconhecimento de padrões, frequentemente encontram-se tarefas para as quais dados de treinamento, contendo rótulos de classes conhecidos, não estão disponíveis. Supondo que um conjunto de dados  $X$  é conhecido, um algoritmo que agrupe os vetores característicos tal que os vetores mais similares pertençam a um mesmo agrupamento (do inglês, *cluster*) e vetores menos similares pertençam a agrupamentos diferentes, pode ser definido como um algoritmo de classificação não-supervisionada, ou também, um algoritmo de *clustering* (THEODORIDIS et al., 2010). Este tipo de tarefa é tipicamente encontrada em diversas aplicações, tais como aplicações em processamento de sinais, sensoriamento remoto, análise de dados sociais e também visão computacional (THEODORIDIS et al., 2010). Particularmente, no âmbito de visão computacional, métodos de

clustering têm sido explorados para resolver o problema de segmentação de imagens coloridas. No Capítulo 3 desta dissertação nós apresentamos uma revisão de trabalhos que vem sendo feitos explorando métodos de *clustering* aplicados a segmentação de imagens coloridas.

Uma questão importante na classificação não supervisionada reside na definição de similaridade entre os vetores característicos (THEODORIDIS; KOUTROUMBAS, 2008). Tipicamente, a similaridade é definida em função de uma medida entre os vetores característicos no espaço de feições, tal como a distância Euclideana e a distância Bhattacharyya (DUDA; HART; STORK, 2012). Ressalta-se que diferentes escolhas para medidas de similaridade, combinados ao mesmo algoritmo de *clustering*, podem implicar em diferentes resultados de agrupamento (THEODORIDIS et al., 2010).

### 3 REVISÃO DA LITERATURA

Embora desafiador, o problema de segmentar imagens coloridas tem sido extensivamente estudado ao longo dos últimos anos. De forma geral, existem duas grandes classes de abordagens para resolver o problema da segmentação de imagens: métodos baseados em regiões e métodos baseados em contornos (MALIK et al., 2001).

Métodos baseados em regiões buscam encontrar partições de uma imagem, agrupando *pixels*, em conjuntos, que partilham de propriedades discriminativas similares, tais como brilho, cor e texturas. Métodos baseados em contornos tipicamente são inicializados utilizando um detector de bordas, seguido por um processo de completamento das mesmas, afim de explorar a continuidade das curvas de um objeto (MALIK et al., 2001). Estas duas classes de métodos não necessariamente se distanciam uma da outra. Por exemplo: se um algoritmo impõe que uma borda deve ser completada, naturalmente obter-se-ão contornos de uma imagem. Uma vez que os contornos são obtidos, os conjuntos de regiões, formadas por *pixels* similares, são também obtidos. A diferença entre estas classes de métodos, reside na ênfase que se dá para obter a segmentação, isto é, algumas estratégias naturalmente cairão em um domínio ou em outro (MALIK et al., 2001).

As seções 3.1 e 3.2 deste Capítulo são dedicadas a revisão de métodos baseados em contornos e métodos baseados em regiões para abordar o problema da segmentação de imagens coloridas. A fim de estabelecer um escopo para revisão de estudos, a Seção 3.1 é dividida em duas subclasses de métodos de contorno: métodos que exploram o completamento de bordas (Seção 3.1.1) e métodos que exploram informações de escala das imagens (Seção 3.1.2). A Seção 3.2 é dividida em duas subclasses de métodos de regiões: méto-

dos de limiarização de histogramas (Seção 3.2.1) e métodos de *clustering* (Seção 3.2.2). Destacamos que uma vez que a metodologia proposta nesta dissertação (desenvolvida no Capítulo 5) enquadra-se como pertencente a classe dos métodos de *clustering*, uma maior ênfase, em relação as demais subclasses de métodos, é dada a esta subclasse de métodos na Seção 3.2.2. Na Seção final deste Capítulo (Seção 3.3), revisamos métodos de sobresegmentação publicados na literatura que estão diretamente relacionados com a metodologia desenvolvida no Capítulo 5 desta dissertação.

### 3.1 Métodos Baseados em Contornos

Métodos baseados em contornos localizam os *pixels* que correspondem as bordas dos objetos que compõem a imagem. O resultado é uma imagem binária com os *pixels* de borda detectados. Com a finalidade de detectar bordas, operadores tipicamente utilizados são os operadores Sobel (GONZALEZ; WOODS, 2002), Prewitt (GONZALEZ; WOODS, 2002) e o operador Laplaciano (GONZALEZ; WOODS, 2002). Tais procedimentos, entretanto, são tipicamente adequados para imagens sem ruído que sejam simples (não possuam objetos de formas e nem texturas variadas). A seguir, apresentamos duas sub-classes de métodos baseados em contornos os quais utilizam diferentes estratégias para segmentação de imagens coloridas: completamento de bordas e exploração de escalas de objetos em uma cena.

#### 3.1.1 Completamento de Bordas

Métodos da literatura baseados em contornos tem explorado a abordagem de completamento de bordas de forma substancial (MALIK et al., 2001; MARTIN; FOWLKES; MALIK, 2004; KIM et al., 2005; DOLLAR; TU; BELONGIE, 2006; MAIRAL et al., 2008; ARBELAEZ et al., 2009; REBOUÇAS FILHO et al., 2014). Recentemente, Arbelaez et al. (ARBELAEZ et al., 2011) exploraram uma abordagem unificada para detectar bordas e segmentar imagens utilizando atributos de textura e cor, juntamente a Histogramas de Gradientes Orientados (do inglês, *Histogram of Oriented Gradients*) (DALAL; TRIGGS, 2005).

Nestas técnicas, inicialmente um detector de borda é aplicado de forma local na imagem. Sequencialmente, estratégias globais de conexão e completamento de bordas são adotadas de forma a explorar a continuidade da curva dos objetos. Este tipo de abordagem, frequentemente, é criticada devido a forma prematura com que se decide se há ou não bordas em determinadas partes da imagem (MALIK et al., 2001). Por exemplo: para detecção de contornos em regiões de pouco contraste (onde a magnitude do gradiente é pequena), deve-se definir um limiar muito baixo para que essas bordas sejam detectadas. Definido este limiar baixo, um detector de bordas irá detectar bordas por toda a imagem, fazendo com que o processo de conexão das bordas (completamento de bordas) se torne muito complexo. Igualmente, se a imagem possuir texturas salientes em abundância, terá que se definir um limiar alto para minimizar a detecção das bordas da região saliente. Tal decisão fará com que contornos presentes em regiões de baixo contraste sejam perdidas.

### **3.1.2 Exploração da Informação de Escala**

Outra linha de trabalho em métodos baseados em contornos que vem sendo investigada, relaciona-se a variedade de escalas em que objetos em uma imagem colorida podem aparecer. Por exemplo, Ren et al. (REN; FOWLKES; MALIK, 2008) investigaram as estatísticas associadas as bordas de regiões de uma imagem colorida apresentadas em múltiplas escalas, obtendo assim, um esquema de segmentação de imagens coloridas invariante a escala. Em outra publicação (REN, 2008), operadores aplicados a características de brilho, cor e textura são combinados para múltiplas escalas. Estas informações, juntamente a características de localização espacial e contraste são utilizadas como atributos para construção de um classificador de bordas.

As abordagens acima mencionadas nas seções 3.1.1 e 3.1.2 podem detectar informações de borda para uma variedade de aplicações. Todavia, ressaltamos que métodos de segmentação baseados em contornos, tendem a se precipitar quanto a existência, ou não, de bordas em certas partes de uma imagem.

## 3.2 Métodos Baseados em Regiões

Métodos baseados em regiões prestam-se mais facilmente à definição de uma função objetiva global, podendo fazer uso de Campos Aleatórios de Markov (*MRF*, do inglês, *Markov Random Fields*) (GEMAN; GEMAN, 1984) assim como utilizar uma formulação variacional para segmentação de imagens (MUMFORD; SHAH, 1989). Uma vez proposta uma função objetiva global, decisões são tomadas apenas quando toda a informação contida na imagem é levada em conta ao mesmo tempo. Dentro da classe de métodos baseados em regiões, existem ainda diversas abordagens disponíveis na literatura, tais como métodos baseados em teoria de grafos (FELZENSZWALB; HUTTENLOCHER, 2004; VANHAMEL; PRATIKAKIS; SAHLI, 2006), abordagens de fusão estatística de regiões (WONG; SCHARCANSKI; FIEGUTH, 2011; MEDEIROS; SCHARCANSKI; WONG, 2013; VÁSQUEZ; SCHARCANSKI; WONG, 2015) e abordagens baseadas em corte de grafos (SHI; MALIK, 2000; TAO; JIN; ZHANG, 2007; DENG; DAI; ZHANG, 2011; SALAH; MITICHE; AYED, 2011; COUR; BENEZIT; SHI, 2005). A seguir destacamos duas sub-classes utilizadas para segmentação de imagens coloridas: métodos baseados em limiarização de histogramas e métodos baseados em algoritmos de *clustering*.

### 3.2.1 Limiarização de Histogramas

A segmentação por limiarização de histogramas (GONZALEZ; WOODS, 2002) corresponde ao procedimento de designar o conjunto dos *pixels* de uma imagem, como pertencente a um segundo plano (do inglês, *background*) da imagem, ou, a um primeiro plano (do inglês, *foreground*). Para tanto, uma prática comum nos procedimentos de limiarização consiste em definir algum valor correspondente a um limiar. Dependendo do critério condicional adotado, compara-se os valores dos *pixels* (intensidade, ou, cor) com o limiar e atribui-se correspondentemente valores para cada coordenada do *pixel* comparado de 0, ou, 1, sendo respectivamente pertencentes ao segundo plano, e ao primeiro plano. Quando o limiar definido é igual a um valor constante, esse procedimento é denominado limiarização global, se não, é denominado limiarização local. Métodos baseados

em limiarização global são muito dependentes da iluminação. Assim, imagens localmente saturadas tendem a apresentar segmentações ruins.

Métodos de limiarização tendem, por essa razão, a utilizar múltiplos limiares a fim de compensar a iluminação não uniforme das imagens. Ainda que simples e computacionalmente eficientes, os métodos baseados na limiarização de histograma se fazem adequados para algumas aplicações (BERNART et al., 2015; BERNART; SCHARCANSKI; BAMPI, 2016). Métodos baseados em limiarização de histogramas tendem a falhar quando as regiões de interesse são adjacentes e possuem *pixels* com níveis de cinza muito próximos. Acrescenta-se ainda que métodos tradicionais de segmentação por limiarização de histogramas, tendem a falhar quando os histogramas correspondem a distribuições unimodais.

### 3.2.2 Clustering

O uso de algoritmos de *clustering* para a segmentação de imagens também se destaca entre as abordagens disponíveis na literatura corrente. O agrupamento das características de uma imagem no espaço de feições (do inglês, *feature space*), implica na obtenção de regiões no espaço da imagem. Sendo assim, algoritmos de *clustering* classificam-se como métodos baseados em regiões.

Em métodos de *clustering* aplicados a segmentação de imagens, *pixels* são agrupados de acordo com a similaridade referente a atributos de cor, textura e luminosidade. Dentro desta subclasse, destacamos três abordagens amplamente utilizadas: *K-means* (HARTIGAN; WONG, 1979), *Fuzzy C-means* (BEZDEK, 2013), *Gaussian Mixture Model and Estimation Maximization* (FU; WANG, 2012).

O algoritmo *K-means* é uma técnica iterativa que pode ser utilizada para segmentação de imagens. O algoritmo particiona os vetores característicos de um conjunto de dados  $X$  associado a uma imagem  $I$  (com  $M \times N$  *pixels*) em  $K$  agrupamentos (*clusters*) no espaço de feições. Para cada vetor característico  $\mathbf{x}_i \in X$  define-se um indicador binário  $\bar{r}_{i\bar{k}} \in \{0, 1\}$ , onde  $\bar{k} = 1, 2, \dots, K$  descrevendo quais dos  $K$  *clusters* irá conter o vetor  $\mathbf{x}_i$ . Assume-se que  $\bar{r}_{i\bar{k}} = 1$  se o vetor  $\mathbf{x}_i$  é atribuído ao *cluster*  $\bar{k}$  e, conseqüentemente,  $\bar{r}_{i\bar{j}} = 0$ , para todo  $\bar{j} \neq \bar{k}$ . O algoritmo *K-means* minimiza a medida de distorção (do

inglês, *distortion measure*) definida por:

$$\bar{V} = \sum_{i=1}^{M \times N} \sum_{k=1}^K \bar{r}_{ik} \|\mathbf{x}_i - \bar{\boldsymbol{\mu}}_k\|^2, \quad (2)$$

que representa a soma dos quadrados das distâncias de cada vetor  $\mathbf{x}_i$  em relação ao centróide do *cluster*  $\bar{\boldsymbol{\mu}}_k$ , onde  $\|\cdot\|$  é a distância Euclideana (BISHOP, 2006). Para este método, tipicamente, faz-se uso de vetores característicos que contém informações referentes a cor, textura e localização espacial. Embora a convergência do algoritmo seja garantida de forma ótima, a solução obtida depende fortemente do número de *clusters* definido (parâmetro  $K$ ), assim como do conjunto de vetores característicos escolhidos para descrever os *pixels* da imagem.

Métodos explorando a característica de convergência em ótimos locais do algoritmo *K-means* (HARTIGAN; WONG, 1979) vêm sendo propostos por meio de abordagens híbridas (LI; HE; WEN, 2015) baseadas em *Particle Swarm Optimization* (OMRAN; ENGELBRECHT; SALMAN, 2005), bem como propostas para estimar fatores de ponderação de acordo com o grau de importância das variáveis para a composição de um dado *cluster*.

Algoritmos do tipo *clustering-fuzzy* permitem que elementos dos dados a serem analisados possam pertencer a mais de um agrupamento simultaneamente. Para isso, associado a cada ponto dos dados, define-se um conjunto de pesos de participação desses agrupamentos. Tais pesos indicam o quão fortemente ligado está um ponto dos dados e um agrupamento em particular. O algoritmo *FCM* particiona um conjunto de dados em uma coleção de *C-fuzzy* agrupamentos minimizando a variância entre os atributos de um determinado *pixel* e os atributos referentes a um *pixel* localizado no centro de um agrupamento. A diferença entre os algoritmos *K-means* e *FCM* encontra-se no conjunto de pesos de participação atribuídos aos agrupamentos.

Trabalhos incorporando o algoritmo *Fuzzy C-means (FCM)* (BEZDEK, 2013) vem sendo explorados para segmentar imagens ruidosas de forma robusta (MÚJICA-VARGAS; GALLEGOS-FUNES; ROSALES-SILVA, 2013), sendo combinados a classificadores supervisionados aplicados a características de cor e textura (WANG et al., 2012), e melhora-

dos, buscando estimar a inicialização dos centróides dos *clusters* para posterior aplicação do algoritmo *FCM* (YU et al., 2010). Ainda explorando o algoritmo *FCM*, alguns estudos (TAN; ISA, 2011; TAN; ISA; LIM, 2013) apresentam algoritmos de classificação não supervisionado adaptativos, que se inicializam por meio da obtenção de partições homogêneas originadas pela técnica de limiarização do histograma. Faz-se a fusão das regiões obtidas para posterior utilização do método *FCM*.

O método *GMM-EM*, aplicado no contexto de segmentação de imagens (FU; WANG, 2012), consiste em um algoritmo cujo objetivo é modelar os atributos dos *pixels* de uma imagem (representada por características de cores, texturas) como a combinação linear de Gaussianas multivariadas ponderadas (THEODORIDIS et al., 2010; DUDA; HART; STORK, 2012). Para encontrar os parâmetros deste modelo (média, covariância e pesos) utiliza-se o algoritmo *Expectation Maximization* (DEMPSTER; LAIRD; RUBIN, 1977). Tal algoritmo consiste em um método de otimização iterativo que se divide em duas etapas. Tipicamente, os parâmetros de média e covariância são inicializados aleatoriamente. Iterativamente o algoritmo atualiza os valores de média e covariância até que haja convergência para uma solução ótima local. Tal como o algoritmo *K-means*, o método de segmentação via *GMM-EM* não só depende fortemente do número de *clusters* definidos como parâmetro de entrada, mas também do conjunto de vetores característicos escolhidos para descrever os *pixels* da imagem.

Apesar do bom desempenho dos métodos acima mencionados, técnicas de agrupamento tradicionais tais como *K-means*, *FCM* e *GMM-EM* tendem a não ser confiáveis para estimar o número verdadeiro de *clusters* (JAIN; DUBES, 1988). Sendo assim, uma prática comum de pesquisas recentes, consiste em abordar o problema de segmentação de imagens coloridas, fundindo regiões iniciais obtidas por procedimentos de sobresegmentação, usando, por exemplo, o método *Mean Shift* (COMANICIU; MEER, 2002), ou, algoritmos de *superpixel* (ACHANTA et al., 2012; MORI et al., 2004)<sup>1</sup>. Esta proposta produz regiões locais altamente semelhantes que são unidas subsequentemente, a fim de obter uma segmentação da imagem original (REN; SHAKHNAROVICH, 2013;

---

<sup>1</sup>Na Seção 3.3 deste Capítulo, nós apresentamos os fundamentos do método *Mean Shift*, bem como os fundamentos de dois algoritmos de *superpixel* explorados nesta dissertação.

MIGNOTTE, 2014b; DONOSER et al., 2009; MOBAHI et al., 2011; KHELIFI; MIGNOTTE, 2016a,b; HUANG et al., 2011; YANG et al., 2008; WANG et al., 2016). Tal proposta também será explorada nesta dissertação quando apresentarmos o método das Árvores Bayesianas Adaptativas no Capítulo 5.

### 3.3 Trabalhos Relacionados

Nas Seções 3.2 e 3.1 desta dissertação, foram revisadas abordagens para segmentar imagens coloridas divididas em duas classes de abordagens: métodos baseados em contornos e métodos baseados em regiões. Dentro da subclasse de métodos baseados em regiões, revisamos abordagens que fazem uso de *clustering* (Seção 3.2.2), e particularmente, ressaltamos que uma prática comum de pesquisas recentes consiste em abordar o problema de segmentação de imagens coloridas mediante a fusão de regiões inicialmente obtidas por procedimentos de sobresegmentação.

Nesta dissertação a metodologia desenvolvida (Capítulo 5) segue esta proposta, e para tanto, explora o uso de métodos de sobresegmentação. A seguir revisamos o método *Mean Shift* de sobresegmentação e dois algoritmos de *superpixel* que são utilizados na metodologia desta dissertação.

#### 3.3.1 Mean Shift

O método *Mean Shift* aplicado a segmentação de imagens foi popularizado por Comaniciu e Meer (COMANICIU; MEER, 2002). Esta técnica é composta de duas etapas: a localização dos máximos de densidade do conjunto de dados de uma imagem (no espaço de feições) e uma etapa consecutiva de *clustering* dos dados. Dado um espaço de feições constituído por  $M \times N$  vetores característicos  $\mathbf{x}_i \in \mathbb{R}^D$ ,  $i = 1, 2, \dots, M \times N$ , define-se o estimador de densidade  $\hat{f}(\hat{\mathbf{x}})$  como (COMANICIU; MEER, 2002):

$$\hat{f}(\hat{\mathbf{x}}) = \frac{\hat{c}_{\hat{k},D}}{i\hat{h}^D} \sum_{i=1}^{M \times N} \hat{k}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right), \quad (3)$$

onde  $\hat{h}$  é a largura de banda,  $\hat{\mathbf{x}}$  é o centro do kernel (janela) cuja função perfil  $\hat{k}(\|x\|^2)$ , definida para  $x \geq 0$ , deve ser normalizada tal que:

$$\int_0^\infty \hat{c}_{\hat{k},D} \hat{k}(\|x\|^2) = 1, \quad (4)$$

e  $\hat{c}_{\hat{k},D}$  é uma constante que deve ser escolhida de forma a satisfazer a Equação 4.

Dada uma função  $\hat{f}(\hat{\mathbf{x}})$ , os máximos de densidade podem ser estimados por meio da computação do gradiente  $\nabla \hat{f}(\hat{\mathbf{x}})$  (COMANICIU; MEER, 2002):

$$\nabla \hat{f}(\hat{\mathbf{x}}) = \frac{2\hat{c}_{\hat{k},D}}{i\hat{h}^{D+2}} \sum_{i=1}^{M \times N} (\hat{\mathbf{x}} - \mathbf{x}_i) \hat{k}'\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right). \quad (5)$$

Definindo a função  $\hat{g}(x) = \hat{k}'(x)$ , e assumindo que a derivada de  $\hat{k}$  existe para todo  $x \in [0, \infty)$ , podemos reescrever a Equação 5:

$$\nabla \hat{f}(\hat{\mathbf{x}}) = \frac{2\hat{c}_{\hat{k},D}}{i\hat{h}^{D+2}} \left[ \sum_{i=1}^{M \times N} \hat{g}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right) \right] \left[ \frac{\sum_{i=1}^{M \times N} \mathbf{x}_i \hat{g}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right)}{\sum_{i=1}^{M \times N} \hat{g}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right)} - \hat{\mathbf{x}} \right]. \quad (6)$$

O primeiro termo da Equação 6 é proporcional ao estimador de densidade definido em 3 com a função perfil de kernel  $\hat{g}(x)$ . O segundo termo corresponde ao vetor *mean shift*  $\mathbf{m}_{\hat{h},\hat{g}}$  (COMANICIU; MEER, 2002):

$$\mathbf{m}_{\hat{h},\hat{g}} = \left[ \frac{\sum_{i=1}^{M \times N} \mathbf{x}_i \hat{g}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right)}{\sum_{i=1}^{M \times N} \hat{g}\left(\left\|\frac{\hat{\mathbf{x}} - \mathbf{x}_i}{\hat{h}}\right\|^2\right)} - \hat{\mathbf{x}} \right] \quad (7)$$

Uma vez encontrados os máximos de densidade, a etapa de *clustering* é realizada por meio da substituição dos vetores característicos pelos vetores localizados nos máximos de densidade mais próximos. Na implementação utilizada nesta dissertação, o espaço de feições é caracterizado pela localização espacial dos *pixels* (referente as coordenadas dos *pixels* em uma imagem) bem como as cores associadas aos *pixels* oriundas do espaço de cores.

### 3.3.2 Superpixels

*Superpixels* são constituídos por conjuntos de *pixels* vizinhos e similares que capturam em nível local redundâncias em uma imagem (ACHANTA et al., 2012). Uma vez que o número de *superpixels* em uma imagem tende a ser muitas vezes menor que o número total de *pixels* em uma imagem, esta etapa de pré-processamento tende a reduzir substancialmente a complexidade de diversas aplicações de visão computacional, tais como: estimação de profundidade em problemas estéreo (ZITNICK; KANG, 2007), segmentação de imagens (MOBAHI et al., 2011) e localização de objetos (FULKERSON; VEDALDI; SOATTO, 2009).

Uma comparação detalhada de diversos algoritmos de *superpixel* presentes na literatura pode ser encontrada em (ACHANTA et al., 2012). Nesta dissertação, serão investigados os algoritmos *SLIC* (do inglês, *Simple Linear Iterative Clustering*) (ACHANTA et al., 2012) e o algoritmo proposto por Mori e colaboradores (MORI et al., 2004), que são apresentados a seguir.

O algoritmo *SLIC* é baseado em uma adaptação do algoritmo *K-means* para geração de *superpixels*. No algoritmo convencional *K-means*, distâncias são computadas de cada centro de *cluster* para cada *pixel* na imagem, enquanto que no algoritmo *SLIC* se computam distâncias de cada centro de *cluster* para *pixels* delimitados por um espaço de busca pré-definido, reduzindo a complexidade desta técnica. Além disso, o algoritmo *SLIC* faz uso de uma distância ponderada que combina características de cor a informações de proximidade espacial, balanceando o controle sobre o tamanho e a compactação dos *superpixels* (ACHANTA et al., 2012).

Declarado o número de *superpixels* desejados como parâmetro de entrada, o método *SLIC* define o vetor característico  $\mathbf{x}_i = [L_i, a_i, b_i, m_i, n_i]$  para cada *pixel* de uma imagem colorida  $I$ , onde  $L, a, b$  são os canais de cores do espaço CIE  $L^*a^*b^*$ , respectivamente, e  $m$  e  $n$  correspondem às coordenadas espaciais de um dado *pixel* na grade da imagem. Dado um espaço de busca pré-delimitado, iterativamente, cada vetor  $\mathbf{x}_i$  é associado ao *cluster* mais próximo, e uma etapa de pós-processamento garante a obtenção de um mapa de *superpixels* de tamanho similar. A implementação do algoritmo

*SLIC* utilizada neste trabalho encontra-se disponível em: <http://ivrl.epfl.ch/research/superpixels>.

O algoritmo proposto por Mori e colaboradores (MORI et al., 2004)<sup>2</sup> consiste em continuação dos trabalhos de Ren e Malik (REN; MALIK, 2003). Esta abordagem faz uso do algoritmo de corte normalizado em grafos (SHI; MALIK, 2000) para obtenção de *superpixels*. Neste método, a matriz de similaridades construída para representar os grafos utiliza atributos normalizados da imagem, que se fundamentam nos princípios da escola *Gestalt*. Para tanto, os autores utilizam informações de textura e de contorno para definir o vetor característico associado a cada segmento (*superpixel*) da imagem.

---

<sup>2</sup>A implementação do algoritmo proposto por Mori utilizada neste trabalho encontra-se disponível em: <http://www.cs.sfu.ca/~mori/research/superpixels/>.

## 4 FUNDAMENTAÇÃO TEÓRICA

Este Capítulo é dedicado a introdução dos conceitos matemáticos utilizados na construção do algoritmo de *clustering* desenvolvido, que se encontra fundamentado no Capítulo 5 desta dissertação. A seguir, introduziremos: a regra de decisão Bayesiana na Seção 4.1; o critério da máxima estimativa da probabilidade a *posteriori* na Seção 4.2; o conceito de misturas de Gaussianas na Seção 4.3; e definições relevantes, para o nosso trabalho, relacionadas a teoria de grafos na Seção 4.4.

### 4.1 A Regra de Decisão Bayesiana

A regra de decisão Bayesiana proposta pelo matemático Thomas Bayes, que data o ano de 1764, tem sido amplamente utilizada nos campos de visão computacional e reconhecimento de padrões (BISHOP, 2006). A regra define que a probabilidade de ocorrência de um evento aleatório  $A$ , dada a ocorrência de um evento aleatório  $B$ , pode ser computada por meio da expressão (BISHOP, 2006):

$$p(A | B) = \frac{p(B | A)p(A)}{p(B)}, \quad (8)$$

onde  $p(B) = \sum_B p(B | A)p(A)$ . Se  $p(B) = 0$ , então  $p(A | B)$  não é definido.

A probabilidade marginal  $p(A)$ , no contexto da regra de decisão bayesiana, é denominada de probabilidade a *priori*. A probabilidade condicional  $p(A | B)$  é denominada de probabilidade a *posteriori*, e a probabilidade condicional de  $p(B | A)$  é denominada de verossimilhança. A seguir, utilizando diagramas de Venn, deduz-se a Equação 8.

A Figura 6 representa os conjuntos  $A \in U$  e  $B \in U$ . Onde  $A$  e  $B$  representam eventos

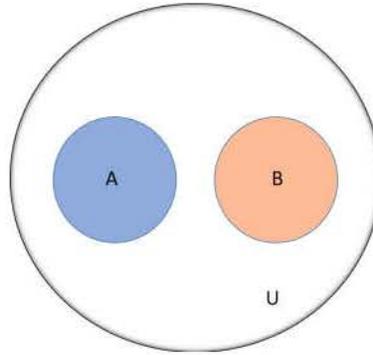


Figura 6: Diagrama de Venn representando dois possíveis eventos  $A$  e  $B$  em um universo de eventos  $U$ .

possíveis dentro de um universo de eventos  $U$ . Utilizando o diagrama de Venn, é possível verificar que a probabilidade ao acaso de um evento em  $A$  ocorrer, corresponde a razão entre a cardinalidade do conjunto  $A$  pela cardinalidade do conjunto  $U$ :

$$p(A) = \frac{\text{card}(A)}{\text{card}(U)}, \quad (9)$$

onde a função  $\text{card}(\cdot)$  representa a cardinalidade de um dado conjunto. De forma análoga, a probabilidade ao acaso associada a ocorrência de um evento em  $B$  é:

$$p(B) = \frac{\text{card}(B)}{\text{card}(U)}. \quad (10)$$

A probabilidade ao acaso associada a ocorrência de um evento que interseccione as áreas de  $A$  e  $B$  é:

$$p(A \cap B) = \frac{\text{card}(A \cap B)}{\text{card}(U)}. \quad (11)$$

Se considerado o cenário representado na Figura 7, a probabilidade condicional  $p(A | B)$  pode ser entendida como a probabilidade ao acaso associada a ocorrência do evento  $A \cap B$  na região  $B$ , isto é:

$$p(A | B) = \frac{\text{card}(A \cap B)}{\text{card}(B)}. \quad (12)$$

Dividindo o numerador e o denominador da Equação 12 por  $\text{card}(U)$ , obtemos:

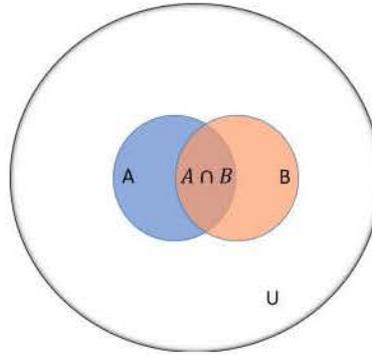


Figura 7: Diagrama de Venn representando a intersecção de dois possíveis eventos  $A$  e  $B$  em um universo de eventos  $U$ .

$$p(A | B) = \frac{\frac{\text{card}(A \cap B)}{\text{card}(U)}}{\frac{\text{card}(B)}{\text{card}(U)}} = \frac{p(A \cap B)}{p(B)}. \quad (13)$$

De forma análoga podemos escrever a probabilidade condicional  $p(B | A)$ :

$$p(B | A) = \frac{\frac{\text{card}(B \cap A)}{\text{card}(U)}}{\frac{\text{card}(A)}{\text{card}(U)}} = \frac{p(B \cap A)}{p(A)}. \quad (14)$$

Do diagrama de Venn é possível concluir que  $p(A \cap B) = p(B \cap A)$ . Igualando as expressões em 13 e 14, escrevemos:

$$p(A | B)p(B) = p(B | A)p(A), \quad (15)$$

e isolando a probabilidade condicional  $p(A | B)$  obtemos a expressão 8:

$$p(A | B) = \frac{p(B | A)p(A)}{p(B)}.$$

A probabilidade marginal  $p(B)$  garante que a soma das probabilidade a posteriori  $p(A | B)$  para todos os possíveis valores de  $A$  será igual a 1. Por esta razão, é comum interpretar  $p(B)$  como uma constante de normalização (BISHOP, 2006).

De forma genérica, no contexto de Reconhecimento de Padrões, a regra de decisão bayesiana toma a forma a seguir. Seja  $\Omega = \{\omega_1, \omega_2, \dots, \omega_Q\}$  um conjunto finito de  $Q$  rótulos e  $\Lambda = \{\alpha_1, \alpha_2, \dots, \alpha_W\}$  um conjunto finito de  $W$  possíveis ações, definimos uma função custo  $\lambda(\alpha_i | \omega_j)$  que descreve o custo sujeito à ação  $\alpha_i$  para um dado rótulo

$\omega_j$  (DUDA; HART; STORK, 2012). Considera-se que o vetor característico  $\mathbf{x}$  seja uma variável aleatória com  $D$  componentes ( $\mathbf{x} \in \mathbb{R}^D$ ), e que a probabilidade condicional  $p(\mathbf{x} | \omega_j)$ , seja a função densidade de probabilidade de  $\mathbf{x}$  condicionado a um rótulo  $\omega_j$ . Denotando  $p(\omega_j)$  como sendo a probabilidade a *priori* associada a ocorrência do rótulo  $\omega_j$ , reescreve-se a probabilidade a *posteriori*  $p(\omega_j | \mathbf{x})$  definida na Equação 8 (DUDA; HART; STORK, 2012):

$$p(\omega_j | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_j)p(\omega_j)}{p(\mathbf{x})}, \quad (16)$$

onde a probabilidade marginal  $p(\mathbf{x})$  é:

$$p(\mathbf{x}) = \sum_{j=1}^Q p(\mathbf{x} | \omega_j)p(\omega_j). \quad (17)$$

Uma vez que  $p(\omega_j | \mathbf{x})$  corresponde a probabilidade de que o rótulo verdadeiro para  $\mathbf{x}$  seja  $\omega_j$ , e assumindo que  $\omega_j$  seja de fato o rótulo verdadeiro, podemos avaliar o custo  $\lambda(\alpha_i | \omega_j)$  de uma determinada ação  $\alpha_i \in \Lambda$  dado um  $\mathbf{x}$  observado, por meio do valor esperado da função custo (DUDA; HART; STORK, 2012):

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1}^Q \lambda(\alpha_i | \omega_j)p(\omega_j | \mathbf{x}). \quad (18)$$

A função  $R(\alpha_i | \mathbf{x})$  é denominada risco condicional. Para um dado  $\mathbf{x}$  conhecido, podemos minimizar o valor esperado da função custo, selecionando a ação que minimiza a função risco condicional. Para minimizar o risco condicional, computa-se  $R(\alpha_i | \mathbf{x})$  tal como descrito na Equação 18 e seleciona-se a ação  $\alpha_i$  para a qual  $R(\alpha_i | \mathbf{x})$  é mínimo (DUDA; HART; STORK, 2012).

Em problemas de classificação, a ação  $\alpha_i$  corresponde em decidir se a verdadeira classe de  $\mathbf{x}$  corresponde a  $\omega_j$ . Se a ação  $\alpha_i$  é tomada, e a verdadeira classe de  $\mathbf{x}$  é  $\omega_j$ , então a decisão esta correta se  $i = j$ , e errada se  $i \neq j$ . Uma função custo de interesse<sup>1</sup> para este caso pode ser definida como (DUDA; HART; STORK, 2012):

<sup>1</sup>Observamos que outras funções custo podem ser definidas de modo a minimizar o risco condicional sujeitos a outras restrições. Na Seção 4.1 deste trabalho utilizamos como exemplo a função custo definida em 19, mas outras definições para função custo podem ser encontradas em (DUDA; HART; STORK, 2012).

$$\lambda(\alpha_i | \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \quad \text{com } i, j = 1, 2, \dots, Q. \quad (19)$$

A função custo, definida em 19, assinala um custo nulo para uma decisão correta e um custo unitário para uma decisão errada (todos os erros “possuem a mesma importância”).

Computando o risco condicional correspondente a essa função custo:

$$\begin{aligned} R(\alpha_i | \mathbf{x}) &= \sum_{\forall j} \lambda(\alpha_i | \omega_j) p(\omega_j | \mathbf{x}) \\ &= \sum_{j \neq i} p(\omega_j | \mathbf{x}) \\ &= 1 - p(\omega_i | \mathbf{x}). \end{aligned} \quad (20)$$

Observa-se que para minimizar o risco condicional definido em 20, deve-se selecionar o índice  $j$  que maximiza a probabilidade a *posteriori*  $p(\omega_j | \mathbf{x})$ . Para a função custo adotada em 19, a ação  $\alpha_j$  que minimiza o risco condicional, equivale a ação que atribui o rótulo  $\omega_j$  ao padrão observável  $\mathbf{x}$  quando:

$$p(\omega_j | \mathbf{x}) > p(\omega_k | \mathbf{x}), \quad \forall j \neq k. \quad (21)$$

Posto isto, a regra de decisão Bayesiana corresponde a um classificador ótimo, pois ao maximizarmos a probabilidade a *posteriori*, estaremos minimizando o valor da função risco condicional em 20 (DUDA; HART; STORK, 2012).

## 4.2 Máxima Probabilidade a *posteriori*

Suponhamos que os rótulos contidos em  $\Omega = \{\omega_1, \omega_2, \dots, \omega_Q\}$  não sejam conhecidos. É possível estimar os valores em  $\Omega$  considerando que vetores característicos  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{\bar{N}}$  sejam observados. Seja o conjunto de dados  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{\bar{N}}\}$ , rescrevemos a probabilidade a *posteriori* definida na Equação 16 (THEODORIDIS; KOUTROUMBAS, 2008) como:

$$p(\Omega | X) = \frac{p(X | \Omega)p(\Omega)}{p(X)}. \quad (22)$$

A máxima estimativa da probabilidade a *posteriori*, (*MAP*, do inglês *maximum a posteriori probability*)  $\hat{\Omega}_{MAP}$ , é definida quando  $p(\Omega | X)$  é máximo, isto é (THEODORIDIS; KOUTROUMBAS, 2008):

$$\hat{\Omega}_{MAP} = \arg \max\{p(\Omega | X)\}. \quad (23)$$

Uma condição necessária que  $\hat{\Omega}_{MAP}$  deve satisfazer, para que esta seja uma estimativa de máxima probabilidade, é de que o gradiente de  $p(\Omega | X)$  em relação a  $\Omega$  deve ser igual a zero, isto é,

$$\hat{\Omega}_{MAP} : \frac{\partial p(\Omega | X)}{\partial \Omega} = 0 \quad \text{ou} \quad \frac{\partial (p(X | \Omega)p(\Omega))}{\partial \Omega} = 0. \quad (24)$$

A formulação da máxima probabilidade a *posteriori* indica que a probabilidade de associar os padrões representados, pelo conjunto de dados  $X$ , ao seu conjunto de rótulos verdadeiros  $\Omega$ , será aquela que maximiza o valor de probabilidade a *posteriori*.

## 4.3 Distribuições Gaussianas

### 4.3.1 Distribuição Gaussiana Unidimensional

A distribuição Gaussiana, avaliada para uma única variável de valor real  $y$ , é definida pela expressão (BISHOP, 2006):

$$\mathcal{N}(y; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(y-\mu)^2/2\sigma^2}, \quad (25)$$

e é governada pelos parâmetros da média ( $\mu$ ) e da variância ( $\sigma^2$ ). Da Equação 25, pode-se concluir que a distribuição Gaussiana satisfaz a condição:

$$\mathcal{N}(y; \mu, \sigma^2) > 0, \quad (26)$$

e o termo  $\frac{1}{(2\pi\sigma^2)^{1/2}}$  garante a normalização da Gaussiana tal que:

$$\int_{-\infty}^{\infty} \mathcal{N}(y; \mu, \sigma^2) dy = 1. \quad (27)$$

Diz-se que o máximo de uma distribuição é a moda da mesma (BISHOP, 2006). Para uma Gaussiana, a moda coincide com a média da distribuição considerada. A Figura 8 apresenta dois exemplos de distribuições Gaussianas unidimensionais. Em (a), observa-se que a distribuição está centrada no ponto 0. Isto ocorre quando a média é definida como  $\mu = 0$ . Observa-se em (b) que, ao definirmos  $\mu = 2$ , deslocamos a distribuição em duas unidades para a direita, em relação a distribuição mostrada em (a). Ao definir diferentes valores para  $\mu$ , portanto, pode-se deslocar o valor máximo da distribuição (a moda), e, conseqüentemente, a média. A variância ( $\sigma^2$ ), por sua vez, definirá a forma desta distribuição. Nota-se na Figura 8 que a forma das Gaussianas apresentadas em (a) e em (b) se mantém a mesma, uma vez que  $\sigma^2$  permanece constante para ambos os casos.

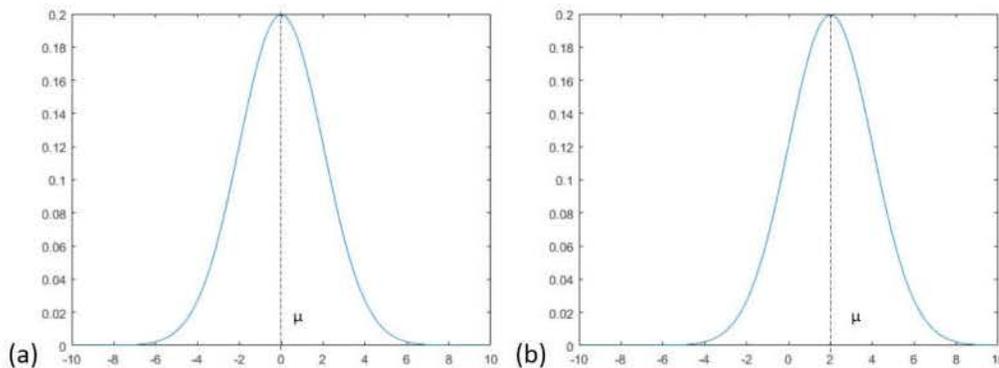


Figura 8: Representação gráfica de distribuições Gaussianas para uma variável contínua e unidimensional: (a) representação de uma distribuição Gaussiana com parâmetros  $\mu = 0$  e  $\sigma^2 = 4$ ; (b) representação de uma distribuição Gaussiana com parâmetros  $\mu = 2$  e  $\sigma^2 = 4$ .

#### 4.3.2 Distribuição Gaussiana Multivariada

A distribuição Gaussiana avaliada em um vetor  $\mathbf{y} \in \mathbb{R}^D$  de D-dimensões é denominada de Gaussiana Multivariada, e pode ser computada por meio da expressão (BISHOP, 2006):

$$\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2\pi^{D/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right]. \quad (28)$$

Onde o vetor  $\boldsymbol{\mu} \in \mathbb{R}^D$  é um vetor coluna de dimensionalidade  $D$  e representa a média da distribuição,  $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$  é uma matriz de tamanho  $D \times D$  denominada de matriz de covariância, e o operador  $|\cdot|$  denota o operador determinante.

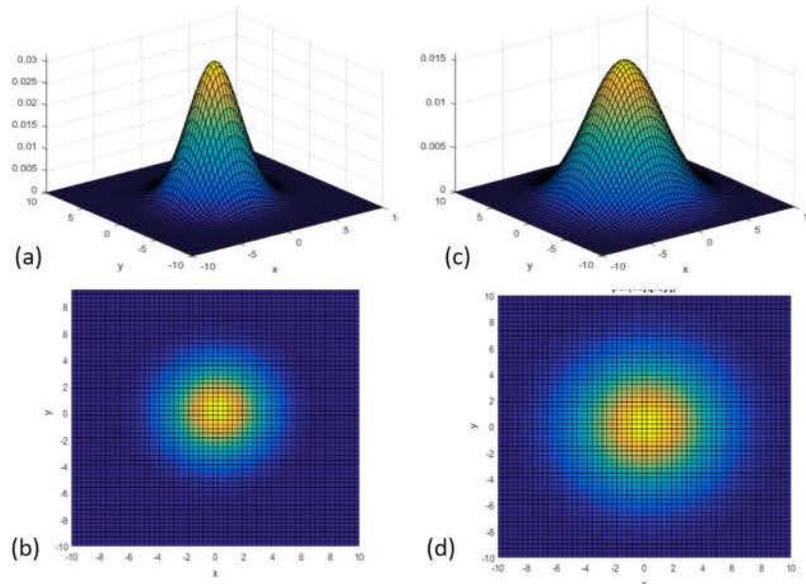


Figura 9: Representação gráfica de distribuições Gaussianas: (a) Representação tridimensional de uma distribuição Gaussiana com parâmetros  $\boldsymbol{\mu} = [0, 0]$  e  $\boldsymbol{\Sigma} = 5 \times \mathbb{I}$ , onde  $\mathbb{I}$  é a matriz identidade de tamanho  $2 \times 2$ ; (b) Representação da distribuição Gaussiana mostrada em (a) numa perspectiva bidimensional; (c) Representação tridimensional de uma distribuição Gaussiana com parâmetros  $\boldsymbol{\mu} = [0, 0]$  e  $\boldsymbol{\Sigma} = 10 \times \mathbb{I}$ , onde  $\mathbb{I}$  é a matriz identidade de tamanho  $2 \times 2$ ; (d) Representação da distribuição Gaussiana mostrada em (c) numa perspectiva bidimensional.

A Figura 9 apresenta dois exemplos de distribuições Gaussianas multivariadas. Uma distribuição Gaussiana multivariada com parâmetros  $\boldsymbol{\mu} = [0, 0]$  e  $\boldsymbol{\Sigma} = 5 \times \mathbb{I}$  é representada graficamente em (a) por meio de uma representação tridimensional, e em (b), por meio de uma representação bidimensional. Representações tridimensional e bidimensional de uma distribuição Gaussiana multivariada com parâmetros  $\boldsymbol{\mu} = [0, 0]$  e  $\boldsymbol{\Sigma} = 10 \times \mathbb{I}$  são apresentadas em (c) e em (d) respectivamente. A matriz  $\mathbb{I}$  denota a matriz identidade de tamanho  $2 \times 2$ . Observa-se que ambas distribuições estão centradas nas coordenadas  $(0, 0)$ , uma vez que possuem o mesmo vetor média  $\boldsymbol{\mu} = [0, 0]$ . Observa-se também a

influência do parâmetro  $\Sigma$ . Como já dito, a covariância, análogo da variância para uma distribuição multivariada, definirá a forma que a distribuição Gaussiana multivariada irá tomar. Em (c) e em (d), nota-se um maior alcance dos valores de probabilidade em relação a distribuição representada em (a) e em (b). Isto decorre dos valores que configuram a matriz de covariância mostrada em (c) e em (d), onde  $\Sigma = 10 \times \mathbb{I}$ , enquanto que a matriz de covariância mostrada em (a) e em (b) é  $\Sigma = 5 \times \mathbb{I}$ . No exemplo mostrado na Figura 9, quanto maior o número real multiplicando a matriz identidade  $\mathbb{I}$ , maior será o alcance da distribuição, e maior será a influência da distribuição em sua vizinhança.

### 4.3.3 Mistura de Gaussianas Multivariadas

Um modelo bastante utilizado em âmbito de visão computacional e reconhecimento de padrões para modelar a distribuição de dados que residem em um espaço de  $D$  dimensões é denominado de mistura de Gaussianas multivariadas (THEODORIDIS et al., 2010). Este modelo consiste em construir uma distribuição a partir da combinação linear de  $J$  distribuições Gaussianas multivariadas.

Dado um vetor  $\mathbf{y} \in \mathbb{R}^D$  de dimensionalidade  $D$ , uma mistura de Gaussianas multivariadas pode ser escrita como uma função  $G(\cdot)$  que combina linearmente  $J$  Gaussianas multivariadas, tal que (THEODORIDIS et al., 2010):

$$G(\mathbf{y}) = \sum_{j=1}^J \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_j, \Sigma_j) \Pi_j, \quad (29)$$

onde  $\Pi_j$  é o fator de ponderação associado a Gaussiana multivariada  $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_j, \Sigma_j)$  centrada em  $\boldsymbol{\mu}_j$ , com covariância  $\Sigma_j$  e avaliada em  $\mathbf{y}$ .  $\Pi_j$  deve ser normalizado de forma que:

$$\Pi_j : \begin{cases} 0 \leq \Pi_j \leq 1; \\ \sum_{j=1}^J \Pi_j = 1. \end{cases} \quad (30)$$

Por meio da expansão indicada na Equação 29, pode-se originar distribuições multimodais mais complexas (que contém múltiplos picos) a partir de  $J$  Gaussianas multivari-

adas. Cada uma das  $j$  Gaussianas multivariadas denotadas por  $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ , podem ser computadas por meio da expressão 28 individualmente.

A Figura 10 apresenta um exemplo simples de uma mistura de três ( $J = 3$ ) Gaussianas multivariadas com parâmetros:  $\boldsymbol{\mu}_1 = [1, 2]$ ;  $\boldsymbol{\mu}_2 = [-1, -2]$ ;  $\boldsymbol{\mu}_3 = [-5, -4]$ ;  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2 \times \mathbb{I}$ ; e  $\Pi_1 = \Pi_2 = \Pi_3 = 1/3$ . A partir das representações gráficas mostradas em (a) e em (b), observa-se a característica multimodal típica da mistura de múltiplas Gaussianas multivariadas. É possível notar também, que a medida que  $J$  aumenta, a distribuição oriunda da mistura torna-se cada vez mais complexa e mais Gaussianas influenciarão os valores de probabilidade nas vizinhanças próximas.

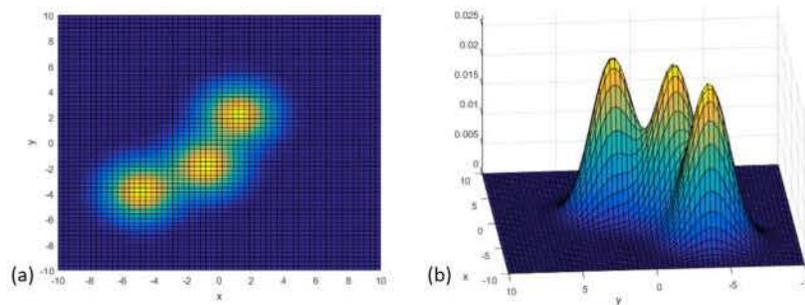


Figura 10: Representação gráfica de uma mistura de três ( $J = 3$ ) distribuições Gaussianas multivariadas: (a) Representação bidimensional da combinação linear de três Gaussianas com parâmetros:  $\boldsymbol{\mu}_1 = [1, 2]$ ;  $\boldsymbol{\mu}_2 = [-1, -2]$ ;  $\boldsymbol{\mu}_3 = [-5, -4]$ ; e  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = 2 \times \mathbb{I}$ , onde  $\mathbb{I}$  é a matriz identidade de tamanho  $2 \times 2$ ; (b) Representação da mistura de Gaussianas mostrada em (a) numa perspectiva tridimensional.

#### 4.4 Conceitos Relevantes de Teoria de Grafos

Um grafo  $G$  consiste em um conjunto de vértices/nodos ( $V$ ) e arestas ( $E$ ) que conectam vértices de um grafo. Arestas podem ser direcionadas (as arestas possuem uma seta indicando um sentido único entre um vértice e outro) ou não-direcionadas. O grafo que possuir apenas arestas direcionadas é denominado de grafo direcionado <sup>2</sup> (do inglês, *directed graph*), e o grafo que possuir apenas arestas não-direcionadas é denominado de grafo não-direcionado (do inglês, *undirected graph*) (BARBER, 2012).

<sup>2</sup>As denominações referentes aos tipos de grafos, os que possuem arestas direcionadas e os que possuem arestas não-direcionadas, podem variar na língua portuguesa. Grafos direcionados podem ser denominados também de grafos direcionais ou dígrafos (CARVALHO, 2005).

Um exemplo de um grafo não direcionado, cujas arestas não possuem direção, pode ser observado na Figura 11. Particularmente, grafos não-direcionados se farão relevantes quando o conceito de um grafo de regiões adjacentes em imagens for introduzido no Capítulo 5.

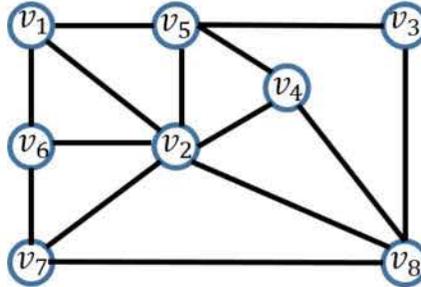


Figura 11: Representação de um grafo não-direcionado acíclico com 8 vértices (em azul) e 14 arestas (em preto).

#### 4.4.1 Árvores Direcionadas

Um grafo direcionado é um par  $(V, E)$  de conjuntos de vértices  $V = \{v_1, v_2, v_3, \dots, v_b\}$  e arestas  $E = \{e_1, e_2, e_3, \dots, e_b\}$ , onde cada aresta parte de um vértice inicial  $v_a$  para um vértice final  $v_f$ , com  $a \neq f$ . Diz-se que um conjunto de arestas  $\{e_1, \dots, e_b\}$  é definido como um caminho direcionado de  $v_a$  para  $v_f$  se  $v_a$  é o vértice inicial da aresta  $e_1$ ,  $v_f$  é o vértice final de  $e_b$ , e o vértice final da aresta  $e_w$  é o vértice inicial da aresta  $e_{w+1}$  para  $w = 1, \dots, b - 1$  (KOONTZ; NARENDRA; FUKUNAGA, 1976).

Uma árvore direcionada é um grafo direcionado no qual as arestas estão todas direcionadas para um vértice específico  $v_r$ , tal que (KOONTZ; NARENDRA; FUKUNAGA, 1976):

- todo vértice  $v_a$  diferente de  $v_r$  é o vértice inicial de exatamente uma aresta;
- $v_r$  não é o vértice inicial de aresta alguma;
- não há caminho direcionado entre um vértice  $v_a$  para ele mesmo (árvores direcionadas não possuem ciclos).

Em uma árvore direcionada, o vértice  $v_r$  é denominado de raiz da árvore direcionada. O vértice final de uma aresta cujo vértice inicial é  $v_a$  é denominado de vértice pai  $\hat{P}(v_a)$

de  $v_a$ . O vértice pai  $\hat{P}(v_a)$  de  $v_a$  é único se  $v_a \neq v_r$ . Há um único caminho partindo de cada vértice  $v_a \neq v_r$  até  $v_r$ , constituído em arestas que partem de  $v_a$  para  $\hat{P}(v_a)$ , de  $\hat{P}(v_a)$  para  $\hat{P}(\hat{P}(v_a))$ , e assim sucessivamente (KOONTZ; NARENDRA; FUKUNAGA, 1976).

A Figura 12 apresenta um exemplo de duas árvores direcionadas, cujas raízes são  $v_{r=1}$  e  $v_{r=7}$ . Observa-se que cada vértice (com exceção das raízes) possui um único pai. Esta estrutura de dados é hierárquica, e tal propriedade será explorada na metodologia desenvolvida no Capítulo 5.

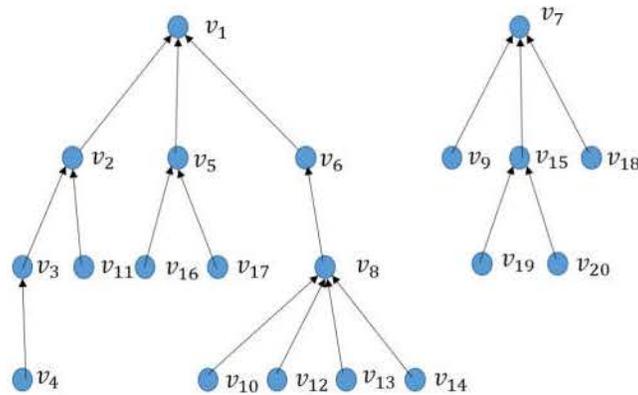


Figura 12: Representação de duas árvores direcionadas. A primeira árvore possui o vértice raiz  $v_{r=1}$  e a segunda árvore possui o vértice raiz  $v_{r=7}$ .

## 5 METODOLOGIA

Neste Capítulo, nós desenvolvemos a metodologia das Árvores Bayesianas Adaptativas (*ABT*, do inglês *Adaptative Bayesian Trees*) para segmentar imagens coloridas. Para tanto, redesenhamos o método de busca de máximos de densidade em uma distribuição de dados discreta, desenvolvido por Koontz et al. (KOONTZ; NARENDRA; FUKUNAGA, 1976), por meio da construção de uma abordagem Bayesiana, e adaptamos-a com o propósito de segmentar imagens coloridas. Apresentamos uma representação hierárquica para imagens coloridas, baseada na estrutura da hierarquia de árvores direcionadas, onde o conteúdo da imagem se encontra acessível em diferentes níveis de detalhamento. Para cada nível desta hierarquia, nodos de árvores direcionadas são unidos, formando *clusters* no espaço de feições, e as probabilidades *a priori* associadas a estes *clusters* são atualizadas até que ocorra a convergência do algoritmo.

A seguir, na Seção 5.1, nós formulamos o problema da segmentação de imagens e resumimos a estrutura do nosso método proposto. Na Seção seguinte (Seção 5.2), nós apresentamos o método das Árvores Bayesianas Adaptativas (*ABT*) aplicado ao problema de segmentação de imagem coloridas.

### 5.1 Formulação do Problema e Resumo do Algoritmo

Permita-nos recapitular as seguintes notações previamente expostas na Introdução desta dissertação. Seja uma imagem  $I$  com  $M \times N$  *pixels*, e  $I(m, n) = \mathbf{x}_i$  o vetor característico  $i$  (e.g. vetor de cor) no *pixel*  $(m, n)$ , com  $m = 1, 2, \dots, M$  e  $n = 1, 2, \dots, N$ . Formalmente, podemos definir o problema de segmentação de imagens como um pro-

cesso de particionamento da imagem  $I$  em um conjunto de subconjuntos  $\{R_1, R_2, \dots, R_J\}$ , tal que:

$$\bigcup_{j=1}^{J \leq M \times N} R_j = I, \quad R_{j'} \cap R_j = \emptyset, \quad j' \neq j. \quad (31)$$

Neste trabalho, nós propomos que partições de uma imagem podem ser obtidas por meio da construção de *clusters* no espaço de feições constituído pelos vetores característicos (i.e. cores) da imagem  $I$ . Nós assumimos que o espaço de feições é a função densidade de probabilidade empírica para os vetores característicos na imagem  $I$ , e fazemos estimativas Bayesianas dos máximos locais (i.e. os máximos de densidade) desta distribuição, por meio de uma estimativa máxima da probabilidade a *posteriori* (MAP). Uma vez determinadas as localizações dos máximos de densidade, nós construímos *clusters* associados a cada máximo, baseados na estrutura local do espaço de feições (i.e. as regiões no espaço de feições de maior densidade correspondem a *clusters* centrados nos máximos de densidade da distribuição de probabilidade dos dados).

### 5.1.1 Resumo da Abordagem Proposta

Existem quatro parâmetros distintos que controlam o método proposto. Estes parâmetros controlam: como nós inicializamos o nosso método ( $\mathcal{I}$ ), quais vetores característicos nós usamos para configurar o espaço de feições ( $\mathcal{FV}$ ), e como a similaridade é definida, que depende de um limiar  $\Phi$  e de uma medida de distância entre pontos de dados ( $\mathcal{D}$ ).

Nosso método começa com um procedimento de inicialização ( $\mathcal{I}$ ) que origina as regiões iniciais  $R_j$  da imagem  $I$ . Tipicamente, imagens contém muitos *pixels*, e um procedimento de inicialização nos permite reduzir o custo computacional do nosso algoritmo de *clustering*. Diversos métodos podem ser utilizados para a inicialização, e escolhas populares são *superpixels* e métodos de sobresegmentação. Neste trabalho definimos o parâmetro  $\mathcal{I}$  como o algoritmo de *superpixel SLIC* (ACHANTA et al., 2012). Portanto, ao invés de agrupar vetores característicos associados a *pixels* da imagem  $I$ , nós agrupamos vetores característicos associados a regiões  $R_j$  da imagem  $I$ .

Na sequência, vetores característicos são extraídos (controlados pelo parâmetro  $\mathcal{FV}$ )

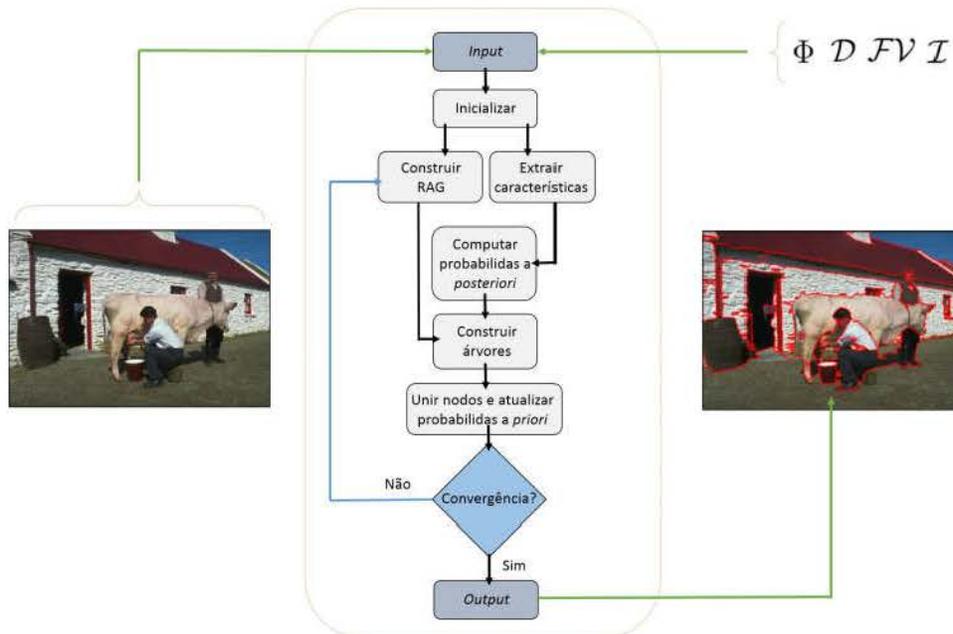


Figura 13: Visão geral do algoritmo de segmentação de imagem proposto.

para cada região da imagem  $R_j \in I$ . Em âmbito de segmentação de imagens, escolhas comuns para vetores característicos são cores e texturas (MALIK et al., 2001). Neste trabalho, vetores característicos  $\mathbf{x}_j$  associados a regiões  $R_j$  correspondem à média das características de cor, no espaço de cores RGB. Para extrair a média das características de cor, nós simplesmente computamos o valor médio das características de cor dos *pixels* em cada região  $R_j$  da imagem  $I$ . Sendo assim, os vetores característicos  $\mathbf{x}_j$  obtidos correspondem a vetores 3-D com os valores médios das cores RGB.

O parâmetro  $\Phi$  de limiar e o parâmetro referente a medida de distância ( $\mathcal{D}$ ) são utilizados para definir a similaridade dos vetores característicos no espaço de feições. O parâmetro  $\Phi$  é um escalar (tipicamente definido como um número real positivo), enquanto que  $\mathcal{D}$  é configurado como a distância Euclideana.

Utilizando a similaridade da média dos vetores característicos de cor, juntamente a restrições espaciais da imagem, nós realizamos estimativas Bayesianas dos máximos de densidade dos dados, construímos árvores direcionadas, e agrupamos os vetores médios de cor no espaço de feições. Em cada iteração do esquema proposto, uma hierarquia de árvores direcionadas é obtida (iterativamente), formando vetores característicos, em diferentes níveis da hierarquia, no espaço de feições. O efeito de agrupamento de vetores

característicos no espaço de feições também origina uma hierarquia de árvores direcionadas no domínio da imagem, e após a convergência do nosso algoritmo, nós obtemos a segmentação final da imagem correspondente aos *clusters* construídos no espaço de feições.

## 5.2 Método Proposto

Nesta seção, apresentamos como as estimativas Bayesianas de máximos de densidade são usadas para construir árvores direcionadas e *clusters* no espaço de feições. Primeiramente, nós descrevemos nossa abordagem Bayesiana para construir árvores direcionadas adaptativas e *clusters*, e, em seguida, com base em nosso esquema de representação hierárquico, mostramos como segmentar uma imagem colorida.

### 5.2.1 Árvores Bayesianas Adaptativas

Suponha que a imagem colorida  $I$  (com  $M \times N$  pixels) é inicialmente particionada em  $J$  regiões não sobrepostas  $\mathcal{R} = \{R_1, R_2, \dots, R_J\}$ , tal que  $\bigcup_{j=1}^J R_j = I$ , com  $R_j \subseteq I$  e  $J \leq M \times N$ . Denotamos a coleção de vetores característicos  $\mathbf{x}_j$  (média das cores RGB) associados às regiões  $R_j \in I$  como sendo o conjunto  $X = \{\mathbf{x}_j \mid \mathbf{x}_j \in \mathbb{R}^D\}$ , com  $j = 1, 2, \dots, J$  regiões da imagem, mapeados em um espaço de feições de dimensionalidade  $D$  (i.e., se  $\mathbf{x}_j$  corresponde a média das cores RGB, então  $D = 3$ ).

Definimos a vizinhança no espaço de feições  $\eta_{\Phi}^j$  de  $\mathbf{x}_j$  como:

$$\eta_{\Phi}^j = \{\mathbf{x}_u \mid d_{ju} \leq \Phi, \forall \mathbf{x}_u \in X\}, \quad (32)$$

com  $d_{ju}$  sendo a distância Euclideana entre  $\mathbf{x}_j$  e  $\mathbf{x}_u$  e,  $\Phi$  um dado limiar. Denotamos  $\eta_{\Phi} = \{\eta_{\Phi}^1, \eta_{\Phi}^2, \dots, \eta_{\Phi}^J\}$  como sendo a coleção de vizinhos do espaço de feições dos vetores característicos  $\mathbf{x}_j \in X$ .

A fim de considerar a localização espacial das regiões  $R_j \in I$ , nós construímos um grafo de regiões adjacentes (*RAG*, do inglês *region adjacency graph*)  $\mathcal{G} = (\bar{V}, \bar{E})$ , associado a vetores característicos  $\mathbf{x}_j$ , onde cada nodo  $\bar{v}_j \in \bar{V}$  corresponde a uma região  $R_j$ , e uma aresta  $\bar{e}_{j,c} \in \bar{E}$  conecta as regiões  $R_j$  e  $R_c$  que são espacialmente adjacentes

na imagem  $I$  (MOBAHI et al., 2011). Posto isto, a vizinhança espacial da região  $R_j$  é expressa como o conjunto de regiões  $\{R_c \mid \bar{e}_{j,c} \in \bar{E}\}$ . Podemos estender a notação do conjunto mencionado anteriormente, e definir  $\mathcal{E}^j$  como o conjunto de vetores característicos associados aos vizinhos espaciais de  $R_j$ , que também contém  $\mathbf{x}_j$ , de modo que:

$$\mathcal{E}^j = \{\mathbf{x}_j, \mathbf{x}_c \mid \bar{e}_{j,c} \in \bar{E}\}. \quad (33)$$

Nós também denotamos  $\mathcal{E} = \{\mathcal{E}^1, \mathcal{E}^2, \dots, \mathcal{E}^J\}$  como sendo a coleção de vetores característicos associados aos vizinhos espaciais das regiões  $R_j \in I$ . Portanto, os vizinhos no espaço de feições de  $\mathbf{x}_j$  são denotados por  $\eta_{\Phi}^j$ , e os vetores característicos dos vizinhos espaciais de  $R_j$  são denotados como  $\mathcal{E}^j$ .

Dadas as notações acima mencionadas, considere que cada vetor característico  $\mathbf{x}_j$  é um nodo de um grafo direcionado no espaço de feições. Na formulação proposta, nós gostaríamos de formar um conjunto de *clusters* dos vetores característicos pertencentes ao conjunto  $X$ , por meio da construção de árvores direcionadas, cujas raízes  $\mathbf{x}_r$ , localizam-se nos máximos de densidade da distribuição dos dados. Sendo assim, a primeira etapa da nossa proposta consiste em determinar quais vetores característicos  $\mathbf{x}_j$ , no espaço de feições, devem ser atribuídos a cada máximo de densidade (cada nodo raiz). Na nossa formulação Bayesiana deste problema, a probabilidade a *posteriori*  $p(\mathbf{x}_{j,r} \mid X)$  do vetor característico  $\mathbf{x}_j$  de corresponder a um nodo raiz  $\mathbf{x}_r$  (i.e.  $\mathbf{x}_{j,r}$ ), dado o conjunto  $X$ , pode ser modelada por meio de uma mistura de Gaussianas:

$$p(\mathbf{x}_{j,r} \mid X) = \frac{\sum_{t=1}^J \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_t, \Sigma = \mathbb{I}) \Pi_t}{\sum_{q=1}^J \sum_{t=1}^J \mathcal{N}(\mathbf{x}_q; \boldsymbol{\mu} = \mathbf{x}_t, \Sigma = \mathbb{I}) \Pi_t}, \quad (34)$$

onde o termo de ponderação ( $\Pi_t$ ) corresponde a frequência de ocorrência do vetor caracte-

terístico  $\mathbf{x}_t \in X$  e deve satisfazer as seguintes restrições:

$$\begin{cases} 0 \leq \Pi_t \leq 1; \\ \sum_{t=1}^J \Pi_t = 1. \end{cases} \quad (35)$$

Cada componente da mistura  $\mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_t, \Sigma = \mathbb{I})$  consiste em uma distribuição Gaussiana multivariada avaliada no vetor característico  $\mathbf{x}_j$  e é expressa como:

$$\mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_t, \Sigma = \mathbb{I}) = \frac{1}{(2\pi)^{D/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_j - \mathbf{x}_t)^T \mathbb{I}(\mathbf{x}_j - \mathbf{x}_t)\right], \quad (36)$$

onde o vetor  $\boldsymbol{\mu} \in \mathbb{R}^D$  é um vetor coluna de dimensionalidade  $D$  e representa a média da distribuição,  $\Sigma \in \mathbb{R}^{D \times D}$  denota a matriz de covariância de tamanho  $D \times D$  e  $\mathbb{I}$  é a matriz identidade<sup>1</sup>.

A probabilidade *a posteriori*  $p(\mathbf{x}_{j,r} | X)$  na Equação 34 pode ser reescrita na forma matricial do seguinte modo:

$$p(\mathbf{x}_{j,r} | X) = \frac{\begin{bmatrix} \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_1, \Sigma = \mathbb{I}) \\ \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_2, \Sigma = \mathbb{I}) \\ \vdots \\ \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu} = \mathbf{x}_J, \Sigma = \mathbb{I}) \end{bmatrix}^T \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_J \end{bmatrix}}{\sum_{q=1}^J \begin{bmatrix} \mathcal{N}(\mathbf{x}_q; \boldsymbol{\mu} = \mathbf{x}_1, \Sigma = \mathbb{I}) \\ \mathcal{N}(\mathbf{x}_q; \boldsymbol{\mu} = \mathbf{x}_2, \Sigma = \mathbb{I}) \\ \vdots \\ \mathcal{N}(\mathbf{x}_q; \boldsymbol{\mu} = \mathbf{x}_J, \Sigma = \mathbb{I}) \end{bmatrix}^T \begin{bmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_J \end{bmatrix}}. \quad (37)$$

Ao definir um modelo Gaussiano multivariável (ver Equação 36), com parâmetros  $\boldsymbol{\mu} = \mathbf{x}_t$  e  $\Sigma = \mathbb{I}$ , nós assumimos uma distribuição de probabilidade simétrica centrada em cada vetor característico (nodo)  $\mathbf{x}_t$  no espaço de feições, que é avaliada em  $\mathbf{x}_j$ . Ao

<sup>1</sup>Nesta dissertação, cabe destacar que identificamos empiricamente, em testes preliminares, que o uso da matriz identidade como matriz de covariância tende a apresentar resultados mais robustos para a segmentação de imagens coloridas, em relação a covariância das cores para cada *superpixel*.

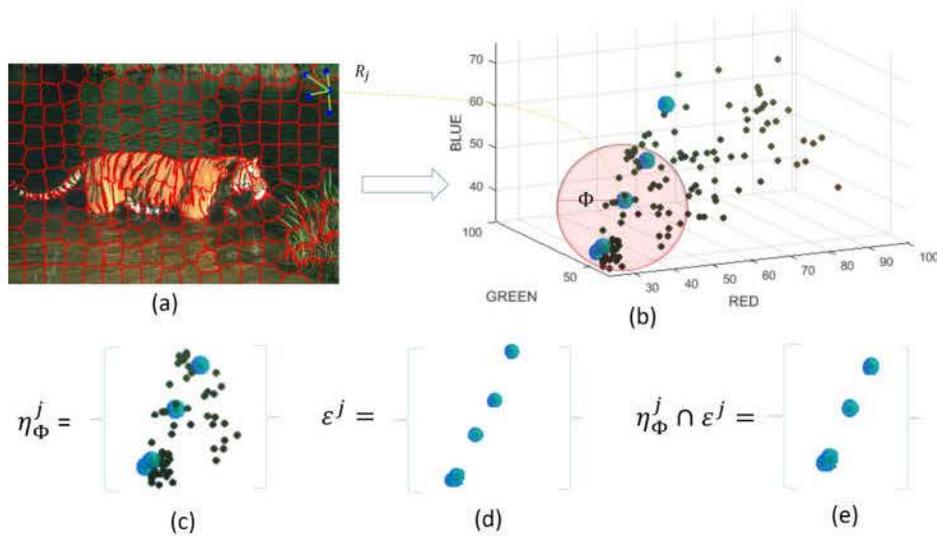


Figura 14: Ilustração da interseção expressa pelo conjunto  $\{\eta_\Phi^j \cap \epsilon^j\}$  no espaço de cores RGB: (a) a imagem colorida  $I$  é decomposta em regiões coloridas homogêneas, obtidas pelo algoritmo *SLIC* (contornos das regiões em vermelho), tal que  $\bigcup_{j=1}^J R_j = I$ . Utilizando um grafo de regiões adjacentes  $\mathcal{G}$ , a vizinhança espacial de uma região arbitrária  $R_j$  é representada (nodos azuis indicam as regiões adjacentes a região  $R_j$  e as arestas que as conectam estão marcadas em verde); (b) cada região da imagem  $I$  esta associada ao valor médio das cores RGB ( $\mathbf{x}_j$ ) dos *pixels* daquela região no espaço de feições. Vetores de cores em destaque azul indicam os valores das características médias dos vizinhos espaciais da região  $R_j$  em  $\mathcal{G}$ , e uma esfera vermelha de raio  $\Phi$  delimita a similaridade entre  $\mathbf{x}_j$  e seus vizinhos no espaço de feições; (c)  $\mathbf{x}_j$  possui vizinhos no espaço de feições definidos pelo conjunto  $\eta_\Phi^j$ ; (d) os vetores característicos associados a vizinhança espacial da região  $R_j$  são dados pelo conjunto  $\epsilon^j$ ; (e) o conjunto  $\{\eta_\Phi^j \cap \epsilon^j\}$  contém os vetores característicos candidatos que potencialmente podem se tornar nodos raiz no espaço de feições.

assumirmos a probabilidade a posteriori  $p(\mathbf{x}_{j,r} \mid X)$  como sendo uma mistura destas Gaussianas multivariadas, estaremos levando em conta a influência das distribuições centradas nos vetores característicos  $\mathbf{x}_t$ , e ponderadas pelas suas probabilidades *a priori*  $\Pi_t$  em  $\mathbf{x}_j$ .

Dada uma vizinhança no espaço de feições  $\eta_\Phi^j$  (ver Equação 32) de  $\mathbf{x}_j \in X$  e dado o conjunto  $\epsilon^j$  (ver Equação 33) de vetores característicos associados aos vizinhos espaciais da região  $R_j \in \mathcal{R}$ , com  $j = 1, 2, \dots, J$ , nós encontramos os nodos raízes  $\hat{\mathbf{x}}_{k,r}$  utilizando a

abordagem da máxima estimativa da probabilidade a *posteriori* (MAP):

$$\hat{\mathbf{x}}_{k,r} = \arg \max_{\forall \mathbf{x}_k \in \{\eta_{\Phi}^j \cap \mathcal{E}^j\}} \{p(\mathbf{x}_{k,r} | X)\}. \quad (38)$$

Procuramos maximizar as probabilidades a *posteriori*  $p(\mathbf{x}_{k,r} | X)$  dos vetores característicos  $\mathbf{x}_k$  no conjunto  $\{\eta_{\Phi}^j \cap \mathcal{E}^j\}$  de corresponder a um nodo raíz (i.e. os vetores característicos disponíveis na vizinhança  $\eta_{\Phi}^j$  de  $\mathbf{x}_j$  no espaço de feições e em  $\mathcal{E}^j$ , i.e. considerando os vetores característicos associados aos vizinhos espaciais de  $R_j$  no grafo de adjacências  $\mathcal{G}$ ) (ver Figura14). Assim sendo, nós garantimos que os vetores característicos  $\mathbf{x}_k$ , candidatos a nodos raízes, de fato estão associados a regiões adjacentes no domínio espacial da imagem, e também, que estes são similares no espaço de feições. Para um dado  $\mathbf{x}_k$ , pode-se calcular a probabilidade a *posteriori*  $p(\mathbf{x}_{k,r} | X)$  utilizando a Equação 34.

De modo a esclarecer as implicações da nossa formulação para a probabilidade a *posteriori* (ver Equação 34), reescrevemos a Equação 36 definindo  $\Delta = \mathbf{x}_j - \mathbf{x}_t$  e  $K = \frac{1}{(2\pi)^{D/2}}$ , onde é assumido que  $\mu = \mathbf{x}_t$ :

$$\mathcal{N}(\mathbf{x}_j; \mu = \mathbf{x}_t, \Sigma = \mathbb{I}) = K \exp\left[-\frac{1}{2}(\Delta^T \mathbb{I} \Delta)\right]. \quad (39)$$

Tomando o limite da Equação 39 com respeito a  $\Delta$  crescente, nós temos:

$$K \exp\left[-\frac{1}{2}(\Delta^T \mathbb{I} \Delta)\right] \rightarrow 0. \quad (40)$$

Sendo assim, a Equação 40 sugere que há um número limitado de componentes de Gaussianas multivariadas, cujos valores são significativamente maior que zero, quando uma mistura é avaliada em cada vetor característico  $\mathbf{x}_j$ . Em nossa proposta, o número de componentes Gaussianas “ativas” avaliadas em  $\mathbf{x}_j$  serão reduzidos quando a diferença  $(\mathbf{x}_j - \mathbf{x}_t)$  aumentar.

Uma vez determinados os valores de probabilidade a *posteriori* para cada vetor característico  $\mathbf{x}_j$ , constrói-se árvores direcionadas, fazendo uso das probabilidades a *posteriori* computadas, por meio do Algoritmo 1. As árvores direcionadas são construídas identi-

ficando um nodo pai para cada  $x_j$ , e os nodos para os quais um nodo pai não pode ser atribuído tornam-se nodos raiz. Posto que árvores direcionadas não contêm ciclos, o procedimento de atribuição dos nodos pai evita grafos cíclicos certificando que se  $x_j$  não é uma raiz, então há apenas um caminho de  $x_j$  para o seu pai, ou, para a sua raiz. O número de árvores é igual ao número de raízes, e cada  $x_j$  pertence apenas a uma árvore. O Algoritmo 1 atribui cada vetor  $x_j$  para exatamente uma árvore. Se  $x_j$  é raiz  $x_{j,r}$  da árvore, o mesmo é atribuído à árvore que tem essa raiz, caso contrário,  $x_j$  é designado a pertencer a mesma árvore que o seu pai. O Algoritmo 1 pode ser utilizado recursivamente a fim de identificar as raízes das árvores e os vetores característicos associados com cada árvore. Cada *cluster* obtido é uma árvore direcionada unimodal, cuja raiz localiza-se em um máximo de densidade da distribuição dos vetores característicos (KOONTZ; NARENDRA; FUKUNAGA, 1976). Este método de procura de máximos de densidade em uma distribuição de dados, determina o número de *clusters*, assim como, a associação de todos os vetores característicos  $x_j \in X$  com seus respectivos *clusters*, dada uma escolha para o parâmetro  $\Phi$ . O parâmetro  $\Phi$  pode ser usado para selecionar apenas os máximos de densidade dominantes (*clusters*) da distribuição de dados. Koontz et al. (KOONTZ; NARENDRA; FUKUNAGA, 1976) demonstram que que o método proposto, para busca de máximos de densidade, converge para o máximo verdadeiro da distribuição de dados, e a sucessão de ancestrais de  $x_j$  é uma abordagem de ascensão íngreme (do inglês, *steepest ascent*) em direção ao ponto de densidade máxima delimitado pelo conjunto  $\{\eta_{\Phi}^j \cap \mathcal{E}^j\}$  para um dado  $x_j$ , que não necessariamente será o máximo de densidade mais próximo de  $x_j$ .

Utilizando a *array* de árvores computada com o Algoritmo 1, uma partição inicial do espaço de feições pode ser obtida atribuindo os vetores característicos  $x_j$  para os seus associados nodos raízes (ver passo 13 do Algoritmo 2). No espaço de feições, este procedimento é equivalente a atribuir a todos os nodos pertencentes a uma árvore direcionada ao seu nodo raiz. Figura 15 apresenta uma ilustração do procedimento de fusão em um espaço vetorial abstrato de dimensionalidade 3. A Figura 15 a) mostra três diferentes árvores direcionadas com raízes  $x_{1,r}$ ,  $x_{5,r}$  e  $x_{9,r}$ . O número total de descendentes da árvore

**Algorithm 1** Construir Árvores Direcionadas

---

```

1: procedure BUILDTREES( $p(\mathbf{x}_{1,r} | X), \dots, p(\mathbf{x}_{J,r} | X); \eta_{\Phi}; \mathcal{E}$ ).
2:   for  $j \leftarrow 1$  to  $J$  do
3:     Encontrar o índice  $k \neq j$  que corresponde ao vizinho do vetor característico
        $\mathbf{x}_j$  com o valor máximo de probabilidade a posteriori em  $\{\eta_{\Phi}^j \cap \mathcal{E}^j\}$ ;
4:     if  $p(\mathbf{x}_{j,r} | X) > p(\mathbf{x}_{k,r} | X)$  then
5:        $\hat{P}(\mathbf{x}_j) = \mathbf{x}_{j,r}$ ; ▷  $\mathbf{x}_j$  é uma raiz  $\mathbf{x}_{j,r}$ 
6:     else if  $p(\mathbf{x}_{j,r} | X) < p(\mathbf{x}_{k,r} | X)$  then
7:        $\hat{P}(\mathbf{x}_j) = \mathbf{x}_k$ ; ▷ o pai de  $\mathbf{x}_j$  é  $\mathbf{x}_k$ 
8:     else
9:        $\hat{P}(\mathbf{x}_k) = \mathbf{x}_{k,r}$ ; ▷  $\mathbf{x}_k$  é uma raiz  $\mathbf{x}_{k,r}$ 
10:       $\hat{P}(\mathbf{x}_j) = \mathbf{x}_k$ ; ▷ o pai de  $\mathbf{x}_j$  é  $\mathbf{x}_k$ 
11:   return array de árvores  $\hat{P}$ .

```

---

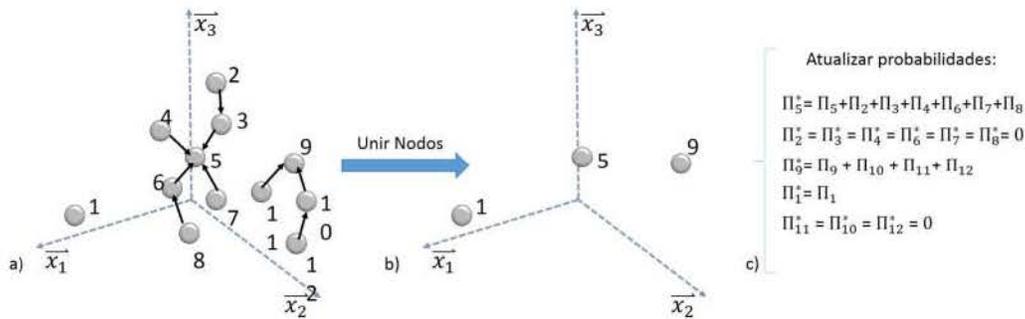


Figura 15: Ilustração do procedimento para unir nodos de árvores direcionadas em um espaço de feições abstrato de três dimensões: a) árvores direcionadas originais com raízes  $\mathbf{x}_{1,r}$ ,  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$ ; b) *clusters* obtidos por meio da atribuição:  $\mathbf{x}_{1,r} \rightarrow \mathbf{x}_{1,r}$ ;  $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8 \rightarrow \mathbf{x}_{5,r}$ ;  $\mathbf{x}_{10}, \mathbf{x}_{11}, \mathbf{x}_{12} \rightarrow \mathbf{x}_{9,r}$ ; c) as probabilidades a priori  $\Pi_1$ ,  $\Pi_5$  e  $\Pi_9$  das raízes  $\mathbf{x}_{1,r}$ ,  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$  são atualizadas para  $\Pi_1^*$ ,  $\Pi_5^*$  e  $\Pi_9^*$ , respectivamente.

cujas raízes  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$  são iguais a 6, o número total de descendentes da árvore cuja raiz é  $\mathbf{x}_{9,r}$  é igual a 3, e a árvore restante é constituída apenas pelo nodo raiz  $\mathbf{x}_{1,r}$ . A Figura 15 b) mostra os *clusters* obtidos por meio da atribuição:  $\mathbf{x}_{1,r} \rightarrow \mathbf{x}_{1,r}$ ;  $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8 \rightarrow \mathbf{x}_{5,r}$ ;  $\mathbf{x}_{10}, \mathbf{x}_{11}, \mathbf{x}_{12} \rightarrow \mathbf{x}_{9,r}$ .

### 5.2.2 Representação Hierárquica de Imagens Coloridas

Vamos supor uma notação mais conveniente para representar uma imagem colorida  $I$  de forma hierárquica. Denotamos uma imagem  $I$  no nível de hierarquia  $s$  por  $I^s$ , onde  $s = 0, 1, 2, \dots, L - 1$  são possíveis níveis desta hierarquia. Nesta notação, a imagem original é denotada como  $I^{s=0}$ , e a imagem obtida pelo procedimento de inicialização  $\mathcal{I}$  (i.e. o método *SLIC*) é denotada como  $I^{s=1}$  (ver Figura 16).

Suponha que a imagem  $I^s$  possa ser particionada em  $\text{card}(\mathcal{R}^s)$  regiões não sobrepostas  $\mathcal{R}^s = \{R_1^s, R_2^s, \dots, R_{\text{card}(\mathcal{R}^s)}^s\}$ , tal que  $R_j^s \subseteq I^s$ . Nós denotamos os vetores característicos associados às regiões  $R_j^s \in I^s$  como o conjunto  $X^s = \{\mathbf{x}_j^s \mid \mathbf{x}_j^s \in \mathbb{R}^D\}$ , com  $j = 1, 2, \dots, \text{card}(\mathcal{R}^s)$ .

Uma versão simplificada ( $X^{s+1}$ ) do conjunto de vetores característicos ( $X^s$ ) pode ser produzida por meio da computação das árvores direcionadas para o conjunto de vetores característicos em  $X^s$ , e pela substituição de cada vetor característico pertencente a  $X^s$ , pelo seu respectivo protótipo de *cluster* (i.e. substituindo os vetores característicos em uma árvore por suas respectivas raízes para obter  $X^{s+1}$ ). O conjunto  $X^{s+1}$  consiste em uma versão simplificada de  $X^s$ , que provavelmente possui menos raízes de árvores direcionadas que  $X^s$  (o número de vetores característicos  $\mathbf{x}_j^s$  serão iguais ao número de raízes quando todos os vetores  $\mathbf{x}_j^s$  tornarem-se raízes de árvores direcionadas). O conjunto  $X^{s+1}$  de vetores característicos  $\mathbf{x}_j^{s+1}$  consiste em um subconjunto de  $X^s$  (i.e.,  $X^{s+1} \subseteq X^s$ ), cujos vetores  $\mathbf{x}_j^{s+1} \in X^{s+1}$  são raízes de árvores direcionadas encontradas no espaço de feições  $X^s$  (i.e.  $\mathbf{x}_{j,r}^{s+1} = \mathbf{x}_z^{s-1}$ ).

O Algoritmo 2 descreve o procedimento de representação hierárquica das Árvores Bayesianas Adaptativas. Este algoritmo converge quando não há mais possíveis fusões (i.e. todos os vetores característicos restantes no espaço de feições nesta fase são nodos raízes de árvores direcionadas). A medida em que os nodos das árvores são representados por raízes de árvore no espaço de feições, o número de vizinhos no espaço de feições  $\eta_{\Phi}^{j,s}$  de  $\mathbf{x}_j^s$ , no nível de hierarquia  $s$ , é reduzido progressivamente (i.e. cada nodo que não é uma raiz deve pertencer a um *cluster* associado a raiz de uma árvore). Portanto, a representação hierárquica depende fortemente de uma escolha para o parâmetro  $\Phi$ , que é utilizado para determinar a vizinhança no espaço de feições  $\eta_{\Phi}^{j,s}$  de  $\mathbf{x}_j^s$ .

Dada uma escolha para o parâmetro  $\Phi$ , o procedimento descrito pelo Algoritmo 2 pode ser usado iterativamente para obter uma hierarquia de árvores direcionadas, que representa as decomposições do conjunto de vetores característicos contidos em  $X^s$ . A medida em que  $s$  aumenta, os dados contidos em  $X^s$  podem ser representados progressivamente por versões mais simples  $X^{s+1}, \dots, X^{L-1}$ , que contém apenas os vetores característicos as-

---

**Algorithm 2** Decomposição de Árvores Bayesianas Adaptativas
 

---

```

1: procedure DECOMPOSEDATA( $I^1; X^1; \Phi$ ).
2:  $I^s \leftarrow I^1$ ;
3:  $X^s \leftarrow X^1$ ;
4:  $X^{s+1} \leftarrow \emptyset$ ;
5:   while  $X^{s+1} \neq X^s$  do
6:     Construir RAG  $\mathcal{G} = (\bar{V}, \bar{E})$  para a imagem  $I^s$ ;
7:     for  $j \leftarrow 1$  to  $J$  do
8:       Computar  $p(\mathbf{x}_{j,r} | X)$  usando a Equação 34;
9:        $\eta_{\Phi}^j \leftarrow$  encontrar os vizinhos no espaço de feições de  $\mathbf{x}_j$ ;
10:       $\mathcal{E}^j \leftarrow$  encontrar os vetores característicos associados as regiões vizinhas
      de  $R_j$  delimitados por  $\mathcal{G}$ ;
11:       $X^s \leftarrow X^{s+1}$ ;
12:      Aplicar o procedimento BuildTrees (no Algoritmo 1);
13:      Atribuir todos os membros de uma árvore para a sua raiz, obtendo  $X^{s+1}$ ;
14:      Atualizar probabilidades a priori.
15: return  $X^{L-1}$ 

```

---

sumidos como estimativas Bayesianas dos máximos de probabilidade da distribuição de dados (i.e. raízes das árvores). Além disso, para cada nível de decomposição  $X^{s+1}$ , os nodos são atribuídos aos nodos raízes, e suas probabilidades a *priori* são atualizadas pela soma das probabilidades a *priori* dos descendentes daquele nodo raiz (ver passo 14 do Algoritmo 2). Na Figura 15 c) as raízes das árvores direcionadas  $\mathbf{x}_{1,r}$ ,  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$  tem suas probabilidades a *priori* atualizadas: as probabilidades a *priori*  $\Pi_1$ ,  $\Pi_5$  e  $\Pi_9$  das raízes  $\mathbf{x}_{1,r}$ ,  $\mathbf{x}_{5,r}$  e  $\mathbf{x}_{9,r}$  são atualizadas para  $\Pi_1^*$ ,  $\Pi_5^*$  e  $\Pi_9^*$ , respectivamente.

No domínio espacial, o efeito de atribuir um conjunto de vetores característicos  $X^s$  pelas suas respectivas raízes, originando um novo conjunto de vetores característicos  $X^{s+1}$ , é que as regiões da imagem  $R_y^s$  e  $R_j^s$  em  $I^s$  são unidas em regiões maiores  $R_{z,r}^{s+1}$  em  $I^{s+1}$ , se seus vetores característicos associados  $\mathbf{x}_y^s$  e  $\mathbf{x}_j^s$  possuem a mesma raiz da árvore direcionada  $\mathbf{x}_{z,r}^s$  no espaço de feições. Isto corresponde a uma união de nodos no espaço de feições (ver Figura 15). Assim, regiões de uma imagem  $R_j^{s+1}$  em  $I^{s+1}$  são pais de regiões  $R_j^s$  em  $I^s$ , formando árvores direcionadas no domínio espacial também (ver Figura 16). A região do céu na Figura 16 ilustra a árvore direcionada no domínio espacial da imagem obtida pela atribuição de um conjunto de vetores característicos  $X^s$  às suas respectivas raízes de árvore  $X^{s+1}$  no espaço de feições. Nodos raízes  $R_{j,r}^s \in I^s$ , no domínio espacial da imagem, podem ser determinados identificando se seus vetores característicos

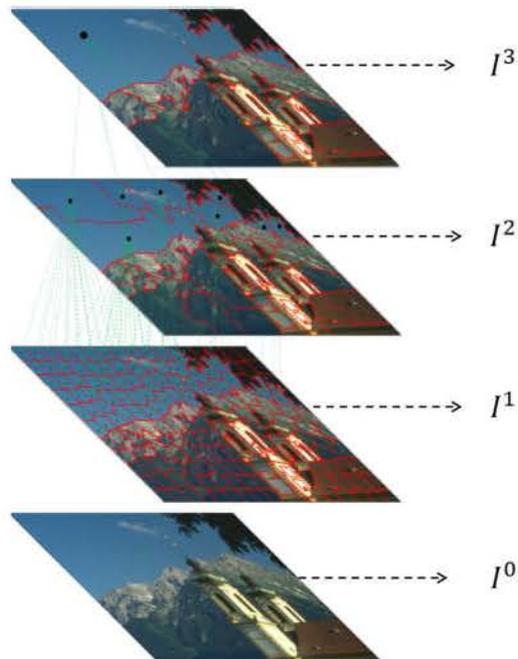


Figura 16: Ilustração da representação hierárquica (com  $L = 4$  decomposições) da região do céu no domínio espacial da imagem.  $I^0$  é a imagem original.  $I^1$  é a imagem obtida pelo método de inicialização *SLIC*.  $I^2$  é originada unindo as regiões de  $I^1$ , e  $I^3$  é a  $L$ -ésima decomposição. Ponto pretos indicam nodos raízes no domínio espacial para cada nível de hierarquia (i.e. regiões  $R_{z,r}^{s+1}$  em  $I^{s+1}$ ) associadas ao vetores característicos  $\mathbf{x}_{z,r}^s$  que são nodos raízes no espaço de feições) as linhas verdes representam os caminhos diretos entre os nodos. Note que regiões da imagem  $I^{s+1}$  são ancestrais de regiões na imagem  $I^s$ . Contornos das regiões são marcados em vermelho.

associados  $\mathbf{x}_{j,r}^s$  são raízes no espaço de feições.

## 6 RESULTADOS EXPERIMENTAIS

De modo a avaliar os resultados de segmentação obtidos pelo método proposto, nós realizamos experimentos utilizando o banco público de dados de segmentação BSD300 (MARTIN et al., 2001). Este banco de dados é constituído de 300 imagens naturais coloridas de tamanho  $481 \times 321$  (ou,  $321 \times 481$ ), sendo 200 imagens parte de um conjunto treino e 100 imagens parte de um conjunto teste. O padrão-ouro deste banco de dados é composto por diferentes segmentações manuais delimitadas por humanos (há em média 6 segmentações manuais para cada imagem do banco de dados). Diferentes segmentações manuais nos permitem investigar como o método de segmentação proposto (Capítulo 5) se assemelha a segmentação de imagens coloridas realizada por seres humanos, no que diz respeito a medidas tipicamente utilizadas para avaliar o desempenho de abordagens de segmentação.

Os experimentos propostos neste Capítulo foram projetados com o propósito de avaliar os efeitos de diferentes configurações dos parâmetros  $\mathcal{L}$ ,  $\mathcal{FV}$ ,  $\Phi$  e  $\mathcal{D}$  nas segmentações obtidas utilizando o método *ABT*. Nós investigamos a significância destas diferentes configurações por meio de uma análise de variâncias (ANOVA) em um projeto fatorial completo (MONTGOMERY, 2008). Utilizando os resultados obtidos da ANOVA, nós otimizamos a nossa configuração por meio da escolha dos parâmetros que maximizam a medida de performance *PRI* (do inglês, *Probabilistic Rand Index*). Uma vez encontrados os parâmetros ótimos, nós computamos também as medidas de performance *VOI*, *GCE*, e *BDE*, com a finalidade de comparar quantitativamente o nosso método com outros métodos disponíveis na literatura. Nós também avaliamos o nosso método qualitativamente junto a abordagens encontradas na literatura.

A seguir, na Seção 6.1, descrevemos o significado qualitativo e mostramos como é possível computar as medidas de performance *PRI* (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005), *VOI* (MOBAHI et al., 2011), *GCE* (MARTIN et al., 2001) e *BDE* (FREIXENET et al., 2002), que serão utilizadas para avaliação dos resultados do nosso método. Na Seção 6.2, nós apresentamos as diferentes configurações exploradas, em nossos experimento, para os parâmetros  $\mathcal{I}$ ,  $\mathcal{FV}$ ,  $\Phi$  e  $\mathcal{D}$ , assim como os efeitos qualitativos dessas configurações na segmentação obtida. Dedicamos a Seção 6.3 para descrever como os parâmetros do método desenvolvido foram otimizados, e por último, na Seção 6.5, comparamos os resultados de segmentação obtidos pelo nosso método com outros métodos disponíveis na literatura de forma qualitativa e quantitativa.

## 6.1 Medidas de Performance

Uma vez que o padrão ouro do banco de dados BSD300 utilizado é constituído de múltiplas segmentações manuais para cada imagem, tipicamente computa-se a média das medidas *PRI*, *VOI*, *GCE* e *BDE* referentes a cada padrão ouro para fins de comparação. Neste trabalho, fez-se uso do código de Allen Y. Yang que se encontra publicamente disponível no endereço: [https://people.eecs.berkeley.edu/~yang/software/lossy\\_segmentation/](https://people.eecs.berkeley.edu/~yang/software/lossy_segmentation/). A seguir, detalhamos o significado qualitativo e mostramos como computar as medidas de performance *PRI*, *VOI*, *GCE* e *BDE*.

### 6.1.1 Probabilistic Rand Index (PRI)

*PRI* (do inglês, *Probabilistic Rand Index*) é uma medida de probabilidade cujos valores situam-se entre  $[0, 1]$ , e um maior valor de probabilidade indica uma melhor performance de segmentação (i.e. maior similaridade entre partições de uma imagem). Dado um par de imagens constituído por uma imagem teste e o padrão ouro, a medida *PRI* conta a fração dos pares de *pixels* cujos rótulos são consistentes entre duas partições. A literatura reporta que a medida *PRI* tende a correlacionar-se melhor, em relação as demais medidas aqui estudadas, com a segmentação manual feita por humanos (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005).

Suponha um conjunto de segmentações manuais<sup>1</sup>  $\{S_k \mid k = 1, \dots, K\}$  de uma dada imagem  $I$  com  $M \times N$  pixels e, suponha que  $S_{teste}$  seja uma segmentação da imagem  $I$  (originada por um sistema digital, por exemplo) que desejamos comparar com o conjunto das segmentações manuais  $\{S_1, S_2, \dots, S_K\}$ . A medida de performance  $PRI$  pode ser calculada como (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005):

$$PRI(S_{teste}, S_k) = \frac{1}{\binom{M \times N}{2}} \sum_{\{u,v \mid u \neq v\}} [\bar{c}_{uv} \bar{p}_{uv} + (1 - \bar{c}_{uv})(1 - \bar{p}_{uv})], \quad (41)$$

onde  $\bar{c}_{uv}$  indica que o par de pixels  $(u, v)$  possuem os mesmos rótulos na segmentação teste, isto é  $\bar{c}_{uv} = \hat{T}(\omega_u^{S_{teste}} = \omega_v^{S_{teste}})$ ,  $\hat{T}(\cdot)$  é um operador binário, e  $\bar{p}_{uv} = \hat{T}(\omega_u^{S_k} = \omega_v^{S_k})$  para um dado  $S_k \in \{S_1, S_2, \dots, S_K\}$  (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005).

### 6.1.2 Variation of Information (VOI)

$VOI$  (do inglês, *Variation of Information*) mede a quantidade de informação que é perdida ou ganha entre duas segmentações (a segmentação associada a imagem teste e a segmentação associada com o padrão ouro). É uma medida não negativa, cujos menores valores indicam um melhor resultado de segmentação.

Considere um par de segmentações diferentes  $(S_{teste}, S_k)$ . A medida de performance  $VOI$  pode ser calculada por meio da expressão:

$$VOI(S_{teste}, S_k) = \frac{1}{2} [H(S_{teste}) + H(S_k)] - MI(S_{teste}, S_k), \quad (42)$$

sendo  $H$  a medida de entropia e  $MI$  a medida de informação mútua (VINH; EPPS; BAILEY, ???).

### 6.1.3 Global Consistency Error (GCE)

$GCE$  (do inglês, *Global Consistency Error*) mede a que extensão a imagem segmentada teste consiste em um refinamento (i.e. um subconjunto) em relação a outra segmen-

<sup>1</sup>No âmbito de segmentação de imagens, o padrão-ouro é tipicamente dado por segmentações manuais realizadas por humanos, tal como ocorre com o banco de dados BSD300.

tação (o padrão ouro). A medida  $GCE$  indica a consistência da imagem segmentada.

Matematicamente, consideramos  $S_{teste}$  e  $S_k$  duas segmentações diferentes de uma imagem  $I$  com  $M \times N$  *pixels*, a primeira segmentação originada por um sistema digital (um método computacional), e a segunda originada por uma segmentação manual. Para um dado *pixel*  $I(m, n) = \mathbf{x}_i$ , considere os segmentos (regiões na imagem) que contém  $\mathbf{x}_i$  em  $S_{teste}$  e  $S_k$ . Denota-se estes conjuntos de *pixels* por  $C(S_{teste}, \mathbf{x}_i)$  e  $C(S_k, \mathbf{x}_i)$ , respectivamente. O Erro Local de Refinamento ( $LRE$ , do inglês *Local Refinement Error*) pode ser definido como (MARTIN et al., 2001):

$$LRE(S_{teste}, S_k, \mathbf{x}_i) = \frac{\text{card}(C(S_{teste}, \mathbf{x}_i) \setminus C(S_k, \mathbf{x}_i))}{\text{card}(C(S, \mathbf{x}_i))}, \quad (43)$$

onde  $\setminus$  denota o operador complemento<sup>2</sup>.

É possível observar que a Equação 43 é não simétrica, isto é, produz diferentes resultados uma vez que o operador complemento é não simétrico. A medida de performance  $GCE$  consiste em uma alternativa à medida  $LRE$  que é simétrica e pode ser computada por meio da expressão (MARTIN et al., 2001):

$$GCE(S_{teste}, S_k) = \frac{1}{M \times N} = \min\left\{ \sum_{i=1}^{M \times N} LRE(S_{teste}, S_k, \mathbf{x}_i), \sum_{i=1}^{M \times N} LRE(S_k, S_{teste}, \mathbf{x}_i) \right\}. \quad (44)$$

Um menor valor da medida  $GCE$  indica uma maior similaridade entre as segmentações (UNNIKRISHNAN; PANTOFARU; HEBERT, 2007).

#### 6.1.4 *Boundary Displacement Error (BDE)*

$BDE$  (do inglês, *Boundary Displacement Error*) mede o erro de deslocamento médio de *pixels* localizados em bordas entre duas imagens (imagem teste e padrão ouro) (FREIXENET et al., 2002).

Seja  $B_{teste}$  um conjunto de *pixels* localizados nas bordas da segmentação  $S_{teste}$  e seja  $B_k$  um conjunto de *pixels* localizados nas bordas da segmentação manual  $S_k$ , denota-se uma distribuição de distâncias partindo de  $B_{teste}$  até  $B_k$  como  $D_{B_{teste}}^{B_k}$ . Esta distribuição

<sup>2</sup>Def:  $A \setminus B := \{z \in A \mid z \notin B\}$ .

caracteriza a discrepância, medida pela distância, partindo de  $B_{teste}$  até  $B_k$ . A distância de um ponto arbitrário  $z \in B_{teste}$  até  $B_k$  é definida como a distância absoluta mínima entre  $z$  e todos os outros pontos em  $B_k$ , isto é (FREIXENET et al., 2002):

$$\hat{d}(z, B_k) = \min\{\hat{d}_E(z, w)\}, \forall w \in B_k, \quad (45)$$

onde  $\hat{d}_E$  denota a distância Euclideana entre os pontos  $z$  e  $w$ . A discrepância entre os *pixels* pertencentes ao conjunto  $B_{teste}$  e os *pixels* pertencentes ao conjunto  $B_k$  é descrita pela forma da distribuição  $D_{B_{teste}}^{B_k}$ , que tipicamente é caracterizada pela sua média ( $\mu_{D_{B_{teste}}^{B_k}}$ ) (FREIXENET et al., 2002). A medida de performance *BDE* pode ser calculada por meio da expressão:

$$BDE(D_{B_{teste}}^{B_k}, D_{B_k}^{B_{teste}}) = \frac{\mu_{D_{B_{teste}}^{B_k}} + \mu_{D_{B_k}^{B_{teste}}}}{2}. \quad (46)$$

Como regra, define-se que quanto menor o valor da medida *BDE*, maior é a similaridade entre a segmentação testada e a segmentação manual (FREIXENET et al., 2002).

## 6.2 Configuração Experimental e Efeitos na Segmentação

Como já dito no Capítulo 5, o método *ABT* proposto, aplicado ao problema de segmentação de imagens, pode ser configurado por meio da definição de quatro parâmetros diferentes: inicialização ( $\mathcal{I}$ ); o procedimento de extração de características ( $\mathcal{FV}$ ); o limiar de similaridade  $\Phi$ ; e a medida de distância  $\mathcal{D}$ .

O parâmetro de inicialização define o primeiro nível ( $I^1$ ) de decomposição da imagem original ( $I^0$ ) (ver Figura 16). Anteriormente, no Capítulo 5, nós configuramos o parâmetro  $\mathcal{I}$  como sendo o algoritmo *SLIC*. A fim de comparar diferentes configurações para o parâmetro  $\mathcal{I}$ , definimos três possíveis estados de configuração para este parâmetro: o algoritmo *SLIC* (ACHANTA et al., 2012), o algoritmo de *superpixels* proposto por Mori (MORI et al., 2004) e o método *Mean Shift* (COMANICIU; MEER, 2002).

Estes inicializadores foram definidos pois apresentaram qualidades notáveis em relação a métodos concorrentes em testes preliminares. O método *SLIC*, por exemplo,

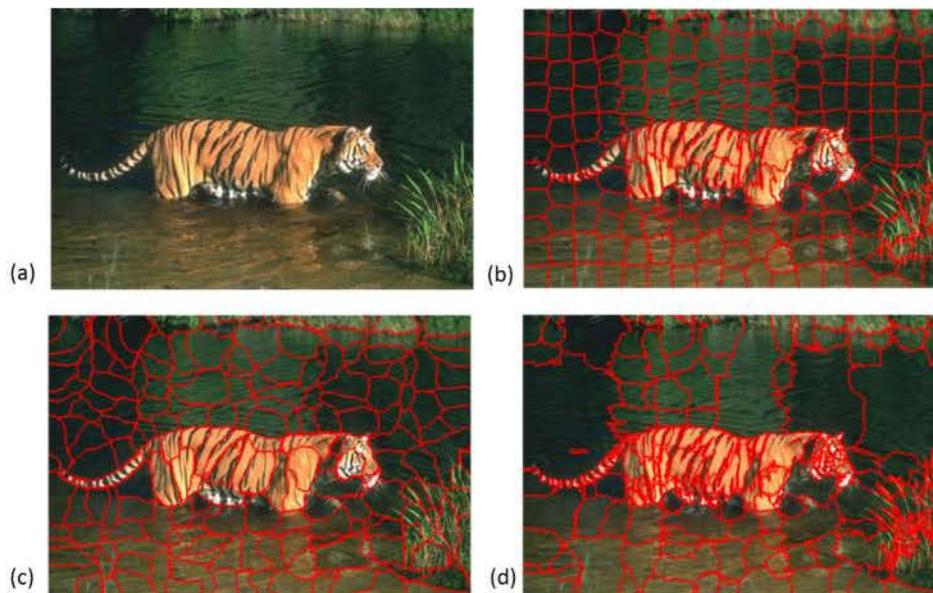


Figura 17: Exemplo de resultado utilizando três tipo de técnicas de inicialização diferentes. Em (a) apresenta-se a imagem original. Em (b) apresenta-se a segmentação obtida mediante a aplicação do algoritmo *SLIC* de *superpixels*. Em (c) apresenta-se a segmentação obtida mediante a aplicação do algoritmo de *superpixels* proposto por Malik. Em (d) apresenta-se a segmentação obtida mediante a aplicação do método Mean Shift.

apresenta uma inicialização rápida, robusta (regiões contendo *pixels* similares são obtidas mesmo para imagens com muitas inomogeneidades) e flexível (pode-se controlar o número de regiões iniciais). O uso do *Mean Shift* origina regiões de alta similaridade por meio da busca de regiões de alta densidade de dados no espaço de feições de forma não paramétrica. O algoritmo proposto por Malik, embora possua um custo computacional maior que o método *SLIC*, se diferencia do mesmo obtendo contornos mais suaves de *superpixels*. Na Figura 17 é possível observar a pré-segmentação obtida por meio da aplicação destes três diferentes métodos.

Em nossa busca pela otimização dos parâmetros de configuração do método proposto, delimitamos 200 regiões iniciais para ambos algoritmos de *superpixel*<sup>3</sup>, e para abordagem *Mean Shift*, fora definido a largura de banda espacial  $h_s = 7$  e a largura de banda referente ao espaço de feições  $h_r = 15$ . Embora o número de regiões seja o mesmo

<sup>3</sup>Nesta dissertação, o número de regiões iniciais para os algoritmos de *superpixel* foi definido por meio da condução de testes preliminares no banco de dados BSD300. Empiricamente verificamos que definindo 200 regiões iniciais, nosso método tende a apresentar maiores resultados para a média da medidade de performance *PRI*.

para ambos os algoritmos de *superpixel*, o tamanho e a forma das regiões originadas por esses métodos podem variar bastante. Em contraste, o método *Mean Shift* determina o número de regiões iniciais de forma automática, de acordo com seu próprio critério de convergência (COMANICIU; MEER, 2002).

Cada estado de configuração, portanto, origina diferentes regiões iniciais que influenciam o procedimento de extração de feições diretamente. Um exemplo do efeito qualitativo do método de inicialização pode ser visto na Figura 18. As segmentações obtidas na Figura 18 foram obtidas alterando apenas o método de inicialização, deixando fixos o limiar de similaridade  $\Phi = 1.0$ , definindo o vetor característico como sendo a média dos canais de cores  $L^*a^*b^*$ , e fazendo uso da distância Euclideana como a medida de distância. Como pode ser visto na Figura 18, as diferentes formas das regiões iniciais originadas pelos métodos *Mori* (a), *Mean Shift* (b) and *SLIC* (c) irão influenciar a etapa de extração de feições. Por consequência, os resultados finais de segmentação obtidos (imagens (d), (e) e (f) na Figura 18) tendem a originar resultados de segmentação diferentes.

O parâmetro  $\Phi$  e a medida de distância  $\mathcal{D}$ , são utilizados para definir a similaridade dos vetores característicos que, por sua vez, dependem do procedimento de extração de características ( $\mathcal{FV}$ ). Na nossa investigação, nós verificamos o comportamento do método proposto para:  $\Phi \in [0.1, 0.2, \dots, 1.2]$ ; e testamos duas medidas de distância  $\mathcal{D}$ , a norma Euclideana ( $\hat{d}_E$ ) e a distância Bhattacharyya ( $\hat{d}_B$ ) (THEODORIDIS; KOUTROUMBAS, 2008), que podem ser calculadas de forma genérica como:

$$\hat{d}_E(\bar{\mathbf{q}}, \bar{\mathbf{p}}) = \sqrt{\sum_i (\bar{p}_i - \bar{q}_i)^2}, \quad (47)$$

$$\hat{d}_B(\bar{\mathbf{q}}, \bar{\mathbf{p}}) = -\ln \left( \sum_i \sqrt{\bar{q}_i \bar{p}_i} \right). \quad (48)$$

para vetores característicos arbitrários distintos  $\bar{\mathbf{q}}$  e  $\bar{\mathbf{p}}$ .

Tendo fixado uma medida de distancia  $\mathcal{D}$ , para um dado vetor característico  $\mathbf{x}_j^s \in X^s$ , a medida que o valor de  $\Phi$  diminui, o número de vizinhos no espaço de feições de  $\mathbf{x}_j^s$  também diminui. Em contraste, à medida que  $\Phi \rightarrow \infty$ , a vizinhança nos espaço de feições  $\eta_\Phi^{j,s}$  tende a conter todos os vetores característicos disponíveis no espaço de feições. O

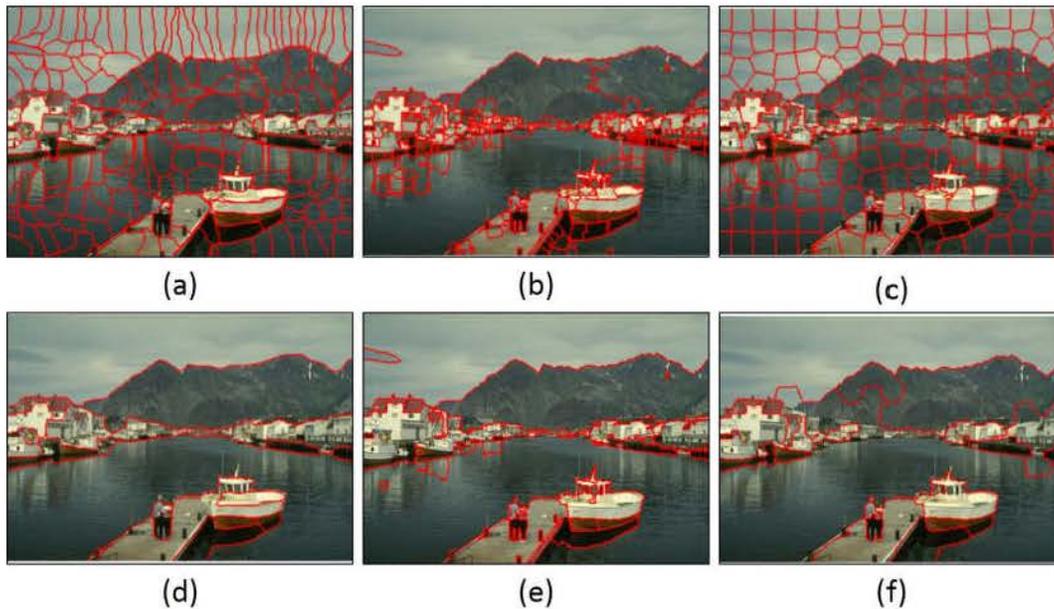


Figura 18: Efeito visual da inicialização no método *ABT*: o algoritmo proposto por Mori, a abordagem *Mean Shift* e o método de inicialização *SLIC* são apresentados em (a), (b) e (c), respectivamente. A segmentação final obtida pelo método *ABT* para o algoritmo proposto por Mori, a abordagem *Mean Shift* e o método *SLIC* são apresentadas em (d), (e) e (f), respectivamente (contornos das regiões em vermelho).

efeito prático de aumentar o valor de  $\Phi$  pode ser observado na Figura 19, que foi obtida fixando: *Mean Shift* como método de inicialização; a média dos canais de cores  $L^*a^*b^*$  como vetor característico; e a distância Euclidiana como medida de similaridade. Na Figura 19 (a), (b) e (c), o parâmetro  $\Phi$  foi definido como 0.1, 0.5 e 1.0 respectivamente.

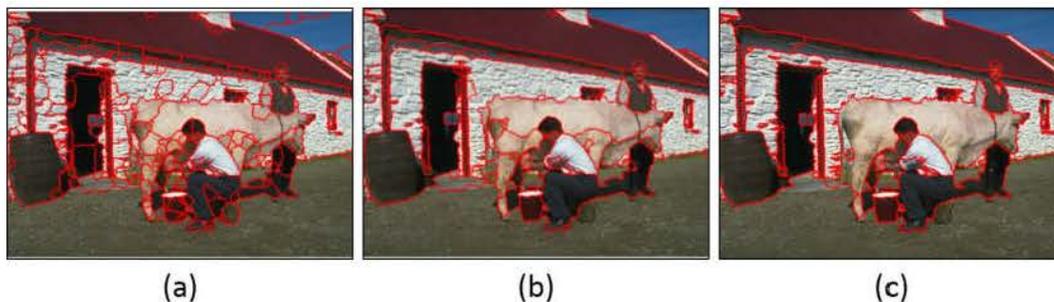


Figura 19: Efeito visual do aumento do parâmetro  $\Phi$ . Foram configurados os parâmetros: *Mean Shift* como o método de inicialização; a média dos canais de cores  $L^*a^*b^*$  como o vetor característico; e a norma Euclidiana como a medida de similaridade: (a) Segmentação com  $\Phi = 0.1$ , (b) Segmentação com  $\Phi = 0.5$ , (c) Segmentação com  $\Phi = 1$  (contornos das regiões em vermelho).

Com o intuito de avaliar o efeito do procedimento de extração de características ( $\mathcal{FV}$ )

no resultado de segmentação, comparamos: características de cor média, nos espaços de cores RGB e L\*a\*b\*; e texturas, incluindo histogramas de cor como descritores de textura.

Como descrito no Capítulo 5, vetores característicos de cor  $\mathbf{x}_j$ , associados às regiões  $R_j$ , são definidos como o valor médio das características de cor dos *pixels* em cada região de imagem. Portanto, há duas configurações possíveis explorando características de cor média: vetores com as médias de cor RGB e vetores com as médias de cores L\*a\*b\*.

Também avaliamos nosso método configurando o parâmetro  $\mathcal{FV}$  como um procedimento de extração de texturas. Neste trabalho, o descritor de textura adotado para comparação consiste em uma mistura de histogramas de cores RGB requantizados. Em tal modelo, o espaço de cores RGB é dividido em  $N_b = \beta^3$  cubos (intervalos) com arestas ( $\beta$ ) de mesmo tamanho, e contam-se quantos vetores de cor RGB associados a *pixels* em uma região da imagem encontram-se dentro de cada cubo. Matematicamente, denotamos  $b(\mathbf{x}_i) \in \{1, 2, \dots, N_b\}$  como sendo o índice do intervalo associado ao vetor de cor  $\mathbf{x}_i = [R_i^*, G_i^*, B_i^*]$  da imagem  $I \in \mathbb{R}^{M \times N \times 3}$ , com  $i = 1, 2, \dots, M \times N$ . Posto isto, o descritor  $\mathbf{h}_j = [h_{j,1}, h_{j,2}, \dots, h_{j,N_b}]$  associado a região  $R_j$  é dado, para  $n = 1, \dots, N_b$ , por meio do procedimento de contagem:

$$h_{j,n} = \frac{1}{\text{card}(R_j)} \sum_{\forall \mathbf{x}_i \in R_j} \delta(b(\mathbf{x}_i) - n), \quad (49)$$

onde  $\frac{1}{\text{card}(R_j)}$  é um termo de normalização e  $\delta$  é a função delta de Kronecker delta function. Este procedimento origina um vetor característico descritor de textura  $\mathbf{h}_j$  de dimensionalidade  $\beta^3$  para cada região da imagem  $R_j$ . Em experimentos preliminares, foi observado que fixando  $\beta = 7$  obtém-se maiores valores para a média da medida de performance *PRI*.

### 6.3 Ajuste de Parâmetros e Análise de Variâncias

Nós conduzimos um projeto de experimentos do tipo fatorial completo, para avaliar o comportamento do nosso método sob diferentes configurações. O projeto de experimentos

Tabela 1: Níveis dos Fatores Controláveis.

$\Phi$	$\mathcal{I}$	$\mathcal{FV}$	$\mathcal{D}$
0.1	S-MS-M	L-R-H	B-E
0.2	S-MS-M	L-R-H	B-E
0.3	S-MS-M	L-R-H	B-E
0.4	S-MS-M	L-R-H	B-E
0.5	S-MS-M	L-R-H	B-E
0.6	S-MS-M	L-R-H	B-E
0.7	S-MS-M	L-R-H	B-E
0.8	S-MS-M	L-R-H	B-E
0.9	S-MS-M	L-R-H	B-E
1.0	S-MS-M	L-R-H	B-E
1.1	S-MS-M	L-R-H	B-E
1.2	S-MS-M	L-R-H	B-E

possui quatro fatores controláveis: o método de inicialização ( $\mathcal{I}$ ) (com três níveis: *SLIC* (S), *Mori* (M) and *Mean Shift* (MS)); o vetor característico ( $\mathcal{FV}$ ) (com três níveis: L\*a\*b\* (L), RGB (R), Histograma dos canais RGB (H)); a medida de distância ( $\mathcal{D}$ ) para definição da vizinhança no espaço de feições (com dois níveis: Euclideana (E) e Bhattacharyya (B)); e o parâmetro  $\Phi$  (com doze níveis:  $\Phi = \{0.1, \dots, 1.2\}$ ). Nós definimos a variável de resposta como sendo a medida *PRI*. A Tabela 1 apresenta os quatro fatores controláveis definidos com seus respectivos níveis.

Foram selecionadas 40 imagens pertencentes ao conjunto de treino do banco de dados BSD300 de forma aleatória. Além disso, atendendo aos requisitos de um projeto de experimentos do tipo fatorial completo (MONTGOMERY, 2008), o experimento foi conduzido de forma totalmente aleatória. Cada imagem selecionada para o experimento, foi segmentada três vezes (três repetições), e para cada uma foi introduzido um ruído Gaussiano com média igual a zero e variância igual a 0.0001.

O projeto de experimentos realizados é descrito pelo seguinte modelo (MONTGO-

Tabela 2: Tabela ANOVA para o projeto de experimentos realizado.

Fonte Var.	Soma Qua.	G.L.	Média Qua.	$F_{calc}$	$F_{tab}$	Conclusão
$\mathcal{I}$	0,49024	2	0,24512	13883,89	6.9077	Significativo
$\mathcal{FV}$	0,27944	2	0,13972	7913,931	6.9077	Significativo
$\mathcal{D}$	0,10025	1	0,100247	5678,113	10.8275	Significativo
$\Phi$	0,06386	11	0,005806	328,8506	2.9588	Significativo
$\mathcal{I}*\mathcal{FV}$	0,07599	4	0,018998	1076,053	4.6167	Significativo
$\mathcal{I}*\mathcal{D}$	0,04804	2	0,02402	1360,55	6.9077	Significativo
$\mathcal{I}*\Phi$	0,01965	22	0,000893	50,58734	2.2657	Significativo
$\mathcal{FV}*\mathcal{D}$	0,14729	2	0,073646	4171,392	6.9077	Significativo
$\mathcal{FV}*\Phi$	0,06982	22	0,003173	179,7478	2.2657	Significativo
$\mathcal{D}*\Phi$	0,03579	11	0,003254	184,3082	2.9588	Significativo
$\mathcal{I}*\mathcal{FV}*\mathcal{D}$	0,06867	4	0,017167	972,3641	4.6167	Significativo
$\mathcal{I}*\mathcal{FV}*\Phi$	0,02033	44	0,000462	26,16605	1.8350	Significativo
$\mathcal{I}*\mathcal{D}*\Phi$	0,02095	22	0,000952	53,94823	2.2657	Significativo
$\mathcal{FV}*\mathcal{D}*\Phi$	0,03342	22	0,001519	86,03557	2.2657	Significativo
$\mathcal{I}*\mathcal{FV}*\mathcal{D}*\Phi$	0,01699	44	0,000386	21,87297	1.8350	Significativo
Erro	0,00763	432	1,77E-05	-	-	
Total	1,49836	647	-	-	-	

MERY, 2008):

$$\begin{aligned} \bar{y}_{abcd r} = & \bar{\mu} + \alpha_a + \beta_b + \gamma_c + \nu_d + (\alpha\beta)_{ab} + (\alpha\gamma)_{ac} + (\alpha\nu)_{ad} + (\beta\gamma)_{bc} + \\ & (\beta\nu)_{bd} + (\gamma\nu)_{cd} + (\alpha\beta\gamma)_{abc} + (\alpha\beta\nu)_{abd} + (\alpha\gamma\nu)_{acd} + (\beta\gamma\nu)_{bcd} + \\ & (\alpha\beta\gamma\nu)_{abcd} + \epsilon_{abcd r}, \end{aligned} \quad (50)$$

onde  $\bar{y}$  é uma observação da variável de resposta. O índice  $a$  representa o nível  $a$  do fator controlável  $\Phi$ ,  $a = 0.1, 0.2, \dots, 1.2$  (total de 12 níveis); o índice  $b$  representa o nível  $b$  do fator controlável  $\mathcal{I}$  (total de 3 níveis); o índice  $c$  representa o nível do fator controlável ( $\mathcal{FV}$ ) (total de 3 níveis); o índice  $d$  representa o nível  $d$  do fator controlável  $\mathcal{D}$  (total de 2 níveis); e o índice  $r$  representa o número de repetições (total de 3 repetições).  $\bar{\mu}$  é a média global;  $\alpha_a, \beta_b, \gamma_c, \nu_d$  representam, respectivamente, os efeitos dos níveis  $a, b, c, d$  dos fatores controláveis  $\Phi, \mathcal{I}, \mathcal{FV}, \mathcal{D}$  na variável de resposta  $\bar{y}$ . Analogamente, os termos  $(\alpha\beta)_{ab}, (\alpha\gamma)_{ac}, \dots, (\alpha\beta\gamma\nu)_{abcd}$ , da Equação 50, representam os efeitos da combinação dos níveis dos fatores controláveis na variável de resposta. O termo  $\epsilon_{abcd r}$  corresponde aos

erros aleatórios da medida. Tais erros são normalmente distribuídos e possuem variância constante.

Nós realizamos uma análise de variâncias (ANOVA) no experimento descrito. Nós verificamos se a hipótese  $H_0$  (MONTGOMERY, 2008) era verdadeira, assumindo a distribuição Fisher-Snedecor entre os tratamentos. Tal tarefa pode ser realizada comparando o valor  $F$  calculado com o correspondente valor encontrado na tabela da distribuição Fischer-Snedecor, respectivamente referenciados, na Tabela 2, como  $F_{calc}$  e  $F_{tab}$ . Em seguida, verifica-se para cada valor de  $F$  computado, se  $F_{calc}$  é maior do que o seu valor correspondente tabelado  $F_{tab}$ . Se é maior, afirmamos que a fonte de variância (i.e. o fator controlável, ou a combinação dos fatores controláveis) é significativa. Se não é maior, afirmamos que a fonte de variância não é significativa. A primeira afirmação implica que a fonte de variância altera estatisticamente a variável de resposta, enquanto a segunda implica que a fonte de variância não altera estatisticamente a variável de resposta. A Tabela 2 apresenta a típica tabela ANOVA, computada assumindo a medida  $PRI$  como sendo a variável de resposta, com a conclusão para hipótese  $H_0$ .

A Tabela 2 revela que  $\mathcal{I}$ ,  $\mathcal{FV}$ ,  $\mathcal{D}$ ,  $\Phi$  alteram significativamente a variável de resposta individualmente, assim como a combinação dos fatores controláveis  $\mathcal{I}*\mathcal{FV}*\mathcal{D}*\Phi$ . Isto implica que a otimização destes parâmetros em termos da medida de performance  $PRI$  deve ser realizada de forma correlacionada.

A Figura 20 apresenta o comportamento do método proposto para todas as possíveis configurações dos estados dos parâmetros  $\mathcal{I}$ ,  $\mathcal{FV}$ ,  $\Phi$  e  $\mathcal{D}$ . Nossos testes revelaram que o conjunto de parâmetros que maximizam a média da medida  $PRI$  para o conjunto de dados BSD300 é alcançado quando:  $\Phi = 1.0$ ; o método de inicialização ( $\mathcal{I}$ ) é a abordagem *Mean Shift*; o procedimento de extração de vetores característicos ( $\mathcal{FV}$ ) é dado pela média dos canais de cores  $L*a*b$  das regiões iniciais; e a medida de distância ( $\mathcal{D}$ ) é dado pela norma Euclideana.

Como previamente mencionado na Seção 6.2, o método *Mean Shift* depende de dois parâmetros: a largura de banda do espaço de feições ( $hr$ ) e largura de banda espacial ( $hs$ ). De modo a explorar os efeitos desses dois parâmetros internos no método *ABT*, nós

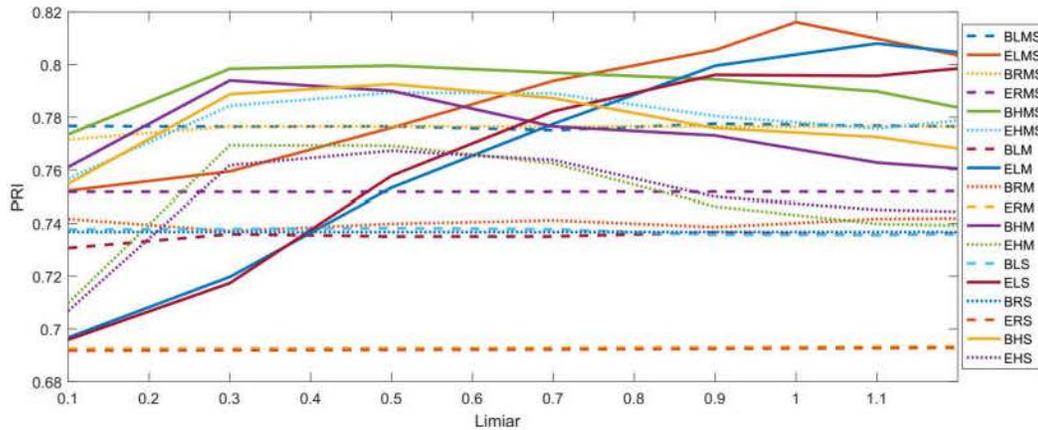


Figura 20: Valores médios de  $PRI$  computados para o banco de dados BSD300 para todas as possíveis configurações testadas em função do parâmetro limiar  $\Phi \in \{0.1, \dots, 1.2\}$ . BLMS: Bhattacharyya- $L^*a^*b^*$ -Mean Shift; ELMS: Euclidean- $L^*a^*b^*$ -Mean Shift; BRMS: Bhattacharyya- $R^*G^*B^*$ -Mean Shift; ERMS: Euclidean- $R^*G^*B^*$ -Mean Shift; BHMS: Bhattacharyya-RGB Histogram-Mean Shift; EHMS: Euclidean-RGB Histogram-Mean Shift; BLM: Bhattacharyya- $L^*a^*b^*$ -Mori; ELM: Euclidean- $L^*a^*b^*$ -Mori; BRM: Bhattacharyya- $R^*G^*B^*$ -Mori; ERM: Euclidean- $R^*G^*B^*$ -Mori; BHM: Bhattacharyya- $R^*G^*B^*$  Histogram-Mori; EHM: Euclidean- $R^*G^*B^*$  Histogram-Mori; BLS: Bhattacharyya- $L^*a^*b^*$ -SLIC; ELS: Euclidean- $L^*a^*b^*$ -SLIC; BRS: Bhattacharyya- $R^*G^*B^*$ -SLIC; ERS: Euclidean- $R^*G^*B^*$ -SLIC; BHS: Bhattacharyya- $R^*G^*B^*$  Histogram-SLIC; EHS: Euclidean- $R^*G^*B^*$  Histogram-SLIC.

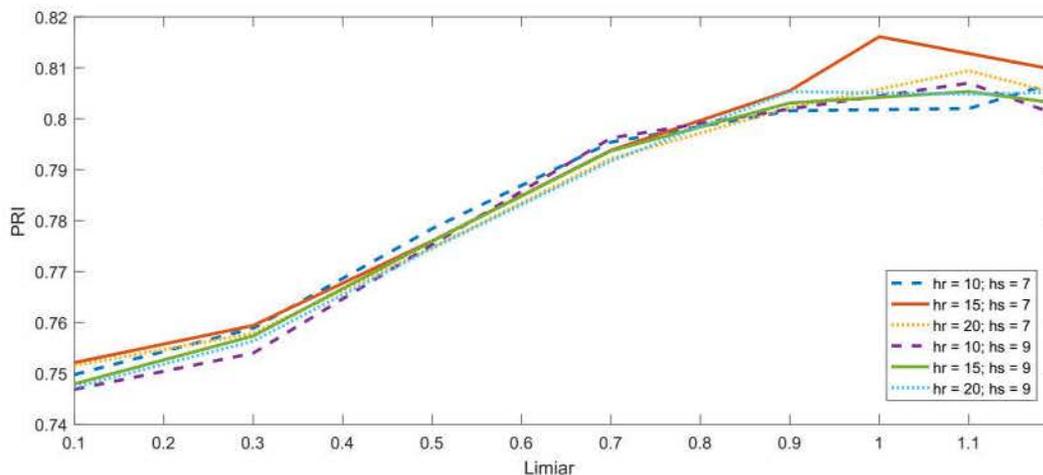


Figura 21: Comportamento da média da medida  $PRI$ , para o banco de dados BSD300, em função de  $\Phi \in \{0.1, \dots, 1.2\}$ . Os parâmetros do método *Mean Shift* foram testados para os intervalos  $hs \in [7, 9]$  e  $hr \in [10, 15, 20]$ . Os demais parâmetros foram fixados como: a média da cor  $L^*a^*b^*$  como vetor característico  $x$ , e como medida de distância a norma Euclideana.

os testamos definindo diferentes configurações. A Figura 21 apresenta o comportamento da média da medida  $PRI$  para os intervalos  $hs \in [7, 9]$  e  $hr \in [10, 15, 20]$ . Como pode ser observado na Figura 21, os parâmetros que maximizam o valor da medida  $PRI$  ocorrem quando  $hs = 7$  e  $hr = 15$ .

## 6.4 Resultados Quantitativos

Na Seção anterior (Seção 6.3), utilizando os resultados da ANOVA, nós encontramos, para o método  $ABT$ , a configuração ótima dos parâmetros que maximizam a média da medida de performance  $PRI$  para o conjunto de dados BSD300. De modo a detalhar o comportamento do nosso método sob esta configuração, calculamos também as medidas de performance  $VOI$ ,  $GCE$  e  $BDE$ .

As Figuras 22, 23, 24 e 25, apresentam o comportamento do método proposto, utilizando o banco de dados BSD300, em termos das medidas de performance  $PRI$ ,  $VOI$ ,  $GCE$  e  $BDE$ , respectivamente. Os resultados dispostos nas Figuras 22, 23, 24 e 25, mostram o limiar  $\Phi \in \{0.1, \dots, 1.2\}$  variando, com a seguinte configuração fixa: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos.

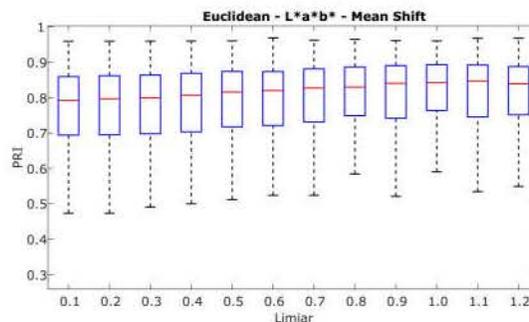


Figura 22: Valores médios de  $PRI$  em função do limiar  $\Phi \in \{0.1, \dots, 1.2\}$  para a configuração: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos.

Com a finalidade de detalhar o comportamento do nosso método sob a sua configuração ótima, apresentamos também a Tabela 3. Utilizando esta tabela, é possível observar que quando  $\Phi = 1.0$ , não só a média da medida de performance  $PRI$  é a maior para o

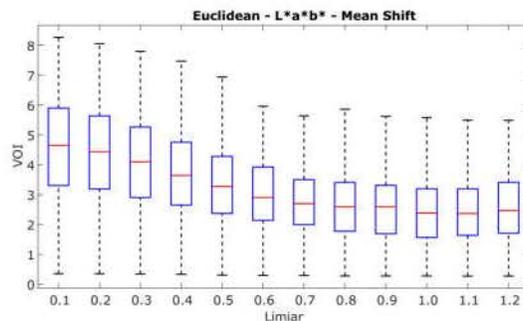


Figura 23: Valores médios de  $VOI$  em função do limiar  $\Phi \in \{0.1, \dots, 1.2\}$  para a configuração: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos.

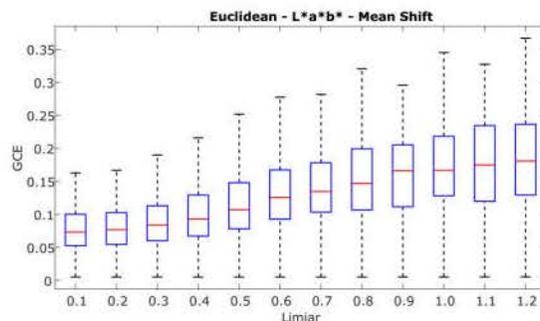


Figura 24: Valores médios de  $GCE$  em função do limiar  $\Phi \in \{0.1, \dots, 1.2\}$  para a configuração: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos.

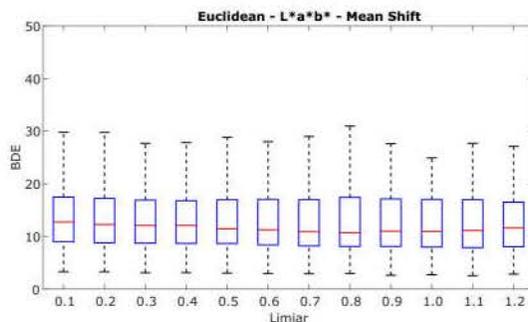


Figura 25: Valores médios de  $BDE$  em função do limiar  $\Phi \in \{0.1, \dots, 1.2\}$  para a configuração: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos.

intervalo de  $\Phi$  investigado (média  $PRI = 0.81480$ , com desvio padrão de  $0.10752$ ), como também a média da medida de performance  $VOI$  é a menor para o mesmo intervalo (média  $VOI = 2.4702$  com desvio-padrão de  $1.0971$ ).

Alguns resultados qualitativos obtidos pelo método *ABT* sob a sua configuração ótima, são apresentados na Figura 26. Nós apresentamos nossos resultados de segmentação dividindo as imagens em três categorias: cenários que apresentam construções na Figura 26 (a); paisagens naturais Figura 26 (b); e imagens que contenham animais Figura 26 (c).

Norma Euclideana - $L^*a^*b^*$ - <i>Mean Shift</i>												
<i>PRI</i>	Limiar ( $\Phi$ )											
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
Mínimo	0.29238	0.29299	0.29353	0.29475	0.29827	0.31466	0.32457	0.39847	0.44394	0.52041	0.44430	0.41296
Mediana	0.79198	0.79660	0.79981	0.80700	0.81571	0.81982	0.82712	0.82956	0.84041	0.84224	0.84672	0.83927
Média	0.75212	0.75410	0.75941	0.76677	0.77603	0.78680	0.79376	0.80323	0.80549	0.81480	0.80954	0.80355
Desvio-padrão	0.14894	0.14876	0.14828	0.14604	0.14057	0.13207	0.12877	0.12019	0.12063	0.10752	0.11476	0.11611
Máximo	0.95908	0.95918	0.95949	0.95988	0.96006	0.96876	0.96205	0.96436	0.96074	0.96007	0.96754	0.96775
<i>VOI</i>	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
Mínimo	0.3575	0.3575	0.3470	0.3360	0.3079	0.2992	0.2944	0.2822	0.2822	0.2806	0.2758	0.2774
Mediana	4.6518	4.4348	4.0982	3.6437	3.2750	2.9062	2.7032	2.5924	2.5953	2.3824	2.3693	2.4658
Média	4.6741	4.4872	4.1443	3.7700	3.4029	3.0828	2.8501	2.6962	2.6050	2.4702	2.4764	2.5391
Desvio-padrão	1.7622	1.7145	1.6333	1.5294	1.3870	1.2534	1.1669	1.1338	1.1182	1.0971	1.0671	1.0692
Máximo	8.2664	8.0486	7.7977	7.4672	6.9358	5.9572	5.8582	5.8582	5.6198	5.5777	5.4952	5.4891
<i>GCE</i>	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
Mínimo	0.0052	0.0053	0.0053	0.0052	0.0052	0.0052	0.0052	0.0052	0.0053	0.0053	0.0054	0.0054
Mediana	0.0733	0.0769	0.0841	0.0929	0.1073	0.1257	0.1348	0.1468	0.1661	0.1667	0.1749	0.1810
Média	0.0832	0.0861	0.0924	0.1029	0.1161	0.1303	0.1432	0.1547	0.1621	0.1701	0.1764	0.1834
Desvio-padrão	0.0447	0.0451	0.0474	0.0506	0.0530	0.0560	0.0629	0.0685	0.0644	0.0670	0.0707	0.0727
Máximo	0.2465	0.2465	0.2479	0.2598	0.2657	0.2775	0.2925	0.3485	0.2959	0.3451	0.3274	0.3669
<i>BDE</i>	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2
Mínimo	3.3080	3.3079	3.1103	3.1251	3.0493	2.9565	2.9331	2.9632	2.6400	2.7285	2.5559	2.8774
Mediana	12.753	12.283	12.101	12.102	11.476	11.277	10.914	10.736	11.019	10.972	11.145	11.633
Média	14.370	14.206	13.960	13.681	13.390	13.263	13.037	12.929	12.934	12.883	13.051	12.908
Desvio-padrão	6.8836	6.8617	6.8854	6.7823	6.6280	6.6083	6.5404	6.4686	6.5855	6.9651	7.3735	6.3286
Máximo	32.755	32.628	32.162	31.756	31.260	31.977	32.691	31.653	32.847	43.123	46.811	33.534

Tabela 3: Comparativo dos resultados obtidos com o método *ABT*, para o banco de dados BSD300, sob a configuração: *Mean Shift* como o método de inicialização, a média dos canais de cor  $L^*a^*b^*$  como o vetor característico e a norma Euclideana como a medida de distância entre pontos em função do limiar  $\Phi \in \{0.1, \dots, 1.2\}$  variando.

## 6.5 Comparação com Outros Métodos

Nas Seções 6.3 e 6.4, nós encontramos, para o método *ABT*, os parâmetros  $\mathcal{I}$ ,  $\mathcal{FV}$ ,  $\Phi$  e  $\mathcal{D}$  que maximizam a média da medida de performance *PRI* calculada para o o banco de dados BSD300. Nesta Seção, nós utilizaremos os valores médios computados para as medidas *PRI*, *VOI*, *GCE* e *BDE* para fins de comparação com outras técnicas disponíveis na literatura.

Nós comparamos o método *ABT* com vinte e um métodos diferentes já publicados e disponíveis na literatura. São estes: gPb-owt-ucm (ARBELAEZ et al., 2011); *Entropy-Weighted Formula Approach Based Multi-criteria Fusion Model* EFA-BMFM (KHELIFI;

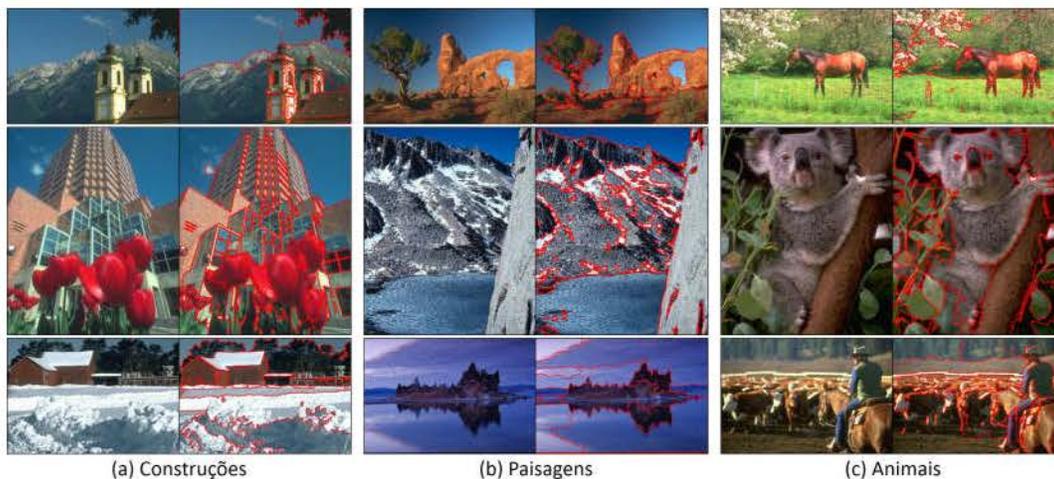


Figura 26: Resultados de segmentação obtidos pelo método *ABT* para diferentes imagens do banco de dados BSD300 (contornos das regiões em vermelho): em (a), são mostradas as segmentações obtidas para imagens que apresentam construções diversas tais como casas, prédios e igrejas; em (b), são mostradas segmentações obtidas para imagens que apresentam algumas paisagens naturais, tais como lagos, montanhas e pedras; em (c), são mostradas segmentações obtidas para imagens que contém animais, tais como bois, um coala, e cavalos.

MIGNOTTE, 2016a); *Image Segmentation Based on Cascade Classifiers and Region Aggregation* (ISCRA) (REN; SHAKHNAROVICH, 2013); *Global Consistency Based Fusion Model* GCEBFM (KHELIFI; MIGNOTTE, 2016b); *Variation of Information based Fusion Model* (VOIBFM) (MIGNOTTE, 2014b); *Texture and Boundary Encoding-based Segmentation* (TBES) (MOBAHI et al., 2011); *Saliency Driven Total Variation Segmentation* (SDTV) (DONOSER et al., 2009); *Model Image Segmentation* (MIS) (KRINIDIS; PITAS, 2009); *Spatially Constrained K-Means* (SCKM) (MIGNOTTE, 2011a); segmentation based on *PRI fusion* (PRIF) (MIGNOTTE, 2008); *Ultrametric Contour Maps* (UCM) (ARBELAEZ et al., 2009); *image Segmentation From a Soft Boundary Map* (SFSBM) (MIGNOTTE, 2014a); *Stochastic Texture Region Merging* (SMTR) (MEDEIROS; SCHARCANSKI; WONG, 2013); image segmentation based on *Multidimensional Scaling* (MD2S) (MIGNOTTE, 2011b); *Mean Shift* (COMANICIU; MEER, 2002); *Felzenszwalb and Huttenlocher graph theoretical approach* (F&H) (FELZENSZWALB; HUTTENLOCHER, 2004); *kernel graph cut method* (KM) (SALAH; MITICHE; AYED, 2011); *Segmentation by Clustering then Labeling* (SCL) (HUANG et al., 2011); *Compression based Texture Merging* (CTM) (YANG et al., 2008); *Stochastic Region Merging*

Tabela 4: Valores médios das medidas de performance *PRI*, *VOI*, *GCE* e *BDE*, para o banco de dados BSD300.

Método	<i>PRI</i>	<i>VOI</i>	<i>GCE</i>	<i>BDE</i>
Proposto	<b>0.8148</b>	2.470	<b>0.1701</b>	12.883
gPb-owt-ucm	0.810	1.650	-	-
EFA-BMFM	0.800	1.870	0.190	8.160
ISCRA	0.810	<b>1.610</b>	-	-
GCEBFM	0.800	2.100	0.190	8.730
VOIBFM	0.810	1.880	0.200	9.300
TBES	0.807	1.705	-	12.681
SDTV	0.776	1.816	0.177	16.240
MIS	0.798	1.935	0.194	<b>7.826</b>
SCKM	0.796	2.110	0.230	10.090
PRIF	0.800	1.970	0.210	8.450
UCM	0.796	1.715	-	10.954
SFSBM	0.790	2.210	0.210	8.870
SNcut	0.780	2.110	-	-
SMTR	0.790	-	-	-
MD2S	0.780	2.360	0.230	10.370
<i>Mean Shift</i>	0.772	2.004	-	13.976
F&H	0.770	2.188	-	14.057
KM	0.765	2.410	-	-
SCL	0.766	2.140	0.182	9.381
CTM	0.7627	2.2035	0.185	9.4902
SRM	0.750	-	-	-

(SRM) (WONG; SCHARCANSKI; FIEGUTH, 2011), *Superpixel Ncut* (SNcut) (WANG et al., 2016); .

A Tabela 4 mostra os resultados obtidos para as medidas de avaliação *PRI*, *VOI*, *GCE* e *BDE* comparadas aos outros vinte e um métodos concorrentes. Destacamos que as medidas de performance mostradas na Tabela 4 foram retiradas das respectivas publicações originais. Entre os algoritmos testados, o método *ABT* obteve um melhor desempenho em respeito a média da medida *PRI* e a média da medida *GCE* ( $PRI = 0.8148$  and  $GCE = 0.1701$ ) para o conjunto de dados BSD300. Com respeito as medidas *VOI* e *BDE*, os métodos *ISCRA* e *MIS* alcançaram melhores desempenhos respectivamente,  $VOI = 1.610$  e  $BDE = 7.826$ .

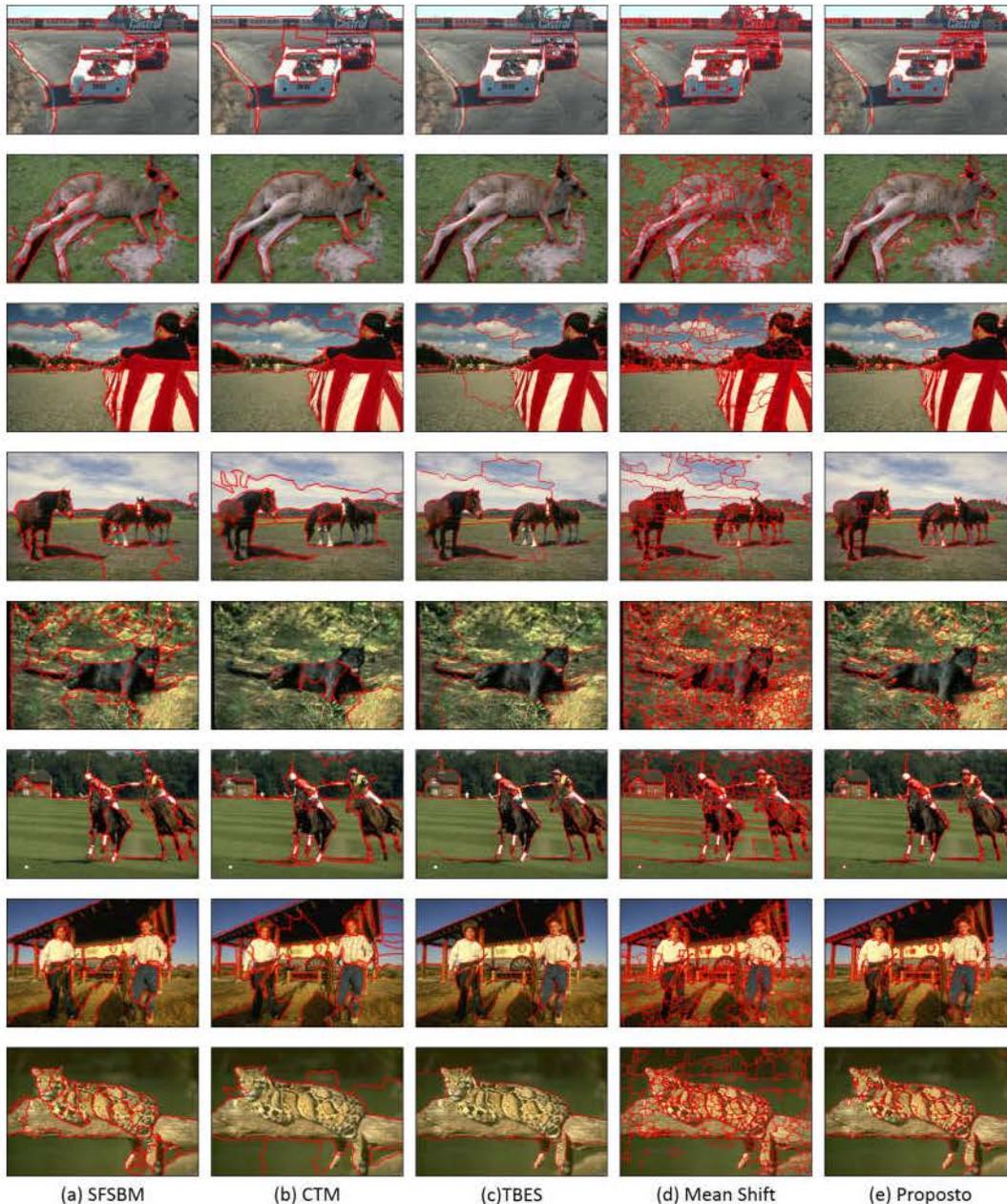


Figura 27: a) Comparação qualitativa do método proposto (e) em relação ao métodos: (a) SFSBM (MIGNOTTE, 2014a); (b) CTM (YANG et al., 2008); (c) TBES (MOBAHI et al., 2011); (d) *Mean Shift* (COMANICIU; MEER, 2002).

Dentre as vinte e uma técnicas utilizadas para comparação das medidas de performance de segmentação, nós selecionamos quatro para comparação visual. Resultados qualitativos são mostrados na Figura 27 para algumas imagens pertencentes ao banco de dados BSD300. Estes métodos foram escolhidos para comparação visual pois estes resultados são públicos e encontram-se disponíveis *on-line*. Particularmente, o método *Mean*

*Shift* foi escolhido para comparação, para que pudéssemos identificar a fusão das regiões que o método *ABT* proposto originou.

## 7 DISCUSSÃO

No Capítulo 6, nós avaliamos os efeitos de diferentes configurações dos parâmetros  $\mathcal{I}$ ,  $\mathcal{FV}$ ,  $\Phi$  e  $\mathcal{D}$  nas segmentações originadas pelo método *ABT*. Utilizando os resultados de uma análise de variâncias, otimizamos a configuração dos parâmetros de modo a maximizar a média da medida de performance *PRI* para o banco de dados BSD300. Os testes conduzidos revelaram que a configuração ótima encontrada ocorre quando fixamos a medida de distância  $\mathcal{D}$  como a norma Euclideana,  $\Phi = 1.0$ , e quando o procedimento de extração de características ( $\mathcal{FV}$ ) é configurado como a média das cores  $L^*a^*b^*$  de regiões inicialmente obtidas (parâmetro  $\mathcal{I}$ ) pela abordagem *Mean Shift*.

Tal como já descrito na Seção 3.3, o método *Mean Shift* consiste em uma técnica não paramétrica, que visa encontrar máximos de densidade em uma distribuição de dados discreta. Esta técnica, baseia-se na análise da função densidade de probabilidade que descreve os dados de uma imagem contidos em um espaço constituído de características espaciais e características de cor. Os pontos de máximo de densidade da distribuição de dados contida neste espaço de feições originará os centróides dos *clusters*, e os pontos de dados próximos a esses máximos de densidade poderão ser agrupados, originando a sobresegmentação de uma dada imagem.

Nós argumentamos que o método das Árvores Bayesianas Adaptativas é similar a abordagem *Mean Shift*, uma vez que *clusters* são formados no espaço de feições, por meio da construção de árvores direcionadas, cujos nodos raízes localizam-se em estimativas Bayesianas dos máximos de densidade. Utilizando a formulação Bayesiana proposta para procurar máximos de densidade, juntamente às restrições de similaridade no espaço de

feições, e utilizando a informação contida no grafo de regiões adjacentes (ver Capítulo 5), formam-se clusters no espaço de feições, e, conseqüentemente, a união de regiões no domínio da imagem. Neste contexto, o limiar de similaridade  $\Phi$  tem um papel análogo a largura de banda  $h_r$ , assim como o grafo de regiões adjacentes possui um papel similar a função que o parâmetro  $h_s$  desempenha na abordagem *Mean Shift*. Sendo assim, nós entendemos que dentre os inicializadores testados, a abordagem *Mean Shift*, combinada a média das cores  $L^*a^*b$  de uma região inicial, tende a funcionar melhor com o método *ABT*, em relação as demais configurações de inicialização e vetores característicos possíveis, pois o método *ABT* continuará buscando os máximos de densidade, tal como ocorre com a técnica *Mean Shift*, de modo a refinar a segmentação da imagem.

No Capítulo anterior (Capítulo 6), nós realizamos uma análise de variâncias em um projeto de experimentos fatorial completo. Para tanto, definimos os fatores controláveis e seus respectivos níveis, tal como apresentado na Tabela 1. É sabido que um experimento do tipo fatorial completo demanda que todos os níveis entre os fatores controláveis sejam combinados. A ANOVA, realizada neste projeto de experimentos, indicou que a combinação dos fatores controláveis, isto é  $\mathcal{L}^*\mathcal{F}\mathcal{V}^*\mathcal{D}^*\Phi$ , e seus respectivos níveis (apresentados na Tabela 1), altera significativamente a variável de resposta (média da medida *PRI*). Conseqüentemente, em nosso processo de otimização de parâmetros, não descartamos as possíveis combinações entre os níveis dos fatores controláveis. Isto justifica porque nós combinamos a medida de distância Bhattacharyya, que é própria para medir distâncias entre histogramas, junto a vetores característicos das cores médias de  $L^*a^*b^*$  e RGB. Da mesma forma, também justifica porque combinamos a medida de distância Euclideana para avaliar a similaridade entre histogramas. O resultado dessas combinações pode ser verificado analisando os resultados indicados na Figura 20.

Observando a Figura 20, nota-se que a distância Bhattacharyya de fato influencia (i.e. tende a funcionar como um critério de similaridade), o valor da média de *PRI* quando o vetor característico de histogramas de textura é fixado (i.e. para as curvas na Figura 20 cujas legendas são: BHMS, BHM e BHS). Em contraste, na Figura 20, esta distância tende a não alterar o valor da média de *PRI* quando fixam-se vetores característicos ba-

seados na média das cores  $L^*a^*b^*$  (i.e. para as curvas na Figura 20 cujas legendas são: BLMS, BLM, BLS). Para estes casos, quando fixa-se o vetor característico como a média dos canais de cores  $L^*a^*b^*$ , a distância Euclideana tende a adequar-se melhor como uma medida de similaridade, visto o comportamento médio da medida *PRI* na Figura 20 (i.e. para as curvas na Figura 20 cujas legendas são: ELMS, ELM, ELS). Posto que a distância Bhattacharyya é uma medida entre distribuições discretas (histogramas), esta observação está em concordância com a literatura conhecida (THEODORIDIS; KOUTROUMBAS, 2008), e explica porque a média dos canais de cores  $L^*a^*b^*$  tende a funcionar melhor com a configuração que define a norma Euclideana como medida de distância.

Ainda observando a Figura 20, nota-se que o comportamento da média da medida *PRI* em função de  $\Phi$ , quando fixa-se o vetor característico como a média das cores RGB para uma região, tende a apresentar uma característica constante, tal como é possível perceber nas curvas da Figura 20 denotadas pelas legendas ERMS, ERM e ERS. Embora a norma Euclideana seja sabidamente apropriada para medir distâncias entre vetores no espaço de cores RGB, isto decorre porque o escopo do intervalo que o limiar de similaridade ( $\Phi$ ) foi avaliado nesta dissertação ( $\Phi \in \{0.1, \dots, 1.2\}$ ) não parece ser apropriado. Tal fato sugere que se estendêssemos o intervalo de avaliação de  $\Phi$ , observaríamos um comportamento para a média da medida *PRI* diferente do observado na Figura 20. Sendo, tipicamente, a escolha de  $\Phi$  um número real positivo, e com base nos dados observados na Figura 20, não podemos descartar a possibilidade de que exista alguma combinação de parâmetros, com a média dos canais de cor RGB fixada como vetor característico (i.e. as combinações denotadas pelas legendas ERMS, ERM e ERS na Figura 20), que para algum intervalo maior que o intervalo aqui investigado (i.e.  $\Phi > 1.2$ ), exista um valor médio de *PRI* maior que o valor ótimo encontrado de  $0.8148 \pm 0.1075$  (ver Tabela 3). Reforçamos, que nesta dissertação, devido ao alto custo computacional de investigar o espaço de parâmetros de forma bruta, fez-se necessário delimitar um escopo de avaliação para o intervalo de  $\Phi$ . O intervalo  $\Phi \in \{0.1, \dots, 1.2\}$  fora definido com base em experimentos preliminares, e que antecederam a geração dos resultados experimentais apresentados no Capítulo 6.

Neste trabalho, a Tabela 3, bem como as Figuras 22, 23, 24, 25, foram construídas

com a finalidade de visualizar com maior detalhamento, e de forma complementar a Figura 20, o comportamento dos nossos resultados. Na Tabela 3, observamos que a a média da medida de performance *PRI* é a maior para o intervalo de  $\Phi$  investigado (média *PRI* = 0.81480, com desvio padrão de 0.10752) e identificamos que a média da medida de performance *VOI* é a menor (média *VOI* = 2.4702 com desvio-padrão de 1.0971) para o mesmo intervalo. Observamos também, na Figura 24, que o comportamento da média da medida de performance *GCE* tende a crescer em função do limiar  $\Phi$ , e que o comportamento da média da medida *BDE* tende a permanecer constante em função do limiar  $\Phi$ .

No que diz respeito aos resultados de comparação com outros métodos, a Tabela 4 revelou que o nosso método obteve um melhor desempenho em relação a média da medida *PRI* e a média da medida *GCE* (*PRI* = 0.8148 and *GCE* = 0.1701) para o conjunto de dados BSD300. Tais dados confirmam a potencialidade do nosso método para resolver o problema de segmentação de imagens coloridas, visto que a literatura reporta que altos valores (valores próximos a 1) para a medida de performance *PRI* tende a correlacionar-se com a segmentação manual feita por humanos (UNNIKRISHNAN; PANTOFARU; HEBERT, 2005). Entretanto, nota-se também, na Tabela 4, que em relação as medidas *VOI* e *BDE*, nosso método apresentou um desempenho inferior aos demais métodos comparados (*VOI* = 1.610 e *BDE* = 12.883).

Resultados qualitativos obtidos pelo método *ABT* são mostrados nas Figuras 26 e 27. Tal como discutido na Introdução desta dissertação, (Capítulo 1), a comparação visual de resultados de segmentação tende a ser uma tarefa subjetiva (UNNIKRISHNAN; PANTOFARU; HEBERT, 2007). Apesar disso, existem qualidades comuns a segmentação, as quais podem ser discutidas sem maiores desacordos alarmantes, como argumentaremos a seguir. Um ser humano tende a não particionar regiões de um mesmo objeto em uma dada cena. No contexto de Visão Computacional, este fenômeno pode ser interpretado como a não existência de bordas "acentuadas" segmentando objetos em uma imagem. Se um objeto, cuja forma é conhecida e nítida em uma cena, é dividido ao meio devido ao resultado de uma segmentação, este resultado contém uma borda "acentuada". Neste sentido, o

método *ABT* aplicado ao problema de segmentação de imagens coloridas tende a promover resultados competitivos (e, para alguns casos, melhores) em relação aos métodos da literatura utilizados para comparação, tal como apresentado na Figura 27. Além disso, o método *ABT*, tal como esperado, tende a unir as regiões oriundas dos resultados de sobre-segmentação do método *Mean Shift*, alcançando potencialmente melhores segmentações semânticas.

Nota-se que para alguns casos, ambos métodos que fazem uso de características de textura e de cores (CTM e TBES), assim como o método SFSBM que faz uso apenas de características baseadas em cor, tendem a falhar (i.e. identificam-se bordas "acentuadas" cortando objetos) na segmentação de regiões homogêneas tais como a pista (imagem da corrida), a grama (imagem do canguru e imagem dos jogadores de polo), e o segundo plano das imagens da pantera e do felino em um tronco, enquanto o método *ABT* apresenta melhores resultados qualitativos em relação a esse quesito.

Observa-se, porém, que o nosso método também apresenta imperfeições quando há uma variação do gradiente de cor em um mesmo objeto em uma cena, tal como é possível ver na imagem das crianças vestidas de *cowboy*. Nesta imagem, a cor azul claro do céu (à direita da imagem) originado pelo método *Mean Shift*, situa-se distante da região de cor azul escuro no espaço de feições (i.e. a distância Euclideana entre os vetores característicos das regiões é maior que o limiar  $\Phi = 1.0$ ). Isto explica porque estas regiões que compõem o céu não se uniram. De modo a consertar este problema de segmentação em particular, ter-se-ia que definir um valor maior para o parâmetro  $\Phi$ . Assim, os vetores característicos associados a estas regiões tenderiam a se aproximar no espaço de feições e as regiões poderiam se unir no domínio espacial da imagem. Em compensação, o efeito de aumentar o limiar  $\Phi$  para todas as imagens de um mesmo banco de dados, tende a prejudicar a qualidade de outras segmentações. Aumentando o valor do limiar  $\Phi$  para o banco de dados, por exemplo, resultaria na união da pata do canguru com o chão de terra que possui cor similar e encontra-se próxima a ele. Este exemplo justifica a necessidade de otimizar o parâmetro  $\Phi$  de forma global para um banco de dados inteiro, de modo a alcançar bons resultados em termos de coerência semântica e em termos da média da

medida *PRI*, e não apenas para uma única imagem.

## 8 CONCLUSÕES

Neste trabalho nós apresentamos uma nova técnica de *clustering* intitulada Árvore Bayesiana Adaptativa (*ABT*). Nós mostramos como se pode abordar o problema de segmentação de imagens coloridas, fazendo uso da estrutura de dados baseada em árvores direcionadas para representar uma imagem de forma hierárquica.

O método *ABT* proposto, utiliza uma abordagem Bayesiana para procurar máximos de densidade em uma distribuição de dados no espaço de feições, na qual a probabilidade *a posteriori* associada a cada amostra dos dados é modelada por meio de uma mistura de Gaussianas multivariadas. Árvores direcionadas, com raízes localizadas nos máximos de densidade, são construídas no espaço de feições, originando *clusters* de vetores de dados. Sendo assim, em nossa abordagem, *clusters* são descritos por misturas de Gaussianas. Neste trabalho, mostramos que, particularmente, o método *ABT* se diferencia de métodos típicos de estimação de misturas de Gaussianas, devido ao fato de que não são limitados os números de componentes Gaussianas em uma mistura associada a um *cluster*. Nós também mostramos um procedimento iterativo para construção de *clusters*, originando uma estrutura hierárquica para representação de características de uma imagem.

Os experimentos conduzidos nesta dissertação revelaram que o método de *clustering* proposto, aplicado ao problema de segmentação de imagens coloridas, tende a originar melhores resultados para as medidas de performance *PRI* e *GCE*, em comparação a vinte e um métodos publicamente disponíveis na literatura. Nós obtivemos um valor médio de  $PRI = 0.8148$  (um valor mais próximo de 1 representa um melhor resultado) e um valor médio de  $GCE = 0.1701$  (um valor mais próximo de 0 representa um melhor resultado)

para o banco de dados BSD300. Corroborando com os resultados quantitativos, as comparações visuais confirmaram a competitividade da nossa abordagem em relação a outros métodos publicamente disponíveis na literatura.

## 8.1 Trabalhos Futuros

Nas etapas seguintes deste trabalho, pretendemos estender o método das Árvores Bayesianas Adaptativas de modo a segmentar objetos em cenas de vídeos. Devido a natureza estatística do método *ABT*, originalmente projetado para segmentar imagens estáticas, e devido a sua fundamentação Bayesiana, a tarefa desafiadora de segmentar diferentes quadros de um vídeo, preservando a segmentação semântica de objetos em cenas de vídeos no decorrer do tempo, poderá ser abordada, acredita-se, mediante a atualização ótima dos parâmetros (média, covariância e probabilidades a *priori*) do modelo de misturas Gaussianas associados aos objetos de um quadro em um vídeo. Sendo assim, as atualizações dos modelos poderão ser obtidas por meio da adaptação do algoritmo *Estimation Maximization (EM)* recentemente proposto por Luo e colaboradores (LUO; CHAN; NGUYEN, 2016). Além disso, para este fim, deverão ser investigadas outras formas de extração de características discriminantes dos *pixels*, incluindo diversos modelos caracterizantes de texturas.

Outra linha de investigação que merece ser destacada, diz respeito as possíveis aplicações do método de segmentação *ABT*. Alguns exemplos são: segmentação de objetos em cenas de câmeras de vigilância, para fins de segurança, e segmentação de objetos em cenas para posterior reconhecimento dos mesmos em sistemas de *image retrieval*. Particularmente, o método das Árvores Bayesianas Adaptativas, aplicado à segmentação de imagens faciais, apresenta um grande potencial para ser incorporado em sistemas de reconhecimento facial, tal como pode ser exemplificado na Figura 28 (Figura retirada do recente trabalho publicado por Soldera et al. (SOLDERA et al., 2017)). Acredita-se, também, que o método poderia ser utilizado para auxiliar na detecção de ataques a sistemas de reconhecimento facial, problema este conhecido na literatura como *Face Spoofing Detection*.



Figura 28: Ilustração da segmentação de faces de três sujeitos diferentes do banco de dados FEI (THOMAZ; GIRALDI, 2010) (as curvas vermelhas indicam os contornos das regiões segmentadas).

Outro aspecto investigativo, que poderia ser de grande acréscimo para o método *ABT*, diz respeito a escolha do parâmetro  $\Phi$ . Nesta dissertação,  $\Phi$  é escolhido de forma global (i.e.  $\Phi$  é constante para todo o conjunto de dados) de modo a maximizar a medida de performance *PRI*, tal como descrito no Capítulo 5. Acreditamos que seja possível selecionar o parâmetro  $\Phi$  ótimo local (i.e. cada imagem do banco de dados poderá ter um parâmetro  $\Phi$  diferente que maximiza a medida *PRI* para cada imagem do banco de dados), utilizando a fundamentação teórica baseada no critério *Minimum Description Length* (DERKSEN et al., 2007; YANG et al., 2008). Estes trabalhos apresentam uma formulação para codificar (em bits) partições de uma imagem, e sugerem que quanto menor o tamanho da codificação, melhor<sup>1</sup> é a partição da imagem. Utilizando este critério, acreditamos que seja possível obter resultados de segmentação qualitativos ainda mais atrativos em termos semânticos e com maiores valores para a medida *PRI*.

De forma análoga, acreditamos que a fundamentação teórica baseada no *Minimum Description Length* também poderia ser útil para estabelecer um critério ótimo de convergência do nosso algoritmo de *clustering*. Utilizando este critério, poderíamos decidir

<sup>1</sup>Partição possivelmente associada a um alto valor para a medida *PRI*.

pela imagem segmentada que minimiza a codificação das partições, dentre as múltiplas imagens originadas a cada iteração do algoritmo, corrigindo assim, o problema de convergência encontrado em nosso método.

## REFERÊNCIAS

- ACHANTA, R. et al. SLIC superpixels compared to state-of-the-art superpixel methods. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 34, n. 11, p. 2274-2282, 2012.
- AMANCIO, D. R. et al. A systematic comparison of supervised classifiers. **PloS ONE**, San Francisco, v. 9, n. 4, p. e94137, 2014.
- ARBELAEZ, P. et al. Contour detection and hierarchical image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 33, n. 5, p. 898-916, 2011.
- ARBELAEZ, P. et al. From contours to regions: an empirical evaluation. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2009, Miami. **Proceedings...** New York: IEEE, 2009. p. 2294-2301.
- BARBER, D. **Bayesian reasoning and machine learning**. Cambridge, UK: Cambridge University Press, 2012.
- BERNART, E.; SCHARCANSKI, J.; BAMPI, S. Segmentation and classification of melanocytic skin lesions using local and contextual features. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2016, Phoenix. **Proceedings...** New York: IEEE, 2016. p. 2633-2637.
- BERNART, E. E. et al. Unsupervised segmentation and categorization of skin lesions using adaptive thresholds and stochastic features. **Vision Letters**, Waterloo, v. 1, n. 1, p. 1-1, 2015.
- BEZDEK, J. C. **Pattern recognition with fuzzy objective function algorithms**. New York: Springer, 2013.
- BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2006.
- CARVALHO, M. A. Garcia de. **Teoria dos grafos: uma introdução**. Limeira, SP: Universidade Estadual de Campinas, 2005.
- COMANICIU, D.; MEER, P. Mean shift: a robust approach toward feature space analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 24, n. 5, p. 603-619, 2002.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Hingham, v. 20, n. 3, p. 273-297, 1995.

COUR, T.; BENEZIT, F.; SHI, J. Spectral segmentation with multiscale graph decomposition. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2005, San Diego. **Proceedings...** New York: IEEE, 2005. p. 1124-1131.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, New York, v. 13, n. 1, p. 21-27, 1967.

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2005, San Diego. **Proceedings...** New York: IEEE, 2005. p. 886-893.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B**, London, v.39, n. 1, p. 1-38, 1977.

DENG, Y.; DAI, Q.; ZHANG, Z. Graph Laplace for occluded face completion and recognition. **IEEE Transactions on Image Processing**, New York, v. 20, n. 8, p. 2329-2338, 2011.

DOLLAR, P.; TU, Z.; BELONGIE, S. Supervised learning of edges and object boundaries. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2006, New York. **Proceedings...** New York: IEEE, 2006. p. 1964-1971.

DONOSER, M. et al. Saliency driven total variation segmentation. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2009, Kyoto. **Proceedings...** New York: IEEE, 2009. p. 817-824.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. New York: Wiley, 2012.

FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. **International Journal of Computer Vision**, Hingham, v. 59, n. 2, p. 167-181, 2004.

FREIXENET, J. et al. Yet another survey on image segmentation: region and boundary information integration. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2002, Copenhagen. **Proceedings...** London: Springer, 2002. p. 408-422.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In: EUROPEAN CONFERENCE ON COMPUTATIONAL LEARNING THEORY, 1995, Orlando. **Proceedings...** London: Springer, 1995. p. 23-37.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. 2. ed. Berlin: Springer, 2001.

- FU, Z.; WANG, L. Color image segmentation using Gaussian mixture model and EM algorithm. In: WANG F.L. et al. **Multimedia and signal processing**. Shanghai: Springer, 2012. cap. 2, p. 61-66.
- FULKERSON, B.; VEDALDI, A.; SOATTO, S. Class segmentation and object localization with superpixel neighborhoods. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2009, Kyoto. **Proceedings**. . . New York: IEEE, 2009. p. 670-677.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 6, n. 6, p. 721-741, 1984.
- GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. Upper Saddle River: Prentice-Hall, 2002.
- HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: a k-means clustering algorithm. **Journal of the Royal Statistical Society. Series C. Applied Statistics**, London, v. 28, n. 1, p. 100-108, 1979.
- HUANG, R. et al. Image segmentation via coherent clustering in l a b color space. **Pattern Recognition Letters**, New York, v. 32, n. 7, p. 891-902, 2011.
- JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. Upper Saddle River: Prentice-Hall, 1988.
- KHELIFI, L.; MIGNOTTE, M. A new multi-criteria fusion model for color textured image segmentation. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2016, Phoenix. **Proceedings**. . . New York: IEEE, 2016. p. 2579-2583.
- KHELIFI, L.; MIGNOTTE, M. GCE-based model for the fusion of multiples color image segmentations. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2016, Phoenix. **Proceedings**. . . New York: IEEE, 2016. p. 2574-2578.
- KIM, J. et al. A nonparametric statistical method for image segmentation using information theory and curve evolution. **IEEE Transactions on Image Processing**, New York, v. 14, n. 10, p. 1486-1502, 2005.
- KOONTZ, W. L. G.; NARENDRA, P. M.; FUKUNAGA, K. A graph-theoretic approach to nonparametric cluster analysis. **IEEE Transactions on Computers**, New York, v. 100, n. 9, p. 936-944, 1976.
- KRINIDIS, M.; PITAS, I. Color texture segmentation based on the modal energy of deformable surfaces. **IEEE Transactions on Image Processing**, New York, v. 18, n. 7, p. 1613-1622, 2009.
- LI, H.; HE, H.; WEN, Y. Dynamic particle swarm optimization and k-means clustering algorithm for image segmentation. **Optik - International Journal for Light and Electron Optics**, Amsterdam, v. 126, n. 24, p. 4817-4822, 2015.

LIU, L. et al. Fusing sorted random projections for robust texture and material classification. **IEEE Transactions on Circuits and Systems for Video Technology**, New York, v. 25, n. 3, p. 482-496, 2015.

LUO, E.; CHAN, S. H.; NGUYEN, T. Q. Adaptive image denoising by mixture adaptation. **IEEE Transactions on Image Processing**, New York, v. 25, n. 10, p. 4489-4503, 2016.

MA, Y. et al. Segmentation of multivariate mixed data via lossy coding and compression. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 29, n. 9, p. 1546-1562, 2007.

MAIRAL, J. et al. Discriminative sparse image models for class-specific edge detection and image interpretation. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 2008, Marseille. **Proceedings...** London: Springer, 2008. p. 43-56.

MALIK, J. et al. Contour and texture analysis for image segmentation. **International Journal of Computer Vision**, Hingham, v. 43, n. 1, p. 7-27, 2001.

MARTIN, D. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2001, Vancouver. **Proceedings...** New York: IEEE, 2001. p. 416-423.

MARTIN, D. R.; FOWLKES, C. C.; MALIK, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 26, n. 5, p. 530-549, 2004.

MEDEIROS, R.; SCHARCANSKI, J.; WONG, A. Natural scene segmentation based on a stochastic texture region merging approach. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2013, Vancouver. **Proceedings...** New York: IEEE, 2013. p. 1464-1467.

MIGNOTTE, M. A de-texturing and spatially constrained k-means approach for image segmentation. **Pattern Recognition Letters**, New York, v. 32, n. 2, p. 359-367, 2011.

MIGNOTTE, M. A label field fusion model with a variation of information estimator for image segmentation. **Information Fusion**, Amsterdam, v. 20, n. 1, p. 7-20, 2014.

MIGNOTTE, M. A non-stationary mrf model for image segmentation from a soft boundary map. **Pattern Analysis and Applications**, London, v. 17, n. 1, p. 129-139, 2014.

MIGNOTTE, M. MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation. **IEEE Transactions on Neural Networks**, New York, v. 22, n. 3, p. 447-460, 2011.

MIGNOTE, M. Segmentation by fusion of histogram-based-means clusters in different color spaces. **IEEE Transactions on Image Processing**, New York, v. 17, n. 5, p. 780-787, 2008.

MOBAHI, H. et al. Segmentation of natural images by texture and boundary compression. **International Journal of Computer Vision**, Hingham, v. 95, n. 1, p. 86-98, 2011.

MONTGOMERY, D. C. **Design and analysis of experiments**. New York: John Wiley & Sons, 2008.

MORI, G. et al. Recovering human body configurations: combining segmentation and recognition. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2004, Washington, DC. **Proceedings...** New York: IEEE, 2004. p. 1-8.

MÚJICA-VARGAS, D.; GALLEGOS-FUNES, F. J.; ROSALES-SILVA, A. J. A fuzzy clustering algorithm with spatial robust estimation constraint for noisy color image segmentation. **Pattern Recognition Letters**, New York, v. 34, n. 4, p. 400-413, 2013.

MUMFORD, D.; SHAH, J. Optimal approximations by piecewise smooth functions and associated variational problems. **Communications on Pure and Applied Mathematics**, New York, v. 42, n. 5, p. 577-685, 1989.

OMRAN, M.; ENGELBRECHT, A. P.; SALMAN, A. Particle swarm optimization method for image clustering. **International Journal of Pattern Recognition and Artificial Intelligence**, River Edge, v. 19, n. 3, p. 297-321, 2005.

PAL, N. R.; PAL, S. K. A review on image segmentation techniques. **Pattern Recognition**, New York, v. 26, n. 9, p. 1277-1294, 1993.

REBOUÇAS FILHO, P. P. et al. Novel adaptive balloon active contour method based on internal force for image segmentation – a systematic evaluation on synthetic and real images. **Expert Systems with Applications**, Tarrytown, v. 41, n. 17, p. 7707-7721, 2014.

REN, X.; MALIK, J. Learning a classification model for segmentation. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2003, Nice. **Proceedings...** New York: IEEE, 2003. p. 10-17.

REN, X.; FOWLKES, C. C.; MALIK, J. Learning probabilistic models for contour completion in natural images. **International Journal of Computer Vision**, Hingham, v. 77, n. 1-3, p. 47-63, 2008.

REN, X. Multi-scale improves boundary detection in natural images. In: EUROPEAN CONFERENCE ON COMPUTER VISION, Marseille. **Proceedings...** London: Springer, 2008. p. 533-545.

REN, Z.; SHAKHNAROVICH, G. Image segmentation by cascaded region agglomeration. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2013, Oregon. **Proceedings...** New York: IEEE, 2013. p. 2011-2018.

SALAH, M. B.; MITICHE, A.; AYED, I. B. Multiregion image segmentation by parametric kernel graph cuts. **IEEE Transactions on Image Processing**, New York, v. 20, n. 2, p. 545-557, 2011.

- SHI, J.; MALIK, J. Normalized cuts and image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 22, n. 8, p. 888-905, 2000.
- SOLDERA, J.; BEHAINE, C. A. R.; SCHARCANSKI, J. Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. **IEEE Transactions on Instrumentation and Measurement**, New York, v. 64, n. 9, p. 2417-2426, 2015.
- SOLDERA, J. et al. Facial biometrics and applications. **IEEE Instrumentation & Measurement Magazine**, New York, v. 20, n. 2, p. 4-10, 2017.
- TAN, K. S.; ISA, N. A. M.; LIM, W. H. Color image segmentation using adaptive unsupervised clustering approach. **Applied Soft Computing**, Amsterdam, v. 13, n. 4, p. 2017-2036, 2013.
- TAN, K. S.; ISA, N. A. M. Color image segmentation using histogram thresholding–fuzzy c-means hybrid approach. **Pattern Recognition**, New York, v. 44, n. 1, p. 1-15, 2011.
- TAO, W.; JIN, H.; ZHANG, Y. Color image segmentation based on mean shift and normalized cuts. **IEEE Transactions on Systems, Man, and Cybernetics, Part B**, New York, v. 37, n. 5, p. 1382-1389, 2007.
- THEODORIDIS, S. et al. **Introduction to pattern recognition: a matlab approach**. San Diego: Academic Press, 2010.
- THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. San Diego: Academic Press, 2008.
- THOMAZ, C. E.; GIRALDI, G. A. A new ranking method for principal components analysis and its application to face image analysis. **Image and Vision Computing**, New York, v. 28, n. 6, p. 902-913, 2010.
- UNNIKRISHNAN, R.; PANTOFARU, C.; HEBERT, M. A measure for objective evaluation of image segmentation algorithms. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2005, San Diego. **Proceedings...** New York: IEEE, 2005. p. 34-41.
- UNNIKRISHNAN, R.; PANTOFARU, C.; HEBERT, M. Toward objective evaluation of image segmentation algorithms. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 29, n. 6, p. 929-944, 2007.
- VANHAMEL, I.; PRATIKAKIS, I.; SAHLI, H. Multiscale graph theory based color segmentation. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2006, Atlanta. **Proceedings...** New York: IEEE, 2006. p. 769-772.
- VÁSQUEZ, D.; SCHARCANSKI, J.; WONG, A. Stochastic color image segmentation using spatial constraints. In: IEEE INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 2015, Via Giacomo Matteotti. **Proceedings...** New York: IEEE, 2015. p. 35-40.

- VINH, N. X.; EPPS, J.; BAILEY, J. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. **Journal of Machine Learning Research**, Cambridge, MA, v.11, n.1, p. 2837-2854, 2010.
- WANG, X.-Y. et al. A pixel-based color image segmentation using support vector machine and fuzzy c-means. **Neural Networks**, Upper Saddle River, v. 33, n. 1, p. 148-159, 2012.
- WANG, Y. et al. Efficient superpixel based segmentation for food image analysis. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, 2016, Phoenix. **Proceedings...** New York: IEEE, 2016. p. 2544-2548.
- WONG, A.; SCHARCANSKI, J.; FIEGUTH, P. Automatic skin lesion segmentation via iterative stochastic region merging. **IEEE Transactions on Information Technology in Biomedicine**, New York, v. 15, n. 6, p. 929-936, 2011.
- YANG, A. Y. et al. Unsupervised segmentation of natural images via lossy data compression. **Computer Vision and Image Understanding**, New York, v. 110, n. 2, p. 212-225, 2008.
- YU, Z. et al. An adaptive unsupervised approach toward pixel clustering and color image segmentation. **Pattern Recognition**, New York, v. 43, n. 5, p. 1889-1906, 2010.
- ZITNICK, C. L.; KANG, S. B. Stereo for image-based rendering using image over-segmentation. **International Journal of Computer Vision**, Hingham, v. 75, n. 1, p. 49-65, 2007.