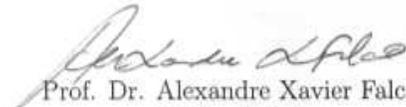


**Aglomeración de pixels pela transformada imagem floresta e sua
aplicación em segmentación de fundo de imagens naturais**

Maíra Saboia da Silva

Este exemplar corresponde à redação final da
Dissertação devidamente corrigida e defendida
por Maíra Saboia da Silva e aprovada pela
Banca Examinadora.

Campinas, 25 de julho de 2011.


Prof. Dr. Alexandre Xavier Falcão
(Orientador)

Dissertação apresentada ao Instituto de Com-
putação, UNICAMP, como requisito parcial para
a obtenção do título de Mestre em Ciência da
Computação.

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MÜLLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Si38a Silva, Maíra Saboia da
Aglomeración de pixels pela transformada imagem
floresta e sua aplicação em segmentação de fundo de
imagens naturais / Maíra Saboia da Silva. - Campinas,
SP: [s.n.], 2011.

Orientador: Alexandre Xavier Falcão.
Dissertação (mestrado) – Universidade Estadual de
Campinas, Instituto de Computação.

1. Segmentação de imagens. 2. Processamento de
imagens . 3. Análise de imagem . 4. Reconhecimento
de padrões. I. Falcão, Alexandre Xavier, 1966-.
II. Universidade Estadual de Campinas. Instituto de
Computação. III. Título.

Informações para Biblioteca Digital

Título em inglês: Clustering of pixels by image foresting transform and its application
in background segmentation of natural images

Palavras-chave em inglês:

Image segmentation

Image processing

Image analysis

Pattern recognition

Área de concentração: Ciência da Computação

Titulação: Mestre em Ciência da Computação

Banca examinadora:

Alexandre Xavier Falcão [Orientador]

Hélio Pedrini

Silvio Jamil Ferzoli Guimarães

Data da defesa: 25-07-2011

Programa de Pós-Graduação: Ciência da Computação

TERMO DE APROVAÇÃO

Dissertação Defendida e Aprovada em 25 de julho de 2011, pela Banca examinadora composta pelos Professores Doutores:



Prof. Dr. Silvio Jamil Ferzoli Guimarães
Instituto de Informática / PUC-MG

Hélio Pedrini

Prof. Dr. Hélio Pedrini
IC / UNICAMP

Alexandre Xavier Falcão

Prof. Dr. Alexandre Xavier Falcão
IC / UNICAMP



Instituto de Computação
Universidade Estadual de Campinas

Aglomerção de pixels pela transformada imagem floresta e sua aplicação em segmentação de fundo de imagens naturais

Maíra Saboia da Silva¹

Julho de 2011

Banca Examinadora:

- Prof. Dr. Alexandre Xavier Falcão (Orientador)
- Prof. Dr. Helio Pedrini
Instituto de Computação, Universidade Estadual de Campinas (IC-UNICAMP)
- Prof. Dr. Silvio Jamil Ferzoli Guimarães
Instituto de Ciências Exatas e Informática, Pontifícia Universidade Católica de Minas Gerais (ICEI- PUC MINAS)
- Prof. Dr. Neucimar Leite (Suplente)
Instituto de Computação, Universidade Estadual de Campinas (IC-UNICAMP)
- Roberto Hirata Junior (Suplente)
Instituto de Matemática e Estatística, Universidade de São Paulo (IME-USP)

¹Bolsista FAPESP, Processo # 01-P-04388/2010, de agosto de 2010 a junho de 2011; Bolsita CNPq, Processo # 136484/2009-3 (IC-Unicamp) de setembro de 2009 a julho de 2010.

Resumo

Esta dissertação apresenta uma metodologia automática para separar objetos de interesse em imagens naturais. Objetos de interesse são definidos como os maiores objetos que se destacam com relação aos pixels em torno deles dentro de uma imagem. Estes objetos não precisam necessariamente estar centrados, mas devem possuir o mínimo possível de pixels na região assumida como fundo da imagem (e.g., borda de imagem com uma dada espessura). A metodologia é baseada em segmentação de fundo e pode ser dividida em duas etapas. Primeiramente, um modelo nebuloso é criado para o fundo da imagem utilizando um método de agrupamento baseado em função densidade de probabilidade das cores de fundo. A partir do modelo é criado um mapa de pertinência, onde os pixels de objeto são mais claros do que os pixels de fundo. Foram investigadas técnicas de agrupamento baseadas em deslocamento médio, transformada imagem floresta, mistura de Gaussianas e maximização da esperança. Três métodos para criação do mapa de pertinência foram propostos e comparados; um inteiramente baseado na transformada imagem floresta, o outro em mistura de Gaussianas e o terceiro em maximização da esperança. Nos dois últimos casos, o agrupamento baseado na transformada imagem floresta foi utilizado como estimativa inicial dos grupos. Em seguida, o mapa de pertinência é utilizado para possibilitar a seleção de pixels sementes de objeto e fundo. Estes pixels geram um agrupamento binário da imagem colorida que separa o fundo do(s) objeto(s). Os experimentos foram realizados com uma base heterogênea composta por 50 imagens naturais. Os melhores resultados foram os obtidos pela metodologia inteiramente baseada na Transformada Imagem Floresta. Para justificar o uso de um agrupamento binário das cores para segmentação, os resultados foram comparados com uma limiarização ótima, aplicada ao mapa de pertinência. Esses testes foram realizados com o algoritmo de Otsu, mas o agrupamento binário apresentou melhores resultados. Também foi proposto um método híbrido de binarização do mapa de pertinência, envolvendo a limiarização de Otsu e a transformada imagem floresta. Neste caso, a limiarização de Otsu reduz o número de parâmetros em relação à primeira.

Abstract

This work presents a new methodology for automatic extraction of desired objects in natural images. Objects of interest are defined as the largest components that differ from their surrounding pixels in a given image. These objects do not need to be centered, but they should contain a minimum number of pixels in the region assumed as background (e.g., an image border of certain thickness). This methodology is based on background segmentation and it can be summarized in two steps. First, a fuzzy model is created by a clustering method based on probability density function of the background colors. This model is a membership map, wherein object pixels are brighter than background pixels. For clustering, the following techniques were investigated: mean-shift, image foresting transform, Gaussian mixture model and expectation maximization. We then propose and compare three approaches to create a membership map; a first method entirely based on the image foresting transform, a second approach based on Gaussian mixture model and a third technique using expectation maximization. The clustering based on image foresting transform was adopted as the initial estimate for the clusters in the case of the two last methods. In a second step, the membership map is used to enable the selection of object and background seed pixels. These pixels create a binary clustering of the color pixels that separates background and object(s). The experiments involved a heterogeneous dataset with 50 natural images. The approach entirely based on the image foresting transform provided the best result. In order to justify the use of a binary clustering of color pixels instead of optimum thresholding on the membership map, we demonstrated that the binary clustering can provide a better result than Otsu's approach. It was also proposed a hybrid approach to binarize the membership map, which combines Otsu's thresholding and image foresting transform. In this case, Otsu's thresholding reduces the number of parameters in regard to the first approach.

Agradecimentos

Ao meu orientador, o Prof. Alexandre Xavier Falcão, pela oportunidade, orientação, dedicação e paciência durante essa jornada.

À minha família, em especial aos meus pais, Paulo Roberto e Mariangela Saboia por todo o suporte e amor, e às minhas irmãs, Naara e Tainá, pela amizade além da fraternidade. Amo todos vocês!

Ao Lucas Bueno, que me fez refletir sobre causas, motivos, razões e circunstâncias, e que esteve ao meu lado na conquista os últimos passos desta caminhada. Aos amigos Lourdes, Samuel e Cristina, Luísa e Pedro que me acolheram e me ajudaram sem restrições.

Ao César Castelo, que sempre esteve ao meu lado, incondicionalmente, e que me deu suporte do começo ao fim desse trabalho.

Ao Laboratório de Informática Visual-LIV. Aos colegas do LIV. Aos colegas das disciplinas. Aos colegas das repúblicas. A todos que colaboraram, direta ou indiretamente, com o meu trabalho.

A todos amigos que fiz pelo meu Brasil!!

À Universidade Estadual de Campinas (UNICAMP) e ao Instituto de Computação (IC) por permitirem e tornarem possível a realização deste trabalho. Aos professores e funcionários desta instituição.

Ao CNPq e à FAPESP, pelo apoio financeiro.

A DEUS!

Sumário

| | |
|---|-----------|
| Resumo | vii |
| Abstract | ix |
| Agradecimentos | xi |
| 1 Introdução | 1 |
| 1.1 Motivação | 3 |
| 1.2 Desafios do trabalho | 3 |
| 1.3 Objetivos | 3 |
| 1.3.1 Objetivos específicos | 4 |
| 1.4 Contribuições do trabalho | 4 |
| 1.5 Organização da dissertação | 5 |
| 2 Agrupamento de Pixels | 7 |
| 2.1 Categorias de métodos de agrupamento | 8 |
| 2.2 Definições gerais | 9 |
| 2.3 Agrupamento de pixels por deslocamento médio | 10 |
| 2.4 Agrupamento de dados por maximização da esperança | 11 |
| 2.5 Transformada imagem floresta | 13 |
| 2.5.1 Agrupamento de múltiplos rótulos | 16 |
| 2.5.2 Agrupamento binário | 20 |
| 2.6 Análise comparativa entre as técnicas baseadas em função densidade de probabilidade | 22 |
| 3 Segmentação de fundo em imagens naturais usando Transformada Imagem-Floresta | 25 |
| 3.1 Modelo de fundo | 27 |
| 3.2 Estimativa de mapa de pertinência com o objeto | 28 |
| 3.2.1 Mapa de pertinência direto da função de conectividade ótima | 28 |
| 3.2.2 Mapa de pertinência com Mistura de Gaussianas | 30 |
| 3.2.3 Mapa de pertinência com maximização de esperança | 31 |
| 3.3 Classificação binária | 34 |
| 3.4 Pós-processamento | 34 |

| | | |
|----------|--|-----------|
| 3.5 | Análise comparativa das abordagens | 35 |
| 4 | Experimentos e Resultados | 37 |
| 4.1 | Base de imagens | 37 |
| 4.1.1 | Medida de acurácia | 38 |
| 4.2 | Metodologia experimental | 39 |
| 4.2.1 | Definição dos parâmetros | 39 |
| 4.3 | Resultados | 40 |
| 4.4 | Binarização do Mapa de pertinência | 56 |
| 4.5 | Análise dos resultados | 57 |
| 5 | Considerações Finais | 63 |
| | Bibliografia | 67 |

Lista de Tabelas

| | | |
|-----|---|----|
| 4.1 | Acurácia média por tamanho do conjunto de amostras da borda | 40 |
| 4.2 | Melhor resultado por algoritmo | 41 |
| 4.3 | Resultado médio por algoritmo | 56 |
| 4.4 | Resultado do teste com Otsu | 57 |

Lista de Figuras

| | | |
|-----|--|----|
| 1.1 | a) Agrupamento de pixels usando transformada imagem floresta, usando restrições de conexidade no domínio da imagem. b) O usuário indica quais regiões formam o objeto de interesse. c) Objeto segmentado. Imagens obtidas de [3] | 2 |
| 1.2 | Na imagem a) a região da borda da imagem destacada é utilizada para selecionar amostras do fundo. Imagem b) exhibe o mapa de pertinência com o objeto e a imagem c) exhibe a imagem segmentada | 3 |
| 1.3 | Fluxograma geral da metodologia proposta | 4 |
| 2.1 | Exemplos das diferentes formas de agrupar um mesmo conjunto de dados com base em diferentes critérios | 7 |
| 2.2 | (a) Espaço de atributos com diferentes concentrações de amostras por grupo. Pode-se identificar diferentes números de grupos dependendo da escala. Soluções razoáveis seriam (b) quatro e (c) cinco grupos, onde dois deles estariam se tocando. Imagem de [3] | 18 |
| 3.1 | Imagens geradas pela metodologia proposta. a) região da imagem definida como borda, b) amostras selecionadas da borda, c) exemplo da imagem de pertinência e d) resultado da classificação binária. | 26 |
| 3.2 | Fluxograma da metodologia proposta para segmentação de fundo de imagens coloridas | 27 |
| 3.3 | Agrupamento de múltiplos rótulos em imagem natural | 28 |
| 3.4 | Exemplos de Mapa de pertinência usando conexidade ótima. | 33 |
| 3.5 | Resultado da metodologia proposta usando como pós-processamento o número de objetos b) $n = 1$ e c) $n = 2$ para a) imagem original | 35 |
| 3.6 | Fechamento morfológico. a) imagem original, b) o resultado da segmentação sem pós-processamento. c) imagem após fechamento morfológico com limiar 300, e d) fechamento morfológico com limiar suficiente para fechar todos os buracos internos ao objeto | 35 |
| 4.1 | Exemplos de imagens que compõem a base de dados e seus respectivos <i>Padrões Ouro</i> | 37 |
| 4.2 | Processo de criação dos <i>Padrões Ouro</i> . a) marcadores desenhados na imagem, b) resultado da segmentação e c) o padrão ouro. | 38 |

| | | |
|-----|---|----|
| 4.3 | Modificação no fluxograma para abordagem em que o mapa de pertinência é binarizado pelo algoritmo de Otsu para gerar o resultado final. | 57 |
| 4.4 | Modificação no fluxograma para abordagem híbrida entre o algoritmo de Otsu e o agrupamento binário por IFT | 58 |
| 4.5 | Exemplo de melhor resultado obtido pela IFT. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d) | 59 |
| 4.6 | Exemplo do caso onde o EM diverge do resultado ótimo. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d) | 60 |
| 4.7 | Exemplo de resultados semelhantes para IFT e EM. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d) | 60 |
| 4.8 | Exemplo de resultados para EM e GMM melhores que IFT. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d) | 61 |
| 5.1 | Melhor configuração da metodologia proposta | 65 |

Capítulo 1

Introdução

Segmentação de imagem é um dos principais problemas em Processamento de Imagens e Visão Computacional, necessária para a análise quantitativa de objetos presentes nas imagens. Esta tarefa requer a localização aproximada dos objetos de interesse e a definição precisa de suas extensões espaciais na imagem. Técnicas de aglomeração de pixels permitem particionar a imagem em regiões, tal que os pixels de uma dada região tenham maior “similaridade” entre si do que com os pixels de regiões distintas. Alguns algoritmos utilizados para agrupamento de pixels em imagens são k-médias [25], Mistura de Gaussianas, FCM (do inglês *Fuzzy C-Means*) [12], Transformada Imagem Floresta [31], Maximização da Esperança [10], Corte em Grafos [16], deslocamento médio (do inglês *Mean-shift*) [8].

Estas técnicas reduzem a segmentação ao problema de reconhecer quais regiões devem ser unidas para formar cada objeto de interesse. Este reconhecimento pode ser feito pelo usuário, ao desenhar marcadores que unem regiões de um dado objeto (Figura 1.1), ou por modelos desses objetos [5, 27].

Quando um modelo de conhecimento de objeto é utilizado para segmentação, a técnica é denominada de segmentação de objeto. Outra forma de interpretar o problema de segmentação é a subtração de fundo [22, 30, 18], cujos objetivos são modelar o fundo da imagem e separar dele todas as regiões que não satisfazem ao modelo. O reconhecimento de quais regiões formam cada objeto é ainda deixado para o usuário ou modelo de objeto. No entanto, em diversas aplicações, a segmentação de fundo pode ser suficiente. Exemplos são o ajuste de foco de uma câmera digital em objetos da cena, a detecção de objetos em movimento e o acompanhamento desses objetos em um vídeo de vigilância [32], busca de imagens baseada em conteúdo [24].

O objetivo desse trabalho foi explorar a técnica de agrupamento baseada na Transformada Imagem Floresta (IFT do inglês *Image Foresting Transform*) no contexto de segmentação de fundo de imagens naturais. Para este propósito, foram combinadas duas técnicas de agrupamentos baseadas em IFT: agrupamento de múltiplos rótulo [31], para criação do modelo de conhecimento do fundo, e o agrupamento binário final, para separar objeto e fundo [29].

As técnicas de agrupamento de pixels por floresta de caminhos ótimos modelam um conjunto de pixels (i.e., amostras da imagem) como um grafo, onde os pixels são os nós,

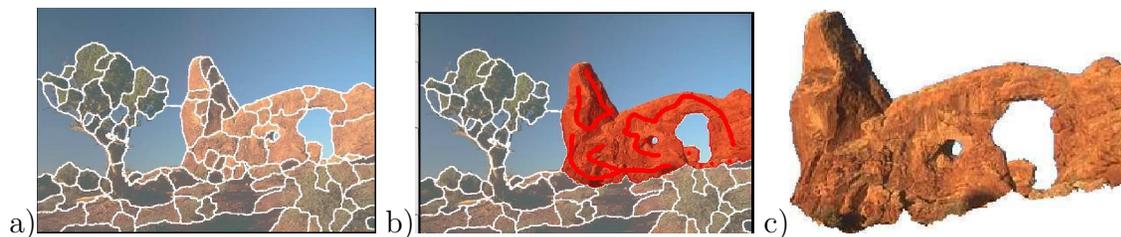


Figura 1.1: a) Agrupamento de pixels usando transformada imagem floresta, usando restrições de conexidade no domínio da imagem. b) O usuário indica quais regiões formam o objeto de interesse. c) Objeto segmentado. Imagens obtidas de [3]

descritos por um vetor de atributos, e os arcos ligam pixels adjacentes no espaço de atributos. A relação de adjacência é um parâmetro da metodologia, o qual pode levar em conta também restrições de conexidade no domínio da imagem (i.e., distâncias máximas entre pixels na imagem). No exemplo da Figura 1.1 a), esta relação de adjacência conecta pixels que estão a certas distâncias no espaço de atributos e na imagem. Uma densidade de probabilidade para cada nó no grafo é estimada com base nas distâncias entre os vetores de atributos de nós adjacentes, de modo que os máximos da Função Densidade de Probabilidade (fdp) das amostras da imagem são representados por alguns nós do grafo. Para uma dada função de conexidade aplicada a caminhos simples neste grafo, os máximos da fdp particionam o grafo em uma floresta de caminhos ótimos, cujas raízes são os nós de máximo e a árvore de cada raiz é um grupo (*cluster*) formado pelos nós mais fortemente conexos com esta raiz do que com qualquer outro máximo da fdp.

Técnicas como Deslocamento Médio (*Mean-Shift*) [15, 6, 8] e Maximização da Esperança (EM do inglês *expectation-maximization*) [10, 40, 1] adotam a mesma estratégia de agrupar as amostras de cada máximo da fdp para particionar um conjunto de amostras dado. No entanto, as abordagens de segmentação baseadas em EM [20, 19] modelam a fdp normalmente como uma mistura de Gaussianas (GMM do inglês *Gaussian mixture model*) forçando o número de máximos, e usam o algoritmo de EM para encontrar os parâmetros ótimos das Gaussianas. Ou seja, estas técnicas assumem um certo número de grupos (um por Gaussiana). As abordagens baseadas em *mean-shift* [23, 38, 17] não assumem modelos para a fdp, nem calculam a fdp explicitamente, elas apenas buscam um ponto de máximo mais próximo de cada amostra, seguindo o gradiente da fdp como estimado localmente. O algoritmo de agrupamento por IFT [31] também não assume modelos para a fdp, mas estima seu valor para cada amostra e, em vez de procurar os máximos da fdp, ele os detecta primeiro e depois propaga seus rótulos para as respectivas amostras mais fortemente conexas.

Neste trabalho propomos uma nova técnica automática de segmentação de fundo de imagens naturais. Esta técnica usa um agrupamento de multi-rótulos pela transformada imagem floresta em um conjunto de pixels assumido como parte do fundo, para várias topologias de grafos, a fim de encontrar a melhor estimativa de fdp. Esse conjunto de pixels do fundo foi selecionado da borda da imagem, pois foi considerada como uma região que contém uma quantidade representativa de amostras do fundo (Figura 1.2a).

Depois utiliza a fdp deste conjunto para calcular um mapa de pertinência com o objeto, de modo que o resultado é uma imagem com pixels claros para objeto e escuros para fundo (Figura 1.2b). A binarização da imagem assume um percentual de pixels claros e escuros como partes do fundo e do(s) objeto(s), treina e aplica um agrupamento binário baseado na IFT [29] para separar objeto(s) e fundo (Figura 1.2c). Na Figura 1.3 é exibido o fluxograma geral da metodologia proposta, esse fluxo possui alguns variantes, que estão descritos no Capítulo 3.



Figura 1.2: Na imagem a) a região da borda da imagem destacada é utilizada para selecionar amostras do fundo. Imagem b) exibe o mapa de pertinência com o objeto e a imagem c) exibe a imagem segmentada

1.1 Motivação

A principal motivação é o grande número de aplicações envolvendo agrupamento de pixels em imagens naturais e os bons resultados recentes da abordagem de agrupamento pela transformada imagem floresta em algumas dessas aplicações, como na segmentação de tecidos do cérebro humano em imagens de ressonância magnética [4] e na interpretação inteligente da marcação feita pelo usuário durante uma tarefa de segmentação interativa de imagens naturais [37].

1.2 Desafios do trabalho

Várias técnicas do estado da arte utilizam grafos para modelar uma imagem, porém a dificuldade do computador para separar objeto e fundo faz com que essas técnicas necessitem da intervenção do usuário. Desta forma, o desafio principal deste trabalho é obter automaticamente um modelo de fundo para poder destacar as partes distintas do modelo como objeto de interesse. Neste contexto, o estado da arte usa técnicas baseadas em deslocamento médio e maximização de esperança. Estamos propondo uma técnica nova que apresenta várias vantagens em relação ao estado da arte.

1.3 Objetivos

Este trabalho teve como objetivo geral o estudo e o desenvolvimento de métodos de segmentação de fundo de imagens baseados em fdp. O agrupamento de pixels pela trans-

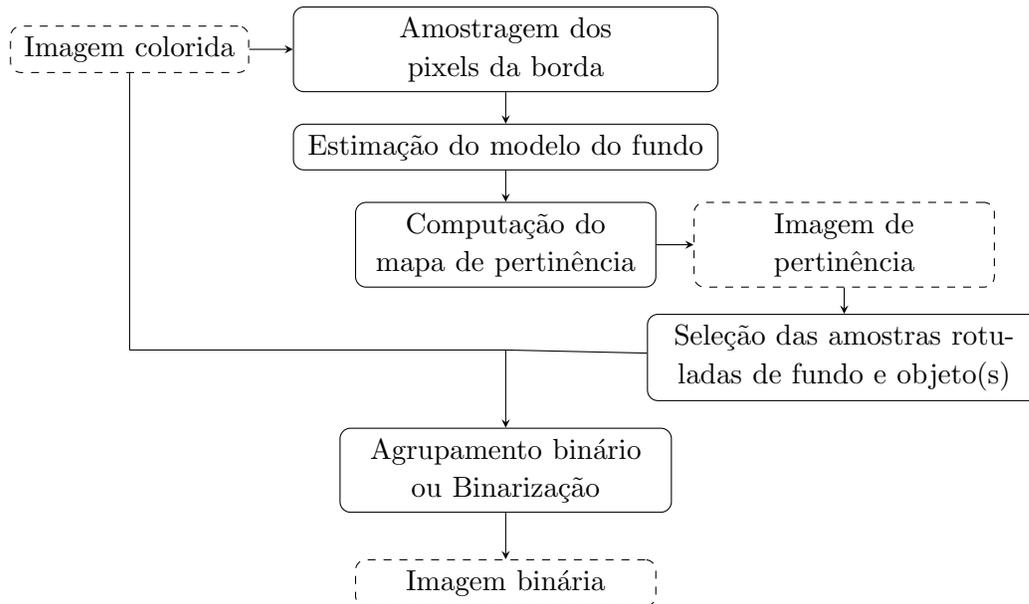


Figura 1.3: Fluxograma geral da metodologia proposta

formada imagem floresta foi a metodologia proposta.

1.3.1 Objetivos específicos

Os objetivos específicos desse trabalho foram:

- Estudo sobre técnicas de agrupamento de pixels baseadas em função densidade de probabilidade (fdp);
- Análise comparativa entre estas técnicas.
- Propor uma técnica de segmentação baseada em agrupamento de pixels por IFT.
- Propor um método híbrido entre agrupamento por IFT e maximização da esperança.
- propor um método híbrido entre agrupamento por IFT e mistura de Gaussianas.
- Propor um método híbrido entre o algoritmo de Otsu e o agrupamento binário por IFT para binarização de imagens em tons de cinza.

1.4 Contribuições do trabalho

Entre as contribuições deste trabalho, podem ser citadas:

- Uso do agrupamento de pixels pela Transformada Imagem Floresta para segmentação de fundo em imagens naturais. O agrupamento foi feito com base em duas topologias de grafos. O agrupamento de múltiplos rótulos usa os k-vizinhos mais próximos como relação de adjacência e o agrupamento binário usa um grafo completo. Em ambos os casos, os nós são amostras de pixels, no primeiro caso são pixels da borda e no segundo são pixels mais claros e mais escuros do mapa de pertinência.
- Desenvolvimento das técnicas híbridas para segmentação binária de imagens coloridas. Os métodos de agrupamentos híbridos foram 1) “transformada imagem floresta e mistura de Gaussianas” (IFT + GMM) e 2) “transformada imagem floresta e maximização da esperança” (IFT + EM), em que a IFT é utilizada para criar um modelo de conhecimento do fundo da imagem. Além do método híbrido entre Otsu e agrupamento binário por IFT para binarização de imagens em tons de cinza.

1.5 Organização da dissertação

Este texto está organizado de forma que primeiramente é apresentado um levantamento bibliográfico dos métodos relacionados e depois os capítulos relacionados com o método proposto.

Como parte da revisão bibliográfica estão os Capítulos 2 e 3. No Capítulo 2 estão descritas as técnicas de agrupamento de dados baseados em funções de densidade de probabilidade, dois métodos de agrupamento bastante utilizados na literatura, o deslocamento médio e maximização da esperança, também estão descritos os agrupamentos baseados na IFT.

Nos capítulos 3 e 4 estão descritos a metodologia proposta para segmentação de fundo baseada na IFT e os experimentos realizados, respectivamente. No capítulo da metodologia proposta é apresentada a ideia geral e as etapas nas quais o método está dividido, e cinco implementações diferentes da metodologia.

Para finalizar, o Capítulo 5 apresenta as considerações finais e os possíveis trabalhos futuros.

Capítulo 2

Agrupamento de Pixels

Uma maneira de visualizar segmentação de imagens é considerá-la como uma forma de particionar a imagem em grupos naturais de pixels utilizando características inerentes aos mesmos [11, 14]. A classificação não supervisionada é o tipo de classificação em que o rótulo dos dados de treinamento não estão disponíveis. Portanto, o desafio para esse tipo de classificação é estimar um padrão de organização dentro dos dados e criar um agrupamento. Esse agrupamento permite a utilização de similaridades e diferenças entre os pixels para possibilitar a formulação de conclusões sobre eles [39].

O processo de atribuir dados aos grupos pode levar a diferentes resultados dependendo dos critérios utilizados para fazer o agrupamento. Veja a Figura 2.1, se o critério de agrupamento fosse baseado em cor, o agrupamento seria o exibido na imagem a), se o critério fosse proximidade espacial, os grupos seriam os exibidos na imagem b), se a forma geométrica fosse o critério, o resultado seria o da imagem c). No entanto, se pelo menos dois dos critérios tivessem sido utilizados ao mesmo tempo, o resultado seria o da imagem d).

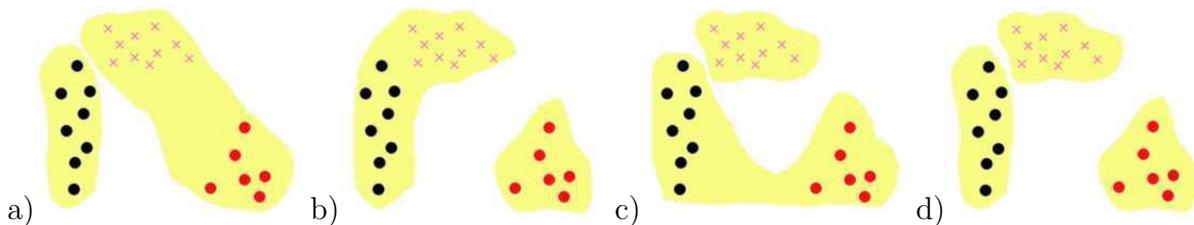


Figura 2.1: Exemplos das diferentes formas de agrupar um mesmo conjunto de dados com base em diferentes critérios

Neste capítulo é apresentada uma breve categorização dos métodos de agrupamento de dados. A seguir são descritos três dos algoritmos que estão relacionados à metodologia proposta, e no final do capítulo é feita uma resenha comparativa entre esses algoritmos.

2.1 Categorias de métodos de agrupamento

Os algoritmos de agrupamento podem ser vistos como sistemas que nos fornecem grupos naturais dos dados. O resultado depende do algoritmo e dos critérios utilizados. Assim, um algoritmo de agrupamento é um processo de aprendizado que tenta descrever como os dados se agrupam. Em termos gerais, os algoritmos de agrupamentos podem ser divididos em categorias, em que cada algoritmo pode pertencer a mais de uma categoria simultaneamente [39]:

- *Algoritmos sequenciais.* Esses algoritmos produzem um agrupamento único. Em geral, os dados são apresentados uma única vez (ou poucas, tipicamente não mais que cinco vezes) e o agrupamento vai sendo formado enquanto eles são apresentados. Um dado novo pode fazer parte de um grupo que já existe, ou pode ser criado um novo grupo para contê-lo.
- *Algoritmos hierárquicos.* Os agrupamentos hierárquicos são divididos em dois tipos, os divisivos ou aglomerativos.
 - aglomerativos: esses algoritmos produzem uma sequência de agrupamentos pela diminuição do número de grupos a cada iteração. O agrupamento produzido em cada iteração é resultado do agrupamento anterior pela união de alguns grupos em um único grupo;
 - divisivos: esses algoritmos agem na direção contrária aos aglomerativos; eles produzem uma sequência de agrupamentos aumentando a cada iteração. O agrupamento gerado em cada iteração é resultado do anterior pela divisão de um grupo em dois ou mais grupos;
- *Algoritmos baseados em otimização de função de custo.* Essa categoria contém os algoritmos cujos agrupamentos são definidos por uma função de custo. A maioria desses algoritmos utiliza conceitos de cálculo diferencial e produz sucessivos agrupamentos tentando otimizar a função de custo. Eles finalizam o procedimento quando um ótimo local da função é determinado. Esta categoria pode ser dividida nos seguintes tipos:
 - partição rígida: a amostra pertence exclusivamente a um grupo. A atribuição das amostras a um único grupo é realizada otimamente, de acordo com o critério de otimização adequado;
 - probabilístico: é um tipo especial de agrupamento rígido que utiliza a classificação Bayesiana. Cada amostra é atribuída ao grupo cuja probabilidade a posteriori dela pertencer ao grupo é máxima;
 - nebuloso: em que uma amostra pertence a um grupo específico até um certo grau;
 - detecção de fronteiras: em vez de determinar os grupos pelas próprias amostras, os algoritmos ajustam iterativamente as fronteiras das regiões que separam os grupos;

- *Outros*: esta categoria contém algumas técnicas especiais de agrupamento que não se encaixam adequadamente em nenhuma das categorias anteriores. Elas são:
 - ramificação e poda: esse tipo de algoritmo fornece um agrupamento ótimo sem utilizar todos os possíveis agrupamentos, para um número fixo de grupos, e para critérios pré-definidos que satisfaçam certas condições.
 - genéticos: esses algoritmos utilizam uma população inicial de possíveis grupos e iterativamente geram novas populações, que em geral, contém melhores grupos que os da população anterior, de acordo com critérios específicos.
 - busca de vales: esses algoritmos tratam as amostras como instâncias de uma (multidimensional) variável aleatória \mathbf{x} . Eles são fundamentados no pressuposto de que regiões de \mathbf{x} em quais muitas amostras estão localizadas correspondem a uma região de crescimento de uma função de densidade de probabilidade. Portanto, a estimativa da fdp deve destacar regiões onde os grupos estão formados; os grupos são vistos como os picos dessa função de densidade e, inspirados nessa consideração, os algoritmos buscam localizar as fronteiras dos grupos nos vales.
 - baseados em densidade: esses algoritmos consideram os grupos como regiões densas nos dados. A maioria dos algoritmos nessa categoria não impõe restrição de forma dos grupos resultantes. Assim, esses algoritmos são capazes de lidar com grupos de formas arbitrárias.

Neste trabalho estamos interessados em algoritmos que são baseados em funções de densidade de probabilidade. Dois algoritmos clássicos para agrupamento e que se enquadram nessa categoria são o algoritmo de maximização da esperança e o deslocamento médio. Esses algoritmos estão descritos e analisados neste capítulo. Ademais, um algoritmo derivado de um *framework* mais geral, a saber, a Transformada Imagem Floresta, também baseado em função de densidade de probabilidade é descrito e analisado.

2.2 Definições gerais

Seja $\hat{I} = (D_I, \vec{I})$ uma imagem colorida, cujos pixels s de seu domínio D_I têm associado um vetor $\vec{I}(s)$ de três componentes de cor. Qualquer transformação para extração de novos atributos, gera uma imagem com o mesmo domínio mas vetor de atributos distinto. Portanto, sem perda de generalidade, podemos assumir que o vetor $\vec{I}(s) = (I_1(s), I_2(s), \dots, I_n(s))$ é um vetor de variáveis aleatórias $\mathbf{x} = (x_1, x_2, \dots, x_n)$, cujos valores vão depender de $s \in D_I$. Neste sentido, podemos usar $\rho(\mathbf{x})$ para indicar o valor de densidade de probabilidade no ponto \mathbf{x} do espaço de atributos \mathbb{R}^n ou $\rho(s) = \rho(\mathbf{x})$ para indicar a densidade de probabilidade de um pixel $s \in D_I$ que é mapeado no ponto \mathbf{x} .

Se soubéssemos o número c de regiões (grupos) de interesse na segmentação de uma imagem, a probabilidade *a priori* $P(j)$ de ocorrência de observações de valores \mathbf{x} de cada grupo $j = 1, 2, \dots, c$, e a densidade de probabilidade $\rho(\mathbf{x} \setminus j)$ condicionada ao conhecimento

que \mathbf{x} vem do grupo j , o teorema de Bayes nos permitiria calcular a probabilidade *a posteriori* de ocorrência do grupo j dada a observação \mathbf{x} :

$$P(j|\mathbf{x}) = \frac{\rho(\mathbf{x}|j)P(j)}{\rho(\mathbf{x})}, \quad (2.1)$$

onde $\rho(\mathbf{x}) = \sum_{j=1}^c \rho(\mathbf{x}|j)P(j)$. A decisão que \mathbf{x} , ou melhor, o pixel $s \in D_I$ que gera a observação \mathbf{x} , vem da classe j^* seria tomada pela regra:

$$P(j^*|\mathbf{x}) = \max_{j=1,2,\dots,c} \{P(j|\mathbf{x})\}. \quad (2.2)$$

Como a densidade $\rho(\mathbf{x}|j)$ não é conhecida, os métodos de agrupamento tentam estimar esse valor. Existem métodos paramétricos, que supõem que cada classe tem uma forma. Existem também os métodos não paramétricos que tentam estimar a função de densidade baseando-se nas informações presentes nas amostras.

Os algoritmos de agrupamento não-paramétricos, tais como deslocamento médio e o agrupamento pela transformada imagem floresta, buscam particionar a imagem em regiões (grupos) tais que cada região é formada por um domo da função densidade de probabilidade $\rho(\mathbf{x})$. Uma estratégia adotada neste trabalho é amostrar os pixels em D_I , formando um conjunto menor $\mathcal{N} \subset D_I$ de amostras. O agrupamento é aplicado ao conjunto \mathcal{N} e depois os rótulos (ou valores de densidade) são propagados para os demais pixels em $D_I \setminus \mathcal{N}$.

Técnicas paramétricas, como maximização da esperança, assumem um modelo matemático para os grupos. Por exemplo, assumem que $\rho(\mathbf{x}|j)$ é uma função Gaussiana com parâmetros, vetor média e matriz de covariância, e assumem que $\rho(\mathbf{x}) = \sum_{j=1}^c \rho(\mathbf{x}|j)P(j)$ é uma mistura de Gaussianas. A função de $P(j|\mathbf{x})$ é denominada verossimilhança entre \mathbf{x} e j . A decisão é feita pelo grupo j com máxima verossimilhança. Portanto, o algoritmo de maximização da esperança busca os parâmetros que melhor aproximam a mistura de Gaussianas dos dados \mathbf{x} observados.

2.3 Agrupamento de pixels por deslocamento médio

O Deslocamento Médio (*Mean-shift*) é uma técnica de agrupamento não-paramétrica baseada no gradiente da função de densidade de probabilidade das amostras no espaço de atributos, e não requer conhecimento *a priori* do número de grupos e nem da forma da distribuição. O deslocamento médio foi proposto inicialmente por Fukunaga e Hostetler em 1975 [15], adaptado por Cheng [6] para análise de imagens e posteriormente estendido por Comaniciu, Meer e Ramesh para problemas de visão computacional, incluindo segmentação, suavização adaptativa [8] e *tracking* [9].

A ideia do algoritmo é considerar uma função densidade de probabilidade (fdp) das amostras no espaço de atributos, onde as regiões densas correspondem aos domos da distribuição. Para cada uma das amostras, o algoritmo iterativamente busca o ponto de máxima densidade local, que é identificado por um processo de deslocamento crescente na

direção de um máximo na fdp. No final, as amostras cujos gradientes estão direcionados para o mesmo máximo local são consideradas membros de um mesmo grupo, e recebem o rótulo do máximo.

Uma estimativa de densidade multivariável computada no ponto \mathbf{x} pode ser obtida com uma função de Kernel. O uso de um kernel diferenciável permite definir a estimativa do gradiente da densidade como o gradiente da estimativa da densidade do kernel [2, 8]:

$$\widehat{\nabla}\rho(\mathbf{x}) \equiv \nabla\hat{\rho}(\mathbf{x}) = \frac{1}{kd^n} \sum_{i=1}^k \nabla K \left(\frac{\mathbf{x} - \mathbf{x}_i}{d} \right) \quad (2.3)$$

Os máximos da função de densidade estão localizados nos zeros da função de gradiente $\nabla\rho(\mathbf{x}) = \mathbf{0}$.

Se a Equação 2.3 for fatorada, um dos termos dessa fatoração é o vetor

$$m(\mathbf{x}) = \frac{\sum_{i=1}^k \mathbf{x}_i K' \left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{d} \right\|^2 \right)}{\sum_{i=1}^k K' \left(\left\| \frac{\mathbf{x} - \mathbf{x}_i}{d} \right\|^2 \right)} - \mathbf{x} \quad (2.4)$$

onde $K'(\cdot)$ é a função de kernel derivada de $K(\cdot)$, d é a largura da janela ou o tamanho do kernel e $m(\mathbf{x})$ é o vetor de deslocamento que aponta em direção do máximo aumento de densidade e é proporcional à estimativa do gradiente da densidade no ponto \mathbf{x} obtido com kernel K .

O procedimento do deslocamento médio é obtido pelo cálculo sucessivo do vetor $m(\mathbf{x})$ e translação da janela formada por este vetor até a convergência em um ponto estacionário, em que o gradiente da função de densidade é zero. Uma vez que o vetor se desloca no sentido e na direção de maior crescimento da densidade, ele define um caminho até o máximo local; todas as amostras cujos deslocamentos convergem para um mesmo máximo definem uma zona de influência deste. Todos os pontos que estão numa mesma zona de influência pertencem ao mesmo grupo.

Alguns problemas podem ser identificados na técnica do deslocamento médio. Há situações em que o gradiente não é bem estimado porque uma determinada região tem sua densidade constante. Outro possível problema é que uma região de máximo pode ser composta de várias amostras com a mesma densidade (platôs na fdp); a zona de influência poderá ainda ser dividida em múltiplas zonas de influência, uma para cada amostra do máximo. Se o platô estiver numa região que não seja um máximo, o gradiente poderá estacionar nessa região e a amostra \mathbf{x} não apontará para máximo algum. A vantagem é que armazena apenas amostras e não requer armazenar as adjacentes de cada amostra.

2.4 Agrupamento de dados por maximização da esperança

O método de Maximização da Esperança (EM do inglês *Expectation Maximization*) [10] é um algoritmo genérico para estimação de parâmetros aplicado a problemas de dados incompletos. Para aplicação em segmentação de imagens, a ideia fundamental do EM é simples, primeiro é necessário escolher um modelo para a distribuição dos *pixels* contidos na imagem. Existem várias formas de modelar uma distribuição, um modelo muito

utilizado é baseado em misturas de Gaussianas [21], onde a distribuição é considerada como composta por várias Gaussianas. Para esse modelo são escolhidos os parâmetros iniciais, que podem ser artificiais ou encontrados a partir de outras técnicas. Em [41] é demonstrado que a qualidade da partição gerada pelo EM é extremamente dependente da partição inicial escolhida. Os *pixels* são atribuídos a um grupo ao qual eles têm maior probabilidade de pertencer, isto é, que apresente máxima verossimilhança.

Considerando que o conjunto de amostras foi particionado em c subconjuntos \mathcal{N}_j , i.e., $\mathcal{N} = \cup_{j=1}^c \{\mathcal{N}_j\}$. Para cada $j = 1, 2, \dots, c$, temos que calcular o vetor média $\mu_j^{(0)}$, a matriz de covariância $\Sigma_j^{(0)}$, e probabilidade *a priori* $P_j^{(0)}$ da primeira iteração $i = 0$, dados por:

$$\mu_j^{(0)} = \frac{1}{|\mathcal{N}_j|} \sum_{\forall s \in \mathcal{N}_j} \vec{v}(s) \quad (2.5)$$

$$\Sigma_j^{(0)} = \frac{1}{|\mathcal{N}_j|} \sum_{\forall s \in \mathcal{N}_j} [\vec{v}(s) - \mu_j^{(0)}][\vec{v}(s) - \mu_j^{(0)}]^T \quad (2.6)$$

$$P_j^{(0)} = \frac{|\mathcal{N}_j|}{|\mathcal{N}|} \quad (2.7)$$

Para as iterações $i = 1, 2, \dots$ temos as distâncias de mahalanobis e probabilidades *a posteriori* calculadas por:

$$D_M^{(i)}(s, j) = \sqrt{[\vec{v}(s) - \mu_j^{(i-1)}]^T \text{Inv}(\Sigma_j^{(i-1)}) [\vec{v}(s) - \mu_j^{(i-1)}]} \quad (2.8)$$

$$P^{(i)}(j \setminus s) = \frac{\left[\text{Det}(\Sigma_j^{(i-1)}) \right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} D_M^{(i)}(s, j)^2\right] P_j^{(i-1)}}{\sum_{k=1}^c \left[\text{Det}(\Sigma_k^{(i-1)}) \right]^{-\frac{1}{2}} \exp\left[-\frac{1}{2} D_M^{(i)}(s, k)^2\right] P_k^{(i-1)}} \quad (2.9)$$

Em seguida, o vetor média, a matriz de covariância e a probabilidade *a priori* são atualizados para cada grupo $j = 1, 2, \dots, c$, visando a próxima iteração.

$$\mu_j^{(i)} = \sum_{\forall s \in \mathcal{N}} \vec{v}(s) \psi_j(s) \quad (2.10)$$

$$\Sigma_j^{(i)} = \sum_{\forall s \in \mathcal{N}} [\vec{v}(s) - \mu_j^{(i)}][\vec{v}(s) - \mu_j^{(i)}]^T \psi_j(s) \quad (2.11)$$

$$P_j^{(i)} = \frac{1}{|\mathcal{N}|} \sum_{\forall s \in \mathcal{N}} \psi_j(s) \quad (2.12)$$

onde $\psi_j(s)$ é o peso que a amostra s tem no grupo j , e seu valor é definido por:

$$\psi_j(s) = \frac{P^{(i)}(j \setminus s)}{\sum_{j=1}^c P^{(i)}(j \setminus s)} \quad (2.13)$$

A desvantagem do EM, quando ele é utilizado em abordagens não supervisionadas, é a necessidade de ser indicado o número de grupos nos quais os dados estão divididos. Além disso, se os grupos iniciais estiverem distantes dos dados, possivelmente, deverá ser necessário um número maior de iterações. Por isso, uma estimativa inicial dos grupos pode ser deixada para o algoritmo mais eficiente.

Uma potencial vantagem desse método é que a complexidade da classificação de novas amostras é proporcional ao número de grupos. Portanto, se o número de grupos for pequeno, a classificação pode ser muito rápida.

2.5 Transformada imagem floresta

A Transformada Imagem Floresta (IFT do inglês *Image Foresting Transform*) [13] é uma ferramenta para projeto, implementação e avaliação de operadores de processamento de imagem baseados em conexidade. A ideia fundamental da IFT é reduzir um problema de processamento de imagens no cálculo de uma floresta de caminhos ótimos em um grafo derivado da imagem. A imagem é vista inicialmente como um grafo, depois o grafo é transformado em uma floresta a partir de uma função de custo e uma relação de adjacência.

Para alguns casos de adjacência este grafo é denso, por exemplo, o grafo k-vizinhos mais próximos (knn) armazena os k-vizinhos mais próximos de cada pixel no espaço de atributos. Dependendo do tamanho do k, não é possível implementá-lo para todos os pixels da imagem. Entretanto, a IFT se aplica eficientemente se o grafo for esparso. Se o grafo for denso, existem duas abordagens para lidar com o problema: ou se impõe uma restrição espacial, onde apenas os knn dentro de um dado raio para um dado pixel são considerados no cálculo; ou é feita uma amostragem dos pixels da imagem. O grafo dos k-vizinhos mais próximos é transformado na floresta e depois os rótulos de cada árvore são propagados para o restante da imagem.

Neste trabalho é apresentado um novo método de classificação binária dos pixels da imagem baseada no *framework* da IFT. O método utiliza um agrupamento de múltiplos rótulos e um agrupamento binário baseados na IFT. O agrupamento de múltiplos rótulos define regiões, onde cada uma dessas regiões equivale a um domo da função de densidade de probabilidade e sua zona de influência. O agrupamento binário utiliza um grafo completo também formado por amostras de pixels da imagem.

Nesta seção estão descritos a Transformada Imagem Floresta e os métodos baseados nela para agrupamento de múltiplos rótulos e para agrupamento binário.

Fundamentação teórica

A IFT define uma floresta de caminhos ótimos, cujos nós são os pixels da imagem e os arcos são definidos por uma relação de adjacência entre os pixels. O custo de um caminho é calculado por uma função dependente da aplicação e das propriedades da imagem. Essas propriedades podem ser globais, provenientes de marcadores ou modelos do objeto, por exemplo; ou propriedades locais, tais como, cor, gradiente, posição do pixel.

Para uma função de custo e uma relação de adjacência adequadas, a IFT associa a cada pixel da imagem um caminho de custo mínimo partindo dos mínimos do mapa de

custo inicial formados por caminhos de um único pixel e particiona a imagem em uma floresta de caminhos ótimos, em que cada árvore tem como raiz um pixel de mínimo e como nós os pixels da imagem mais conexos com a raiz do que com qualquer outra. Como resultado, a IFT gera três atributos por pixel: seu predecessor no caminho ótimo, o custo do caminho e o pixel raiz (ou algum rótulo associado à raiz).

Considere um imagem \hat{I} composta por um par (D_I, \vec{I}) , onde D_I é um conjunto finito de pixels e \vec{I} é um mapeamento que atribui a cada pixel s um valor $I(\vec{s})$ em algum espaço arbitrário de valores. Uma *relação de adjacência* \mathcal{A} é escolhida, e a imagem \hat{I} pode ser interpretada como um grafo direcionado onde os nós são os pixels da imagem e os arcos são os pares em \mathcal{A} . Dizemos que t pertence ao conjunto $\mathcal{A}(s)$ de amostras adjacentes a s ou (s, t) um arco em \mathcal{A} . Alguns exemplos de relações de adjacência são:

- Adjacência Euclidiana:

$$(s, t) \in \mathcal{A} \quad \text{se} \quad \|t - s\| \leq h_i, \quad (2.14)$$

onde $s = (x_s, y_s)$, $t = (x_t, y_t)$, $\|t - s\| = \sqrt{(x_t - x_s)^2 + (y_t - y_s)^2}$ e $h_i > 0$ é o raio de um disco com centro em s .

- Adjacência retangular:

$$(s, t) \in \mathcal{A} \quad \text{se} \quad |x_t - x_s| \leq \frac{a}{2} \text{ e } |y_t - y_s| \leq \frac{b}{2}, \quad (2.15)$$

onde $a > 0$ e $b > 0$ são os lados de um retângulo com centro em s .

- Adjacência baseada em conjunto:

$$(s, t) \in \mathcal{A} \quad \text{se} \quad t - s \in \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}. \quad (2.16)$$

onde $t - s \in \{(-2, 0), (-1, 0), (-1, -1), (0, -2), (1, 0), (2, 0)\}$.

- Adjacência Euclidiana no espaço de atributos com restrição espacial:

$$(s, t) \in \mathcal{A} \quad \text{se} \quad \|t - s\| \leq h_i \text{ e } \|\vec{v}(t) - \vec{v}(s)\| \leq h_f, \quad (2.17)$$

onde a adjacência definida por um é raio $h_f > 0$ no espaço de atributos e um raio h_i no domínio da imagem.

Os casos particulares de vizinhança-4 e vizinhança-8 são obtidos pela relação Euclidiana (Equação 2.14), com os valores de $h_i = 1$ e $h_i = \sqrt{2}$, respectivamente.

Um caminho $\pi_t = \langle t \rangle$ é chamado de caminho trivial e $\pi_t = \pi_s \cdot \langle s, t \rangle$ é a concatenação do caminho π_s e a aresta $(s, t) \in \mathcal{A}$. Para cada caminho π_t , um custo é determinado por uma função f . Para garantir uma floresta de caminhos ótimos, a função deve ser suave. Isto é, para todo pixel t deve existir caminho ótimo π_t que deve ser trivial ou da forma $\pi_t = \pi_s \cdot \langle s, t \rangle$, onde

- (C1) $f(\pi_s) \leq f(\pi_t)$,
- (C2) π_s é ótimo,
- (C3) Para qualquer caminho ótimo π'_s , $f(\pi'_s \cdot \langle s, t \rangle) = f(\pi_t)$.

As funções de conexidade são especificadas por uma regra de inicialização e por uma regra de extensão de caminho. A seguir estão listadas algumas dessas funções.

$$\begin{aligned} f_{max}(\langle t \rangle) &= H(t) \\ f_{max}(\pi_s \cdot \langle s, t \rangle) &= \max\{f_{max}(\pi_s), \omega(s, t)\} \end{aligned} \quad (2.18)$$

$$\begin{aligned} f_{min}(\langle t \rangle) &= H(t) \\ f_{min}(\pi_s \cdot \langle s, t \rangle) &= \min\{f_{min}(\pi_s), \omega(s, t)\} \end{aligned} \quad (2.19)$$

$$\begin{aligned} f_{sum}(\langle t \rangle) &= H(t) \\ f_{sum}(\pi_s \cdot \langle s, t \rangle) &= f_{sum}(\pi_s) + \omega(s, t) \end{aligned} \quad (2.20)$$

$$\begin{aligned} f_{euc}(\langle t \rangle) &= \begin{cases} 0 & \text{se } t \in S \\ +\infty & \text{caso contrário} \end{cases} \\ f_{min}(\langle \pi_s \cdot \langle s, t \rangle \rangle) &= \|t - R(s)\|^2 \end{aligned} \quad (2.21)$$

onde $H(t)$ é o valor inicial e $\omega(s, t)$ é o peso do arco $\langle s, t \rangle$. S o conjunto de possíveis raízes, $R \subset S$ o conjunto das raízes, sendo assim, $R(s)$ a raiz (origem) do caminho que termina em s .

Neste trabalho são utilizadas as funções f_{max} e f_{min} como definidas nas Equações 2.18 e 2.19, respectivamente. Um mapa de conexidade deve ser minimizado com a função f_{max} enquanto a função f_{min} deve maximizar este mapa, conforme as Equações 2.22 e 2.23, respectivamente.

$$V(t) = \min_{\forall \pi_t \in \Pi(D_I, \mathcal{A}, t)} \{f_{max}(\pi_t)\} \quad (2.22)$$

$$V(t) = \max_{\forall \pi_t \in \Pi(D_I, \mathcal{A}, t)} \{f_{min}(\pi_t)\} \quad (2.23)$$

onde $\Pi(D_I, \mathcal{A}, t)$ o conjunto de todos os caminhos do grafo com término em t . Os custos são armazenados em um mapa V de custo ótimo.

Para essas funções, o conjunto S é composto pelos mínimos de $H(t)$ quando é usada a função f_{max} ou composto pelos máximos de $H(t)$ quando é utilizada a função f_{min} . As raízes são o subconjunto dos pixels armazenados em S que não foram conquistados por outros pixels de S , pois a IFT estabelece uma competição entre os elementos de S por pixels mais fortemente conexos a eles em D_I .

Um mapa de predecessores P é uma função que associa a cada pixel $t \in D_I$ ou outro pixel $s \in D_I$, $(s, t) \in \mathcal{A}$, ou uma marca $nil \notin D_I$. No segundo caso, s é dito ser uma raiz do mapa. Uma floresta espalhada é um mapa de predecessores que não contém ciclos - i.e., aquele que leva todo pixel para nil em um número finito de iterações. Para qualquer pixel $t \in D_I$, a floresta P define um caminho $P^*(t)$ recursivamente como $\langle t \rangle$, se $P(t) = nil$, e $P^*(s) \cdot \langle s, t \rangle$ se $P(t) = s \neq nil$.

Para um dada imagem \hat{I} , uma função de custo f e uma relação de adjacência \mathcal{A} a IFT retorna uma *Floresta de Caminhos Ótimos* P - uma floresta espalhada onde $P^*(t)$ tem custo mínimo para todo $t \in D_I$. Como subprodutos, a IFT também gera um mapa de custos V e um mapa de raízes R , onde $V(t)$ é o custo do caminho ótimo até s e $R(t)$ é o pixel inicial deste caminho (ou algum rótulo associado a ele). O algoritmo geral para a IFT, assumindo a minimização de $V(t)$, é apresentado no Algoritmo 1.

Algoritmo 1 – TRANSFORMADA IMAGEM FLORESTA

ENTRADA: Imagem $\hat{I} = (D_I, \vec{I})$, relação de adjacência \mathcal{A} , e uma função de custos f
 SAÍDA: Imagens $\hat{V} = (D_I, \vec{V})$ de custo, $\hat{P} = (D_I, \vec{P})$ de predecessores, e $\hat{R} = (D_I, \vec{R})$ de raízes
 AUXILIARES: Fila Q de prioridades e variável tmp

1. Para todo $t \in D_I$ faça
2. $P(t) \leftarrow nil, V(t) \leftarrow f(\langle t \rangle)$ e $R(t) \leftarrow t$
3. Se $V(t) \neq +\infty$, insira t em Q .
4. Enquanto Q no estiver vazia, faça
5. Remova de Q um pixel s cujo valor de $V(s)$ seja mínimo.
6. Para cada $t \in \mathcal{A}(s)$, tal que $V(t) > V(s)$, faça
7. $tmp \leftarrow f(P^*(s) \cdot \langle s, t \rangle)$
8. Se $tmp < V(t)$ faça
9. $V(t) \neq +\infty$, remova t de Q .
10. $V(t) \leftarrow tmp, P(t) \leftarrow s, R(t) \leftarrow R(s)$, e insira t em Q

O algoritmo inicia com todos os caminhos triviais (linhas 1 e 2). Na linha 3, todos os candidatos a raízes são inseridos na fila Q . As raízes são os mínimos de V , primeiramente detectados na linha 5, ois, ao remover qualquer nó s de Q na linha 5, o caminho $P^*(s)$ é ótimo. As demais, as linhas 7 e 8, calculam e propagam caminhos melhores para os nós adjacentes, seguindo uma ordem não decrescente de custo é aplicando a seguinte regra de expansão: se $f(\pi_s \cdot \langle s, t \rangle) < f(\pi_t)$ (linha 8), então $\pi_t \leftarrow \pi_s \cdot \langle s, t \rangle$

2.5.1 Agrupamento de múltiplos rótulos

Para agrupar pixels de uma imagem com base na similaridade entre eles em um dado espaço de atributos, a relação de adjacência deve levar em conta distâncias entre pixels neste espaço. Devido ao tamanho de D_I em imagens atuais, mesmo considerando a capacidade de memória e processamento dos PCs modernos, fica inviável armazenar explicitamente o grafo resultante de algumas relações de adjacência, ou mesmo processar um grafo denso tendo todos os pixels em D_I como nós deste grafo.

Para contornar o problema mencionado, o conjunto $\mathcal{N} \subset D_I$ de nós do grafo é frequentemente escolhido por amostragem aleatória com distribuição de probabilidade uniforme para os pixels escolhidos. A ideia é, portanto, calcular a floresta de caminhos ótimos em um grafo (N, \mathcal{A}) , onde $\mathcal{A} \subset \mathcal{N} \times \mathcal{N}$, e depois usar atributos da floresta para calcular o rótulo dos pixels de $D_I \setminus \mathcal{N}$.

Para ocupar amostras em \mathcal{N} , a relação de adjacência \mathcal{A} adotada neste trabalho define como adjacentes a s as $k > 0$ amostras t mais próximas de s de acordo com a distância $d(s, t) = \|\vec{I}(t) - \vec{I}(s)\|$.

O valor da fdp $\rho(s)$ (peso do nó) pode ser computado pela equação abaixo

$$\rho(s) = \frac{1}{\sqrt{2\pi\sigma^2}|\mathcal{A}(s)|} \sum_{\forall t \in \mathcal{A}(s)} \exp\left(\frac{-d^2(s, t)}{2\sigma^2}\right) \quad (2.24)$$

onde σ pode ser fixado por

$$\sigma = \max_{\forall (s, t) \in \mathcal{A}} \left\{ \frac{d(s, t)}{3} \right\} \quad (2.25)$$

para garantir que todas as amostras adjacentes sejam consideradas na estimativa da densidade. Note que σ definido como o peso máximo de arco em \mathcal{A} . σ é calculado para garantir que todos os nós em \mathcal{N} serão considerados. O denominador 3 é escolhido porque a Gaussiana cai 99% em 3σ . A Equação 2.26 define um grafo knn $(\mathcal{N}, \mathcal{A})$ e, embora o kernel seja gaussiano, apenas as k -vizinhas amostras mais próximas de s são utilizadas para computar o valor da fdp.

Relações de adjacências simétricas resultam em relações de conexidade simétricas, porém \mathcal{A} é uma relação de adjacência assimétrica. Sendo assim, um máximo da fdp deve ser um subconjunto de amostras adjacentes com o mesmo valor de densidade; necessário para garantir conexidade entre cada par de amostras nesse máximo.

Assim, qualquer amostra do máximo pode ser um representante e conquistar, por um caminho ótimo, outras amostras neste máximo ou em sua zona de influência. Para isso é necessário estender a relação de adjacência \mathcal{A} para ser simétrica nos platôs de ρ a fim de computar os agrupamentos sem quebrar os máximos em múltiplas zonas de influência (árvores).

$$\begin{aligned} \text{Se } t &\in \mathcal{A}(s), \\ s &\notin \mathcal{A}(t) \text{ e} \\ \rho(s) &= \rho(t), \text{ então} \\ \mathcal{A}'(t) &\leftarrow \mathcal{A}(t) \cup \{s\}. \end{aligned} \quad (2.26)$$

Ou seja, se a densidade nas amostras s e t tem o mesmo valor e a amostra t é adjacente à amostra s , mas a amostra s não é adjacente à amostra t , t passa a fazer parte da adjacência de s .

As escolhas da relação de adjacência, fdp, e vetor de atributos permitem detectar agrupamentos com formas arbitrárias, porém eles apresentam problemas com as diferenças na escala e na concentração das amostras. Escolhas adaptativas do σ dependendo da região do espaço de atributos [7] podem ser utilizadas para solucionar esse problema. Ao levar em conta os k -vizinhos mais próximos, é feita uma redução do problema de lidar com diferentes concentrações e reduções na escala ao problema de encontrar o melhor valor de k no intervalo $[1, k_{\max}]$.

A solução é considerar um corte mínimo em $(\mathcal{N}, \mathcal{A})$ conforme os resultados dos agrupamentos para $k \in [k_{\min}, k_{\max}]$, de acordo com a medida de corte $C(k)$ [35].

$$C(k) = \sum_{j=1}^c \frac{W'_j}{W_j + W'_j}, \quad (2.27)$$

$$W_j = \sum_{\forall (s,t) \in \mathcal{A} | L(s)=L(t)=w_j} \frac{1}{d(s,t)}, \quad (2.28)$$

$$W'_j = \sum_{\forall (s,t) \in \mathcal{A} | L(s)=w_j, L(t) \neq w_j} \frac{1}{d(s,t)}, \quad (2.29)$$

onde $L(t)$ o rótulo do grupo da amostra t , W'_j usa todos os pesos dos arcos entre o grupo w_j e os outros grupos, e W_j usa todos os pesos dos arcos no grupo $w_j = 1, 2, \dots, c$.

A Figura 2.2 mostra um exemplo sintético com 340 amostras de treinamento retirado de [31], que forma alguns grupos com diferentes concentrações de amostras em um espaço de atributos 2D. Dependendo da escala, o número natural de grupos é um, quatro ou cinco. Se $k_{\max} \geq 150$, então o corte mínimo vai ser quando todas as amostras estiverem em um único grupo. O corte mínimo para $k_{\max} = 100$ identifica quatro grupos com o melhor $k = 37$ (Figura 2.2b), e limitando a busca do k ao intervalo $[1, 30]$, o corte mínimo identifica cinco grupos, com o melhor $k = 29$ (Figura 2.2c).

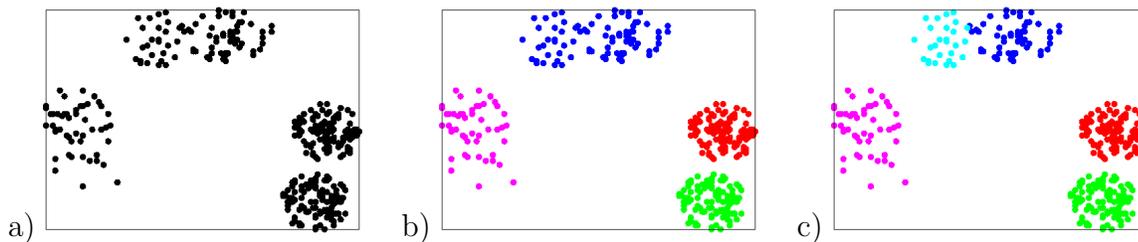


Figura 2.2: (a) Espaço de atributos com diferentes concentrações de amostras por grupo. Pode-se identificar diferentes números de grupos dependendo da escala. Soluções razoáveis seriam (b) quatro e (c) cinco grupos, onde dois deles estariam se tocando. Imagem de [3]

O conjunto \mathcal{R} de raízes é composto por um elemento para cada máximo local da fdp. Entre todos os possíveis caminhos π_t com raiz no máximo da fdp, é desejável encontrar o caminho cuja menor densidade ao logo dele seja máxima. Sendo assim, cada máximo

deve definir uma zona de influência (agrupamento) para selecionar as amostras que são mais fortemente conexas a ele do que a qualquer outro máximo. Os arcos extras em \mathcal{A}' , que garantem simetria na densidade dos platôs, também garantem um único grupo por máximo da fdp. Formalmente, desejável encontrar

$$V(t) = \max_{\forall \pi_t \in \Pi(\mathcal{N}, \mathcal{A}, t)} \{f_{min}(\pi_t)\}, \quad (2.30)$$

onde

$$\begin{aligned} f_{min}(\langle t \rangle) &= \begin{cases} \rho(t) & \text{se } t \in \mathcal{R} \\ \rho(t) - \delta & \text{caso contrário} \end{cases} \\ f_{min}(\langle \pi_s \cdot \langle s, t \rangle \rangle) &= \min\{f_{min}(\pi_s), \rho(t)\} \end{aligned} \quad (2.31)$$

para

$$\delta = \min_{\forall (s,t) \in \mathcal{A}' | \rho(t) \neq \rho(s)} \{|\rho(t) - \rho(s)|\}. \quad (2.32)$$

Valores elevados para delta reduzem o número de máximos.

O Algoritmo 2 identifica uma raiz em cada máximo da fdp ($P(s) = nil$ na Linha 5 implica $s \in \mathcal{R}$), atribui a cada raiz um rótulo de grupo distinto na Linha 6, e calcula a zona de influência (grupo) de cada raiz como uma árvore de caminhos ótimos em P , de tal forma que os nós na árvore recebem o mesmo rótulo que sua raiz no mapa L (Linha 10). Como resultado, temos o mapa de conectividade ótima V , a floresta P e a lista nós \mathcal{L} ordenada em ordem decrescente do seu valor ótimo em V . Esta lista usada posteriormente para propagar rótulos aos pixels $t \in D_I \setminus \mathcal{N}$.

O Algoritmo 2 requer $O(k|\mathcal{N}| + |\mathcal{N}| \log |\mathcal{N}|)$ computações, quando Q um *heap* binário, e a estimativa do melhor k precisa ser computada várias vezes. Isso também explica a necessidade de amostrar um menor conjunto $\mathcal{N} \subset D_I$.

Algoritmo 2 – AGRUPAMENTO DE MÚLTIPLOS RÓTULOS

ENTRADA: Grafo $(\mathcal{N}, \mathcal{A}')$ e função ρ .

SAÍDA: Mapa de rótulos L , mapa de conectividade V , floresta P , e lista \mathcal{L} de nós ordenados.

AUXILIARES: Fila de prioridade Q , variáveis tmp e $l \leftarrow 1$.

1. Para todo $s \in \mathcal{N}$, faça $P(s) \leftarrow nil$, $V(s) \leftarrow \rho(s) - \delta$, insira s em Q .
2. Enquanto Q no estiver vazia, faça
3. Remova de Q a amostra s cujo valor de $V(s)$ seja máximo.
4. Insira s em \mathcal{L} .
5. Se $P(s) = nil$, então
6. $L(s) \leftarrow l$, $l \leftarrow l + 1$, e $V(s) \leftarrow \rho(s)$.
7. Para cada $t \in \mathcal{A}'_s$ e $V(t) < V(s)$, faça
8. Calcule $tmp \leftarrow \min\{V(s), \rho(t)\}$.
9. Se $tmp > V(t)$ então
10. $L(t) \leftarrow L(s)$, $P(t) \leftarrow s$, $V(t) \leftarrow tmp$.

11. $\perp \perp \perp$ Insira t em Q .

De acordo com [31], os rótulos dos grupos obtidos são incrementalmente propagados para o restante das amostras $t \in D_I \setminus \mathcal{N}$ pela seguinte equação.

$$V(t) = \max_{\forall s \in \mathcal{N}, \forall t \in D_I \setminus \mathcal{N}} \{\min\{V(s), \rho(t)\}\}. \quad (2.33)$$

Se $s^* \in \mathcal{N}$ é o nó que satisfaz a equação, então seu rótulo pode ser simplesmente atribuído a t . A condição $(s, t) \in \mathcal{A}'$ na equação anterior implica que t pode ser um k -vizinho mais próximo de s , se t estivesse em \mathcal{N} . Essa condição pode ser facilmente determinada pelo armazenamento da maior distância $\Omega(s)$ entre s e seus k -vizinhos mais próximos no grafo para cada $s \in \mathcal{N}$. Portanto, o nó candidato para s^* deve satisfazer

$$d(s, t) \leq \Omega(s). \quad (2.34)$$

A parte custosa na Equação 2.33 é a computação de $\rho(t)$ para todo $t \in D_I \setminus \mathcal{N}$, que também requer a computação dos k -vizinhos mais próximos de t em \mathcal{N} .

A propagação dos rótulos foi consideravelmente acelerada ao evitar a computação de $\rho(t)$ pela escolha de s^* que satisfaça

$$V(s^*) = \max_{\forall s \in \mathcal{L}, d(s, t) \leq \Omega(s)} \{V(s)\}, \quad (2.35)$$

onde \mathcal{L} é a lista ordenada resultante do Algoritmo 2. É importante salientar que s^* é o primeiro nó em \mathcal{L} que satisfaz a equação, também evitando o cálculo dos $d(s, t)$ para os nós restantes em \mathcal{L} . Este variante é mais rápido e ainda eficaz, produzindo bons resultados.

2.5.2 Agrupamento binário

Conforme descrito no Capítulo 3, o agrupamento de múltiplos rótulos é utilizado para gerar um mapa de pertinência com o(s) objeto(s) de interesse. Neste mapa, podemos assumir que as $|\mathcal{N}_o|$ amostras mais claras, ou seja, com níveis de cinza mais altos, são de objeto e as $|\mathcal{N}_f|$ amostras mais escuras são de fundo. Assim, temos um novo problema de agrupamento binário onde partimos de $\mathcal{N} = \mathcal{N}_o \cup \mathcal{N}_f$ amostras, calculamos uma floresta com dois tipos de rótulos, objeto e fundo, e depois propagamos esses rótulos para os demais pixels em $D_I \setminus \mathcal{N}$. Neste caso, porém, usamos o modelo de grafo proposto em [29], o qual é mais eficiente que o modelo de grafo knn pois não requer achar o parâmetro k . Este modelo tem sido usado com sucesso em classificação supervisionada e como estamos assumindo os rótulos em \mathcal{N} , podemos utilizá-lo diretamente. Adotamos \mathcal{A} de modo que $(\mathcal{N}, \mathcal{A})$ forma um grafo completo, o qual não precisa ser armazenado explicitamente como no caso anterior.

A ideia é estimar amostras representativas de cada grupo (objeto e fundo) em \mathcal{N} , formando um conjunto S de protótipos para estes grupos, e calcular uma floresta de caminhos ótimos com raízes em S , de modo que o grupo da raiz é propagado para as demais amostras de sua árvore no mapa de rótulos L . Adotamos então pesos $d(s, t) = \|\vec{I}(t) - \vec{I}(s)\|$ aos arcos e função f_{max} para resolver este problema:

$$\begin{aligned}
f_{max}(\langle s \rangle) &= \begin{cases} 0 & \text{se } s \in S \\ +\infty & \text{caso contrário} \end{cases} \\
f_{max}(\langle \pi_s \cdot \langle s, t \rangle \rangle) &= \max\{f_{max}(\pi_s), d(s, t)\}
\end{aligned} \tag{2.36}$$

A floresta é um agrupamento binário de caminhos ótimos, o qual pode propagar rótulo de objeto ou fundo para novas amostras $t \in D_I \setminus \mathcal{N}$. O conjunto de protótipos S dentro do conjunto \mathcal{N} é obtido pela computação de uma Árvore Geradora Mínima (MST do inglês *Minimum Spanning Tree*), no grafo $(\mathcal{N}, \mathcal{A})$, e para cada arco (s, t) na MST cujo $\lambda(s) \neq \lambda(t)$ são estabelecidos s e t como os protótipos dos grupos $\lambda(s)$ e $\lambda(t)$, respectivamente. Onde $\lambda(s)$ indica o grupo ao qual $s \in \mathcal{N}$ pertence, podendo ser objeto ou fundo.

Uma vez que o peso total das arestas na MST é mínimo e como f_{max} é utilizado, a MST contém um caminho ótimo para cada $s \in \mathcal{N}$. A escolha dos protótipos como as amostras mais próximas entre grupos objetiva evitar caminhos ótimos para grupos distintos durante o agrupamento. O treinamento é concluído pela execução do Algoritmo 3.

Algoritmo 3 – AGRUPAMENTO BINÁRIO

ENTRADA: Grafo completo \mathcal{N} e protótipos $S \in (\mathcal{N}, \mathcal{A})$ rotulados com a função λ .
SAÍDA: Mapa de rótulos L , mapa de custo V , floresta P , e lista \mathcal{L} de nós ordenados.
AUXILIARES: Fila de prioridade Q , variáveis tmp e $l \leftarrow 1$.

1. Para todo $s \in \mathcal{N}$ faça $V(s) \leftarrow +\infty$
2. Para todo $s \in S$ faça
3. $P(s) \leftarrow nil, V(s) \leftarrow 0$
4. $L(s) \leftarrow \lambda(s)$, insira s em Q .
5. Enquanto Q no estiver vazia, faça
6. Remova de Q a amostra s cujo valor de $V(s)$ seja mínimo.
7. Insira s em \mathcal{L} .
8. Para cada $t \in \mathcal{N}$ e $V(t) > V(s)$, faça
9. Calcule $tmp \leftarrow \max\{V(s), d(s, t)\}$.
10. Se $tmp < V(t)$ então
11. Se $V(t) \neq +\infty$ então remova t de Q .
12. $L(t) \leftarrow L(s)$, $P(t) \leftarrow s$, $V(t) \leftarrow tmp$.
13. Insira t em Q .

A lista \mathcal{L} é usada para agrupamento de novas amostras $t \in D_I \setminus \mathcal{N}$. Seja $s^* \in \mathcal{N}$ a amostra que resolve a equação abaixo para $t \in D_I \setminus \mathcal{N}$

$$V(t) = \min_{\forall s \in \mathcal{N}} \{\max\{V(s), d(s, t)\}\} \tag{2.37}$$

Então $s^* = P(t)$ no caminho ótimo de S para t e a classe $\lambda(R(s^*))$ da raiz $R(s^*)$ deve ser propagada para t (i.e $L(t) \leftarrow \lambda(R(s^*)) = L(s^*)$). Portanto, se os nós $s \in \mathcal{L}$ forem acessados na ordem crescente de $V(s)$ (na qual eles estão armazenados em $s \in \mathcal{L}$), a busca

por s^* pode ser interrompida quando $V(s) < V(s')$ para s' sendo o próximo nó em \mathcal{L} após s .

$$V(t) = \min_{\forall s \in \mathcal{L}} \{\max\{V(s), d(s, t)\}\} \quad (2.38)$$

2.6 Análise comparativa entre as técnicas baseadas em função densidade de probabilidade

O método da IFT de múltiplos rótulos define zonas que equivalem aos domos e suas regiões de influência de uma função de densidade de probabilidade. Este método tem similaridades com os métodos de Deslocamento Médio e da Maximização da Esperança, descritos nas seções 2.3 e 2.4, respectivamente. No entanto, existem diferenças fundamentais entre eles como descritas nesta seção.

A técnica de deslocamento médio objetiva caminhar na direção do gradiente da função de densidade de probabilidade (direção de maior crescimento da fdp) até encontrar um máximo local. Esse gradiente é estimado para cada *pixel* da imagem. O deslocamento médio busca um máximo para cada amostra e rotula a zona de influência de cada máximo em $\rho(\mathbf{x})$; se o máximo for em um platô, amostras irão encontrar pontos diferentes nesse platô e estacionar a busca, e assim, o domo ficar supersegmentado.

Além disso, se um platô definido em uma janela estiver no meio do caminho do vetor deslocamento na busca pelo máximo, ele pode não ser encontrado, pois no platô o gradiente é nulo, satisfazendo a mesma condição de ser um máximo local.

Na técnica de deslocamento médio não existe uma representação em grafos da imagem, isso é uma vantagem pois permite aplicar o algoritmo na imagem inteira, não apenas em um subconjunto de pixels. O algoritmo da IFT não depende do gradiente, pois ele funciona independentemente de estar ou não em um platô. Ele garante isso naturalmente quando define na Equação 2.26 uma adjacência simétrica nos platôs e determina que o agrupamento comece pelos máximos e eleja um ponto máximo para representar o platô. Isso garante que todos os máximos de uma região homogênea sejam conquistados por apenas uma amostra desse platô antes que o rótulo desta seja propagado para o restante das amostras do domo da fdp.

A técnica de Maximização da Esperança é usada para estimação dos parâmetros desconhecidos dos componentes que formam a densidade. Uma das informações que o método necessita é o número de grupos nos quais a imagem será dividida. Contudo, em muitas aplicações essa informação não é conhecida. Se o número de grupos for definido, ele pode forçar que grupos naturalmente diferentes se unam em um único grupo, por exemplo. Se n grupos definem uma classe, é necessária informação adicional sobre o problema (aplicação) para definir isso.

Além disso, o EM encontra a probabilidade *a posteriori* supondo que a probabilidade *a priori* segue um modelo. Nos algoritmos da IFT e do deslocamento médio não é preciso conhecer o número de grupos, pois eles encontram esse número naturalmente, além de não assumirem forma pré-definida para os grupos.

Uma das vantagens potenciais do algoritmo de Maximização da Esperança é no agrupamento de novas amostras. Para cada nova amostra é calculada a probabilidade dela

pertencer a algum grupo nos quais os dados estão divididos. Se o número de grupos for pequeno, o agrupamento se torna proporcionalmente rápido. No entanto, o que é uma vantagem, também pode ser uma desvantagem, se as mostras forem divididas em muitos grupos. A IFT também é rápida, pois o agrupamento é normalmente resolvido sem percorrer a lista toda.

Este trabalho adotou a IFT, porque ela não gera a supersegmentação do deslocamento médio, e porque ela não depende da informação sobre o número de grupos do EM. Porém, o agrupamento por IFT pode ser utilizado como estimativa inicial para o EM, conforme proposto no capítulo 3 e isso é avaliado no capítulo 4.

Capítulo 3

Segmentação de fundo em imagens naturais usando Transformada Imagem-Floresta

A metodologia elaborada para segmentação de fundo de imagens naturais se baseia na possibilidade de gerar um modelo probabilístico para o fundo da imagem. Utilizar este modelo para criar um mapa de pertinência com o objeto. O mapa de pertinência é uma função que associa um valor de 0 a 1 para a pertinência de cada pixel em relação a um objeto de interesse. Depois é realizada uma binarização da imagem usando amostras mais claras e mais escuras selecionadas do mapa de pertinência.

Como na aplicação de segmentação de fundo a classe de interesse é o fundo, um modelo probabilístico deve ser criado para essa classe. Para isso, é preciso definir uma maneira de selecionar um conjunto de amostras da imagem que represente a distribuição de cores do fundo. Como o método proposto é não supervisionado, não existe um usuário indicando onde está o fundo ou onde está o objeto, sendo assim, é utilizada a premissa de que a borda da imagem deve conter amostras do fundo. A imagem a) da Figura 3.1, mostram a região da borda definida como domínio para selecionar as amostras (pixels pintados de branco) exibidas na Figura 3.1b). Portanto, o método se propõe a segmentar imagens cujos objetos de preferência estejam longe da borda, mas isso não é uma limitação total, os objetos podem estar tocando na borda, eles somente não devem ter uma representatividade grande nela.

As amostras do fundo são selecionadas por uma amostragem aleatória com distribuição de probabilidade uniforme da borda da imagem. Um modelo probabilístico é estimado para essas amostras e chamado de *modelo do fundo* da imagem. A partir desse modelo é criada uma imagem nebulosa chamada de *mapa de pertinência* com o objeto. O mapa de pertinência associa a cada pixel da imagem um valor de dissimilaridade com os pixels do fundo, e portanto, uma estimativa de similaridade com a classe dos objetos. Esse mapa é utilizado para a seleção das amostras que serão classificadas como fundo ou como objeto na etapa de classificação binária. O mapa de pertinência com objeto da imagem em questão é exibida na Figura 3.1c).

Na imagem de pertinência, os pixels mais claros têm maior valor de pertinência com



Figura 3.1: Imagens geradas pela metodologia proposta. a) região da imagem definida como borda, b) amostras selecionadas da borda, c) exemplo da imagem de pertinência e d) resultado da classificação binária.

o objeto, por isso um subconjunto desses pixels é escolhido para representar os pixels do objeto. Esses pixels recebem rótulo de objeto. De forma análoga, os pixels mais escuros têm menor valor de pertinência com o objeto, e um subconjunto deles é escolhido para serem os pixels representantes do fundo. Esses pixels recebem rótulo de fundo. A partir desses grupos de amostras, um agrupamento binário, tal como descrito na Seção 2.5.2, é computado e após algum pós-processamento a classificação binária é finalizada.

Para a criação do mapa de pertinência, três variantes do método foram propostos.

1. Mapa direto da função de conectividade ótima;
2. Mapa baseado na mistura de Gaussianas ajustadas aos grupos formados para a borda da imagem;
3. Utiliza a técnica de maximização da esperança para estimar a probabilidade *a posteriori* de um pixel pertencer ou não à classe do fundo e utiliza essa probabilidade como a pertinência.

É importante observar que o *pipeline* proposto para remoção de fundo é extensível para remoção de instâncias de qualquer classe, desde que haja alguma forma de selecionar

algumas amostras dessa classe. Ademais, o método também é extensível a outros variantes, pela associação de valores aos pixels que sejam proporcionais às similaridades deles com determinada classe.

Na Figura 3.2 está exibido um fluxograma com as etapas em que o método está dividido. Todas essas etapas estão descritas e detalhadas neste capítulo.

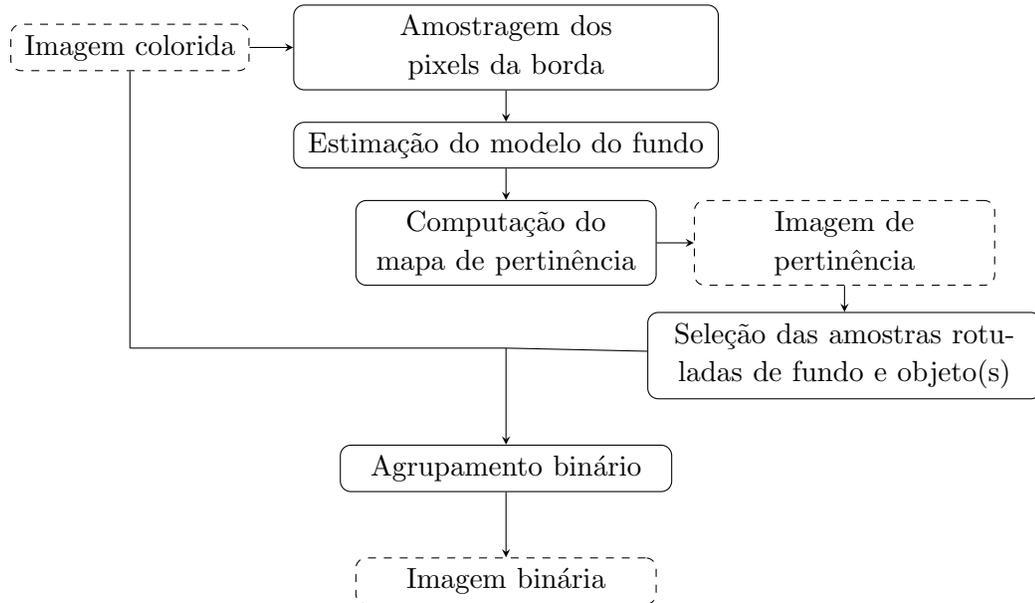


Figura 3.2: Fluxograma da metodologia proposta para segmentação de fundo de imagens coloridas

3.1 Modelo de fundo

A primeira fase do método é computar uma função de densidade de probabilidade para as amostras do fundo da imagem. Esta fase é considerada como um pré processamento para os três variantes do método proposto.

Um conjunto de *pixels* $\mathcal{N}_{\mathcal{B}} \subset D_I$ é selecionado por amostragem aleatória com distribuição de probabilidade uniforme da borda. A espessura da borda deve ser definida de acordo com a aplicação utilizada. Neste trabalho, como as imagens são de bases diferentes e heterogêneas, a espessura da borda passa a ser entrada do usuário. Para essas amostras é criado um agrupamento de múltiplos rótulos baseado na IFT, como descrito na Seção 2.5.1. Esse agrupamento particiona o conjunto $\mathcal{N}_{\mathcal{B}}$ em c grupos \mathcal{N}_j , i.e., $\mathcal{N}_{\mathcal{B}} = \cup_{j=1}^c \{\mathcal{N}_j\}$. Como resultado, além do mapa de rótulos $L(s)$ associado a cada amostra s , é associado também um mapa de conectividade $V(s)$.

O mapa de conectividade V contém para cada amostra $s \in \mathcal{N}_{\mathcal{B}}$ uma força de conectividade com o respectivo máximo, computada utilizando-se as estimativas $\rho(s)$ de densidade no

espaço de atributos. O valor de k escolhido é aquele que gera agrupamento ótimo de acordo com o corte normalizado (Equação 2.27). Como o conjunto de amostras $\mathcal{N}_{\mathcal{B}}$ é selecionado da borda, o mapa de conexidade está sendo chamado de $V_{\mathcal{B}}$, cujo índice \mathcal{B} significa que estão sendo utilizados os pixels de $\mathcal{N}_{\mathcal{B}}$ de amostras da borda.

Na Figura 3.3 é exibido um exemplo de um agrupamento de múltiplos rótulos por IFT para uma imagem natural inteira. O resultado do agrupamento da borda é semelhante ao resultado apresentado nesta imagem. Pode-se observar que o agrupamento separou as “regiões escuras” das “regiões com cor de vegetação” e das “regiões da flor”. Quando apenas os pixels da borda são utilizados no agrupamento tem-se apenas “regiões escuras” da imagem separadas das “regiões de vegetação”, pois não iriam existir grupos representantes da flor, a menos que a borda utilizada fosse muito grande.



Figura 3.3: Agrupamento de múltiplos rótulos em imagem natural

3.2 Estimativa de mapa de pertinência com o objeto

O mapa de pertinência com o objeto indica o quão provável um pixel é de ser considerado ou não amostra do objeto. Para cada variante do método, esse mapa é computado de forma diferente, contudo todos eles são baseados no modelo do fundo, para o qual conhecemos uma estimativa de densidade de probabilidade e as árvores de caminhos ótimos que compõem os grupos.

3.2.1 Mapa de pertinência direto da função de conexidade ótima

Para a criação do mapa de pertinência direto da função de conexidade ótima, o modelo do fundo é a própria função de conexidade encontrada para as amostras $\mathcal{N}_{\mathcal{B}}$, ou seja é o próprio valor de $V_{\mathcal{B}}(s)$ associado a essas amostras.

Para cada amostra $t \in D_I$ é escolhida uma outra amostra $s^* \in \mathcal{N}_{\mathcal{B}}$ que satisfaça

$$V_{\mathcal{B}}(s^*) = \max_{\forall s \in \mathcal{L}, \|\vec{I}(t) - \vec{I}(s)\| \leq \Omega(s)} \{V_{\mathcal{B}}(s)\}, \quad (3.1)$$

onde $\Omega(s)$ é a distância do k-vizinho de s mais distante. \mathcal{L} é a lista ordenada resultante do Algoritmo 2 e s^* é o primeiro nó em \mathcal{L} que satisfaz a equação anterior.

O mapa de pertinência é, portanto, o complemento da ponderação entre o valor de conexidade $V_{\mathcal{B}}(s^*)$ e o valor de t na Gaussiana centrada na amostras s^* :

$$V_{\mathcal{O}}(t) = V_{max} - V_{\mathcal{B}}(s^*) * \exp\left(-\frac{\|\vec{I}(t) - \vec{I}(s^*)\|^2}{K}\right) \quad (3.2)$$

onde

$$K = 2 * \sigma^2 \quad (3.3)$$

cujo σ é definido na Equação 2.25.

Se para um dado $t \in D_I$ não existir nenhuma amostra s^* que satisfaça a Equação 3.1, esse t é considerado um *outlier*, que no contexto da aplicação significa que a amostra t não possui probabilidade nenhuma de ser inserida em algum grupo nos quais a borda foi dividida. Esse *outlier* é automaticamente considerado como amostra do objeto e recebe pertinência máxima com o objeto, denominada de V_{max} .

O Algoritmo 4 contém todas as etapas para a geração do mapa de pertinência com o objeto diretamente da função de conexidade ótima gerada pela IFT.

Algoritmo 4 – GERAÇÃO DO MAPA DE PERTINÊNCIA COM O OBJETO USANDO CONEXIDADE ÓTIMA

ENTRADA: Imagem \hat{I} , lista \mathcal{L} de amostras ordenadas, pixels amostrados da borda $\mathcal{N}_{\mathcal{B}}$, mapa de conexidade $V_{\mathcal{B}}$, constante K e função Ω (armazena Knn mais distante por amostra)
 SAÍDA: Mapa de conexidade $V_{\mathcal{O}}$
 AUXILIARES: Variáveis $dist$ e V_{max}

1. $V_{max} \leftarrow \max_{s \in \mathcal{N}_{\mathcal{B}}} \{V_{\mathcal{B}}(s)\}$
2. Para todo $s \in D_I, V_{\mathcal{O}}(s) \leftarrow V_{max}$
3. Para todo $t \in D_I$, faça
4. $\mathcal{L}' \leftarrow \mathcal{L}$
5. Enquanto \mathcal{L}' não estiver vazia faça
6. Remove o próximo pixel s de \mathcal{L}'
7. $dist \leftarrow \|\vec{I}(t) - \vec{I}(s)\|$
8. Se $dist \leq \Omega(s)$, então
9. $V_{\mathcal{O}}(t) \leftarrow V_{max} - V(s) * \exp(-dist^2/K)$.
10. $\mathcal{L}' \leftarrow \emptyset$

Na linha 1 é encontrado o maior valor de conexidade atribuído aos pixels da borda. Na linha 2, todos os valores de conexidade dos pixels da imagem recebem esse valor de conexidade máxima. Na linha 4 é feita uma cópia da lista ordenada, $\mathcal{L}' \leftarrow \mathcal{L}$. Se a lista \mathcal{L}' não estiver vazia (linha 5), o primeiro elemento s da lista é removido. É calculada a distância entre a cor Lab do pixel s e do pixel t que está sendo observado (linha 3). Se a distância entra as amostras t e s for menor que a distância do k-vizinho mais distante de

s, t é considerado como uma amostra provável de pertencer ao fundo e recebe um valor de pertinência calculada na linha 9; a lista \mathcal{L}' é esvaziada e o algoritmo retorna à linha 3. Observe que se para um certo t não é satisfeita a condição da linha 8, ele é considerado um *outlier* e permanece com seu valor inicial, ou seja, o valor de conectividade máxima.

Na Figura 3.4 são apresentados três exemplos de mapa de pertinência utilizando conectividade ótima.

3.2.2 Mapa de pertinência com Mistura de Gaussianas

Para a criação do mapa de pertinência com Misturas de Gaussianas, também é utilizado o resultado do agrupamento de múltiplos rótulos por IFT, contudo, não é o valor de $V_{\mathcal{B}}$ que é utilizado e sim o particionamento no qual a borda foi dividida.

Uma Gaussiana é ajustada para cada um dos grupos. Para cada grupo $\mathcal{N}_j = 1, 2, \dots, c$ encontrado com o agrupamento de múltiplos rótulos são calculadas a média μ_j , a matriz de covariância Σ_j e a probabilidade *a priori* P_j :

$$\mu_j = \frac{1}{|\mathcal{N}_j|} \sum_{\forall s \in \mathcal{N}_j} \vec{I}(s) \quad (3.4)$$

$$\Sigma_j = \frac{1}{|\mathcal{N}_j|} \sum_{\forall s \in \mathcal{N}_j} [\vec{I}(s) - \mu_j][\vec{I}(s) - \mu_j]^T \quad (3.5)$$

$$P_j = \frac{|\mathcal{N}_j|}{|\mathcal{N}_{\mathcal{B}}|} \quad (3.6)$$

Para cada par (μ_j, Σ_j) existe uma Gaussiana associada criando um modelo de mistura Gaussianas:

$$\sum_{j=1}^c P_j \exp \left[-\frac{1}{2} D_M(s, \mathcal{N}_j)^2 \right] \quad (3.7)$$

Para a computação do mapa de pertinência é preciso calcular a distância de Mahalanobis D_M entre todas amostras $t \in D_I$ e as Gaussianas que representam os grupos \mathcal{N}_j . A distância de Mahalanobis é calculada pela equação

$$D_M(t, \mathcal{N}_j) = \sqrt{[\vec{I}(t) - \mu_j]^T \text{Inv}(\Sigma_j) [\vec{I}(t) - \mu_j]} \quad (3.8)$$

O mapa de pertinência é computado da seguinte maneira

$$V_{\mathcal{O}}(t) = \max \left\{ P_j \exp \left[-\frac{1}{2} D_M(t, \mathcal{N}_j)^2 \right] \right\}, \text{ para } j = 1, 2, \dots, \quad (3.9)$$

ou seja, para cada amostra t a Gaussiana utilizada para o cálculo do mapa de pertinência é aquela que maximiza o valor de $V_{\mathcal{O}}(t)$.

O Algoritmo 5 contém todas as etapas para a geração do mapa de pertinência com o objeto usando misturas de Gaussianas inicializadas pelo agrupamento de múltiplos rótulos por IFT.

Algoritmo 5 – GERAÇÃO DO MAPA DE PERTINÊNCIA UTILIZANDO MISTURAS DE GAUSSIANAS

ENTRADA: Imagem \hat{I} , pixels amostrados da borda $\mathcal{N}_{\mathcal{B}}$, mapa de conexidade $V_{\mathcal{B}}$
 SAÍDA: Mapa de pertinência $V_{\mathcal{O}}$
 AUXILIARES: Variável $dist$ e tem

1. Para todo $\mathcal{N}_j \in \mathcal{N}_{\mathcal{B}}$, calcula μ_j, Σ_j e P_j , $j = 1, 2, \dots, c$
2. Para todo $t \in D_I$, faça
3. $dist \leftarrow 0$
4. Para todo $\mathcal{N}_j \in \mathcal{N}_{\mathcal{B}}$, faça
5. $temp \leftarrow \{P_j \exp[-\frac{1}{2}D(s, \mathcal{N}_j)^2]\}$
6. Se $temp \geq dist$, então
7. $dist \leftarrow temp$
8. $V_{\mathcal{O}}(t) \leftarrow dist$.

Na linha 1 são calculadas a média, a matriz de covariância e a probabilidade a priori para cada grupo \mathcal{N}_j da borda. Na linha 5 é calculada para cada amostra t da imagem (linha 2) a distância ponderada para os grupos \mathcal{N}_j (linha 4). É atribuído ao mapa de pertinência (linha 8) a maior distância ponderada (linha 7) computada para cada grupo. Quanto maior a distância menor é a probabilidade dessa amostra pertencer ao grupo.

Na Figura 3.4 são apresentados três exemplos de mapa de pertinência utilizando Misturas de Gaussianas.

3.2.3 Mapa de pertinência com maximização de esperança

O modelo do mapa de pertinência de Gaussianas misturadas considera que existe uma Gaussiana localizada em cada grupo encontrado pelo agrupamento de múltiplos rótulos por IFT. O modelo de mapa de pertinência usando maximização de esperança utiliza inicialmente o modelo de Gaussianas misturadas e calcula a probabilidade a posteriori de cada pixel pertencer a cada Gaussiana, depois recalcula as Gaussianas (média, matriz de covariância e probabilidade a priori) utilizando a probabilidade a posteriori de cada amostra. Esse processo é repetido n vezes até um determinado critério ser alcançado.

A inicialização do mapa de pertinência é exatamente a mesma do modelo de Gaussianas, ou seja, são utilizadas as equações de média (Equação 3.4), matriz de covariância (Equação 3.5) e probabilidade a priori (Equação 3.6).

Depois, a probabilidade a posteriori de uma amostra s estar contida no grupo \mathcal{N}_j na iteração i é computada da seguinte maneira:

$$P^{(i)}(\mathcal{N}_j \setminus s) = \frac{\left[Det \left(\Sigma_j^{(i-1)} \right) \right]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} D_M^{(i)}(s, \mathcal{N}_j)^2 \right] P_j^{(i-1)}}{\sum_{k=1}^c \left[Det \left(\Sigma_k^{(i-1)} \right) \right]^{-\frac{1}{2}} \exp \left[-\frac{1}{2} D_M^{(i)}(s, \mathcal{N}_k)^2 \right] P_k^{(i-1)}} \quad (3.10)$$

Em seguida, o vetor média, a matriz de covariância e a probabilidade *a priori* são atualizados para cada grupo $\mathcal{N}_j = 1, 2, \dots, c$, visando a próxima iteração:

$$\psi(t) = \frac{P^{(i)}(\mathcal{N}_j \setminus s)}{\sum_{j=1}^c P^{(i)}(\mathcal{N}_j \setminus s)} \quad (3.11)$$

$$\mu_j^{(i)} = \sum_{\forall t \in \mathcal{N}_B} \vec{I}(t) \psi(s) \quad (3.12)$$

$$\Sigma_j^{(i)} = \sum_{\forall t \in \mathcal{N}_B} [\vec{I}(t) - \mu_j^{(i)}][\vec{I}(t) - \mu_j^{(i)}]^T \psi(s) \quad (3.13)$$

$$P_j^{(i)} = \frac{1}{|\mathcal{N}_B|} \sum_{\forall t \in \mathcal{N}_B} \psi(s) \quad (3.14)$$

Quando forem finalizadas as n iterações, o mapa de pertinência fica então calculado da seguinte maneira:

$$V_{\mathcal{O}}(t) = \max \left\{ \exp \left[-\frac{1}{2} D_M(t, \mathcal{N}_j)^2 \right] P_j^{(n)} \right\}, \text{ para } j = 1, 2, \dots, c, \quad (3.15)$$

cuja distância de Mahalanobis utiliza os valores atualizados para média e matriz de covariância.

$$D_M(t, \mathcal{N}_j) = \sqrt{[\vec{I}(t) - \mu_j^{(n)}]^T \text{Inv} \left(\Sigma_j^{(n)} \right) [\vec{I}(t) - \mu_j^{(n)}]} \quad (3.16)$$

Algoritmo 6 – GERAÇÃO DO MAPA DE PERTINÊNCIA UTILIZANDO MAXIMIZAÇÃO DA ESPERANÇA

ENTRADA: Imagem \hat{I} , pixels amostrados da borda \mathcal{N}_B , número de iterações n , mapa de rótulos L_B

SAÍDA: Mapa de conexidade $V_{\mathcal{O}}$

AUXILIARES: Variáveis $dist$, $temp$ e $i \leftarrow 1$

1. Para todo $\mathcal{N}_j \in \mathcal{N}_B, j = 1, 2, \dots, c$, inicializa $\mu_j^{(0)}, \Sigma_j^{(0)}$ e $P_j^{(0)}$ utilizando L_B
2. Enquanto $i \leq it$
3. Para todo $\mathcal{N}_j \in \mathcal{N}_B$, faça
4. Para todo $s \in \mathcal{N}_B$, faça
5. └ calcula $P^{(i)}(\mathcal{N}_j \setminus s)$ utilizando $\mu_j^{(i-1)}, \Sigma_j^{(i-1)}$ e $P_j^{(i-1)}$
6. └ atualiza $\mu_j^{(i)}, \Sigma_j^{(i)}$ e $P_j^{(i)}$
7. └ $i \leftarrow i + 1$
8. Para todo $t \in D_I$, faça
9. └ $dist \leftarrow 0$
10. Para todo $\mathcal{N}_j \in \mathcal{N}_B$, faça
11. └ $temp \leftarrow \{P_j \exp [-\frac{1}{2} D_M(s, \mathcal{N}_j)^2]\}$
12. └ Se $temp \geq dist$, então
13. └ $dist \leftarrow temp$
14. └ $V_{\mathcal{O}}(t) \leftarrow dist$.

O modelo do fundo é inicializado na linha 1, quando são calculados os parâmetros da Gaussiana inicial. Na linha 5, a probabilidade a posteriori de cada pixel (linha 4) pertencer a cada grupo (linha 3) é calculada. Para a primeira iteração (linha 2), são utilizados os valores da Gaussiana criada na linha 1, então, na linha 6, a Gaussiana é atualizada com o valor da probabilidade a posteriori encontrado na linha 5. Uma nova iteração (linha 2) começa, e o processo se repete por it iterações. Esse é o modelo do fundo. Para a criação do mapa de pertinência é calculado para toda amostra $t \in D_I$ um valor de pertinência para cada grupo \mathcal{N}_j na linha 11. É escolhido o maior valor de pertinência para cada amostra (linha 13) e o mapa é formado.

Na Figura 3.4 são apresentados três exemplos de mapa de pertinência utilizando conectividade ótima, mistura de Gaussianas e maximização da esperança. Os mapas de pertinência dessas imagens exibidas são os que geraram melhores acurácias para cada método. Para as três imagens (do cavalo, do homem e da mulher) é notável a diferença entre os mapas gerados usando o algoritmo de maximização da esperança e os mapas gerados usando conectividade ótima e misturas de Gaussianas.

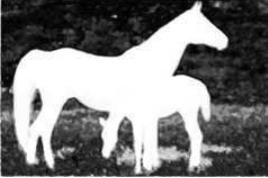
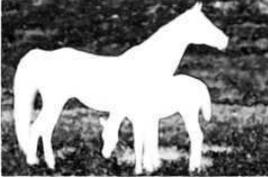
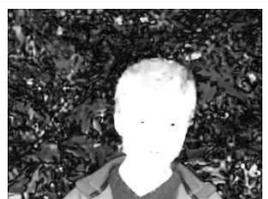
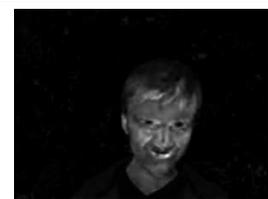
| Imagem Original | Mapa de pertinência baseado em | | |
|---|---|--|---|
| | Conexidade ótima | Mistura de Gaussianas | Maximização da Esperança |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Figura 3.4: Exemplos de Mapa de pertinência usando conectividade ótima.

A imagem do cavalo apresenta os melhores mapas, isso é alcançado porque o objeto não está tocando na borda e porque não existem grupos na borda que sejam semelhantes aos grupos do objeto. Na imagem do homem uma parte do objeto toca na borda, por isso, a diferença entre o objeto e fundo diminui um pouco, mas mesmo assim o objeto está destacado em relação ao fundo. Já na imagem da mulher, o EM classificou todo o vestido e o sapato como sendo de fundo, isso aconteceu porque, existem grupos na borda,

os grupos que contêm as amostras que ficam entre um degrau e outro ou entre uma pedra e outra, que são mais escuros por causa da sombra, e por isso semelhantes aos grupos do vestido e do sapato.

3.3 Classificação binária

Após a transformação da imagem colorida em uma imagem de pertinência, essa imagem é utilizada para auxiliar o processo de classificação binária. Para classificar um pixel da imagem natural como sendo da classe de objetos ou de fundo é utilizado o agrupamento binário por IFT como descrito em 2.5.2. O método de classificação binária considera que os mapas de pertinência das imagens contêm amostras claras associadas às amostras de objetos e as amostras escuras associadas às amostras de fundo (veja Figura 3.4).

Para fazer a classificação, um conjunto de amostras de treinamento $\mathcal{N}_{\mathcal{T}} \subset D_I$ é escolhido. Para compor esse conjunto são escolhidos um subconjunto das amostras mais claras e um subconjunto das amostras mais escuras do mapa de pertinência. Essas amostras são rotuladas como objeto e fundo, respectivamente, protótipos dessas duas classes são selecionados e conquistam cada uma das amostras de $\mathcal{N}_{\mathcal{T}}$ formando uma floresta de caminhos ótimos.

A lista \mathcal{L} é novamente utilizada para classificação de novas amostras $t \in D_I \setminus \mathcal{N}_{\mathcal{T}}$. Cada amostra t é conquistada por uma árvore das amostras de treinamento $\mathcal{N}_{\mathcal{T}}$ e recebem o rótulo do grupo.

3.4 Pós-processamento

Para todas as imagens, em qualquer variante do algoritmo, foi aplicado o mesmo pós-processamento. Como as imagens possuem um padrão variado, ou seja, não são de uma mesma aplicação, o pós-processamento realizado pode ser considerado genérico. Para uma aplicação específica (controlada), o pós-processamento utilizado deve ser adaptado às imagens utilizadas.

Para as imagens apresentadas no capítulo 4, o pós-processamento é dividido em duas etapas:

1. Seleção do número de objetos desejados - baseado em tamanho. O caso padrão é considerar que o objeto de interesse é o que apresentar maior área. Porém, em alguns casos existe mais de um objeto de interesse. Se o número de objetos desejados for n , os n maiores objetos são selecionados. A Figura 3.5 apresenta um caso onde pode ser desejado recuperar mais de um objeto. Neste caso, foi desejado recuperar dois objetos presentes na imagem.
2. Fechamento morfológico. A operação morfológica fechamento é utilizada para que orifício(s) interno(s) ao objeto(s) seja(m) fechado(s). Porém, muitas vezes é desejável que orifício(s) permaneça(m) no(s) objeto(s). Veja um exemplo na Figura 3.6, na imagem b) está o resultado do método sem nenhum pós-processamento; além dos dois orifícios grandes no núcleo da flor, outros pequenos orifícios são observados. O fechamento morfológico vai eliminar os orifícios com área menor ou igual a um certo limiar.



Figura 3.5: Resultado da metodologia proposta usando como pós-processamento o número de objetos b) $n = 1$ e c) $n = 2$ para a) imagem original

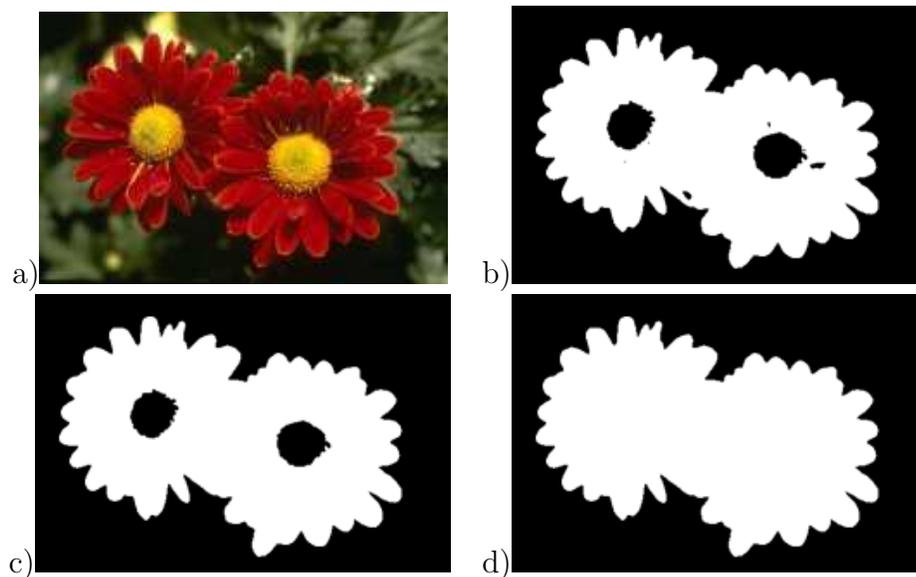


Figura 3.6: Fechamento morfológico. a) imagem original, b) o resultado da segmentação sem pós-processamento. c) imagem após fechamento morfológico com limiar 300, e d) fechamento morfológico com limiar suficiente para fechar todos os burados internos ao objeto

3.5 Análise comparativa das abordagens

Em termos de eficiência, a abordagem utilizando o mapa de pertinência baseado em IFT é relativamente mais rápida que a utilizando misturas de Gaussianas, justamente pela não necessidade do cálculo do modelo de mistura de Gaussianas. O modelo utilizando EM é o menos eficiente que os outros dois modelos na proporção do número de iterações necessárias.

O modelo de misturas de Gaussianas e o modelo EM precisam ser inicializados com o número estimado de Gaussianas cuja borda estaria dividida. Isso é uma desvantagem dessas duas abordagens porque nem sempre é conhecido esse valor. Para resolver esse problema foi proposta uma inicialização dos grupos encontrados pelo agrupamento por Transformada Imagem Floresta.

O agrupamento por Transformada Imagem Floresta cria grupos com formas ar-

bitrárias. Ao ser utilizado o modelo de Gaussianas Misturadas, estamos dizendo que os grupos têm as formas de Gaussiana. Ao utilizarmos o modelo de maximização da esperança, estamos supondo que, além dos grupos terem formas de Gaussianas, eles não estão bem dimensionados ou localizados em relação aos grupos gerados pela IFT.

Portanto, todo o processo começa com o agrupamento Transformada Imagem Floresta, segue para uma mistura de Gaussianas e depois para um modelo baseado em maximização da esperança. Esta é uma forma de tentar ajustar os modelos dos grupos ao que realmente eles representam.

Em relação resultados que podem ser encontrados, podemos analisá-los da seguinte forma:

- Se os melhores resultados forem com o agrupamento por Transformada Imagem Floresta, isto significa que além dos grupos encontrados estarem bem localizados, eles não são necessariamente Gaussianos.
- Se o modelo de Gaussianas apresentar melhores resultados, significa dizer que o agrupamento por Transformada Imagem Floresta seria útil para estimar o número de grupos e onde eles estariam localizados, mesmo que a forma não fosse exatamente a encontrada pela IFT.
- Se o modelo baseado em EM apresentar os melhores resultados, duas conclusões podem ser entendidas:
 1. se o número de iterações for pequeno, o agrupamento por Transformada Imagem Floresta foi útil para inicializar, e o modelo formado pela IFT e EM obteve sucesso;
 2. se o número de iterações for grande, o agrupamento por Transformada Imagem Floresta não foi útil, e não é o ideal para inicialização.

Capítulo 4

Experimentos e Resultados

Nesta seção são apresentados a metodologia para avaliação do método e os resultados obtidos; a base de dados utilizada; uma breve descrição de como os padrões ouro das imagens foram gerados e a medida de acurácia utilizada para avaliar os resultados.

4.1 Base de imagens

Para avaliar o método proposto, a base de imagens utilizada é composta por 50 imagens retiradas de [26] e [34] e de projetos do grupo. Pelo fato da borda da imagem ser utilizada para representar as amostras do fundo, a restrição para escolha das imagens foi a de que os objetos, preferencialmente, estivessem longe da borda, como nas Figuras 4.1a e 4.1b, mas isso não é uma limitação, como pode ser visto nas Figuras 4.1c e 4.1d. O objeto pode estar tocando na borda, ele somente não deve ter uma representatividade grande de pixels nela, e apesar das cenas serem de imagens naturais, os pixels do fundo devem se distinguir dos pixels de objeto.

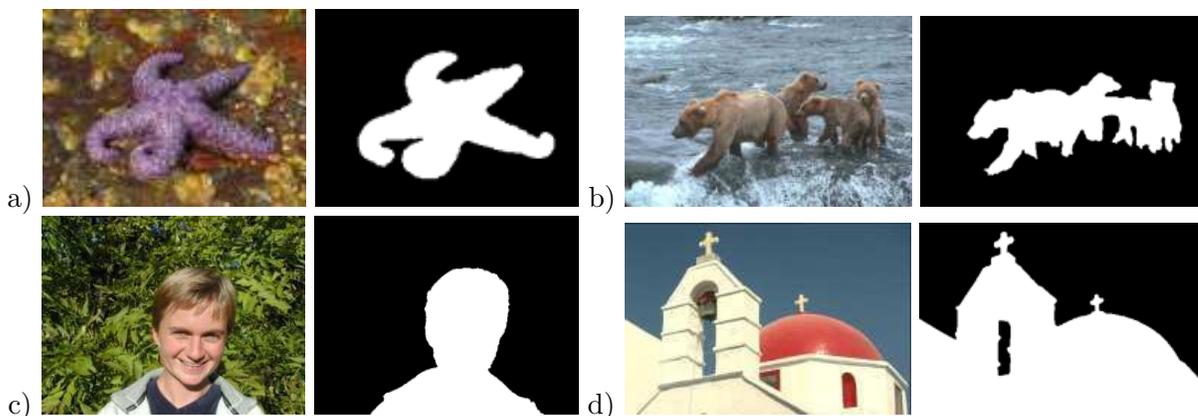


Figura 4.1: Exemplos de imagens que compõem a base de dados e seus respectivos *Padrões Ouro*.

Para cada uma das imagens contidas na base foi criado um *Padrão Ouro*. O padrão ouro é a imagem referência do resultado desejado para a segmentação da imagem. A base

do padrão ouro foi gerada por uma segmentação interativa por IFT com competição de sementes, usando o algoritmo de IFT diferencial utilizando imagem de gradiente [36]. Na Figura 4.2 é mostrado como o padrão ouro é criado.

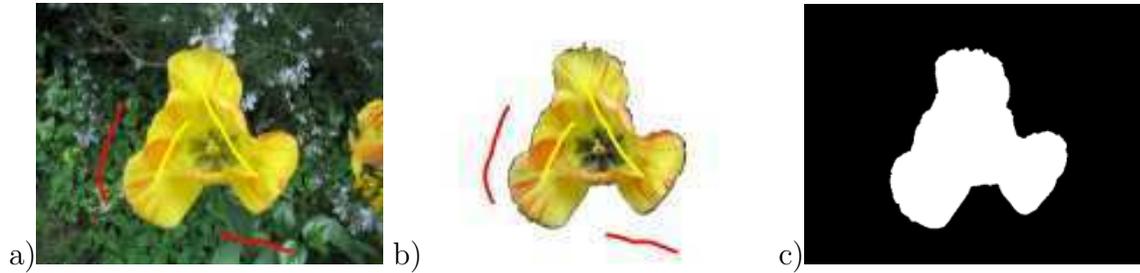


Figura 4.2: Processo de criação dos *Padrões Ouro*. a) marcadores desenhados na imagem, b) resultado da segmentação e c) o padrão ouro.

Marcadores são desenhados nos objetos e no fundo, como na Figura 4.2a, onde os marcadores amarelos são de objeto, os vermelhos são de fundo. Protótipos são escolhidos nas amostras que estão sob os marcadores para conquistarem as outras amostras e segmentar a imagem como uma floresta de caminhos ótimos, 8-conexos, com raízes nos marcadores de objeto, Figuras 4.2b e 4.2c.

4.1.1 Medida de acurácia

A validação da metodologia proposta foi feita pela medição do quão perto uma imagem segmentada está do seu respectivo padrão ouro. Essa medição foi realizada utilizando-se a medida de acurácia da Equação 4.3, que pode ser explicada da seguinte forma. Seja N o número de pixels do objeto na imagem \vec{I} . Erros podem ser definidos pelas seguintes equações

$$e_1 = \frac{FP}{|D_I| - N} \quad e \quad e_2 = \frac{FN}{N}, \quad (4.1)$$

onde FP e FN são falsos positivos e falsos negativos, respectivamente. FP é o número de amostras do fundo que foram equivocadamente definida com parte do(s) objeto(s) e FN é o número de amostras do(s) objeto(s) que foram equivocadamente segmentadas como parte do fundo.

Os erros e_1 e e_2 são utilizados para computar o erro total:

$$E = e_1 + e_2 \quad (4.2)$$

onde E é a soma dos erros parciais, e_1 e e_2 , na classificação do objeto.

Como só existem dois grupos, os erros e_1 e e_2 do objeto correspondem aos erros e_2 e e_1 do fundo, respectivamente. Sendo assim, a soma E dos erros parciais do fundo e do(s) objeto(s) possuem o mesmo valor e por isso não precisa ser novamente calculada. Por fim, a medida de acurácia para o caso binário é computada da seguinte maneira, para uma dada imagem \vec{I} :

$$Acc(\vec{I}) = 1 - \frac{E}{2} \quad (4.3)$$

A versão geral da Equação 4.3 para c classes, onde $c \geq 3$, pode ser encontrada em [29].

4.2 Metodologia experimental

O objetivo da metodologia experimental foi o de encontrar uma boa configuração de parâmetros para os métodos propostos. Para alguns parâmetros foi possível fixar valores que geram bons resultados médios para a base utilizada. Como a base é heterogênea, alguns parâmetros não puderam ser fixados e foram definidos como entrada de usuário.

Para cada parâmetro foi definido um intervalo de valores, e todos os valores nestes intervalos foram testados. Então, para cada imagem foi escolhida a configuração que gerou o melhor resultado. Com base nesses resultados foi tomada a decisão de quais parâmetros poderiam ou não ser fixados.

Após a definição dos valores fixos dos parâmetros, foi feito um treinamento com a permutação apenas dos parâmetros definidos como sendo variáveis para simular um usuário utilizando o método. O melhor resultado possível que o usuário poderia encontrar, dados os parâmetros fixados, foi definido por imagem. Esses resultados são exibidos na Seção 4.3 e analisados na Seção 4.5.

4.2.1 Definição dos parâmetros

Nesta seção são apresentados quais parâmetros foram fixados e quais são entradas do usuário para os três variantes do método.

Parâmetros de entrada do usuário

A simulação da utilização dos parâmetros de usuário foi feita atribuindo-se os valores ótimos definidos para os parâmetros fixos juntamente com o conjunto de possíveis configurações permitidas aos parâmetros variáveis, encontradas na permutação do conjunto de valores definidos a seguir:

- *espessura da borda*: a espessura da borda variou de 1 a 100 pixels de largura;
- *número de objetos*: as imagens que formavam a base apresentavam apenas *um* ou *dois* objetos de interesse;
- *pós-processamento*: escolha entre *fechamento* de todos os orifícios ou apenas dos pequenos, considerados ruídos;
- *número de iterações*: esse número variou de 1 a 100 (parâmetro exclusivo do variante do método baseado em maximização da esperança)

Parâmetros fixos

Para os três variantes, os dois parâmetros fixados foram: número de amostras selecionadas da borda e o intervalo percentual do valor de K (utilizado no computo da fdp do fundo). Para esses parâmetros, os valores definidos foram:

- *número de amostras selecionadas da borda*: para escolha do melhor conjunto de amostras o número máximo de amostras da borda foi fixados em 1000 amostras, e depois foi feita uma busca de 100 até 1000 para verificar se existia outro valor que melhorasse o resultado, e este valor foi fixado em 200. A Tabela 4.1 mostra a acurácia média das imagens para cada tamanho do conjunto.

Tabela 4.1: Acurácia média por tamanho do conjunto de amostras da borda

| Número de amostras | Acurácia média (desvio padrão) |
|--------------------|---|
| 100 | 0,9279 ($\pm 0,1352$) |
| 200 | 0,9664 ($\pm 0,0972$) |
| 300 | 0,9526 ($\pm 0,1066$) |
| 400 | 0,9401 ($\pm 0,1405$) |
| 500 | 0,9328 ($\pm 0,1277$) |
| 600 | 0,9029 ($\pm 0,1840$) |
| 700 | 0,9080 ($\pm 0,1673$) |
| 800 | 0,9203 ($\pm 0,1562$) |
| 900 | 0,9074 ($\pm 0,1420$) |
| 1000 | 0,9099 ($\pm 0,1554$) |

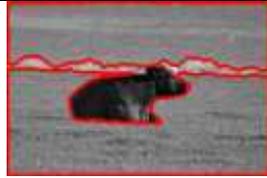
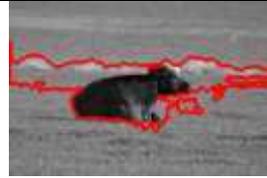
Como pode ser observado na Tabela 4.1, a variação da acurácia média não é homogênea. A conclusão para esse comportamento foi que, para uma amostragem de 100 pixels, a distribuição de probabilidade não fica semelhante a do fundo, por escassez de amostras. Com um número maior de pixels, a amostragem de 200 gera bons resultados. À medida que o número de pixels aumenta, também aumentam as chances de selecionar amostras de regiões não representativas do fundo, como regiões do próprio objeto, e as acurácias diminuem.

- *valor ótimo de k* : valor ótimo de k é obtido por corte mínimo no grafo (Equação 2.27), com base nos agrupamentos obtidos para valores de k entre $k_{min} = 10\%$ e $k_{max} = 40\%$ do número de amostras selecionadas do fundo;

4.3 Resultados

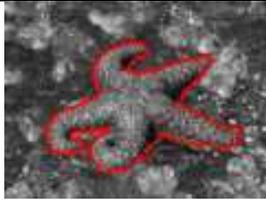
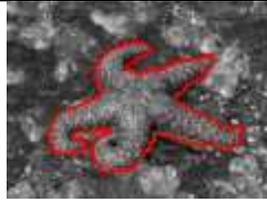
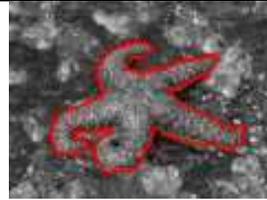
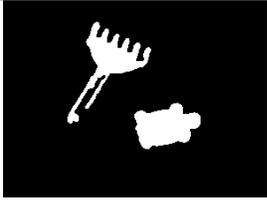
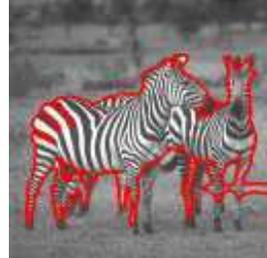
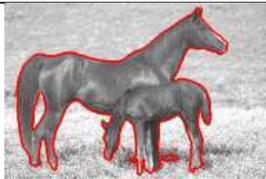
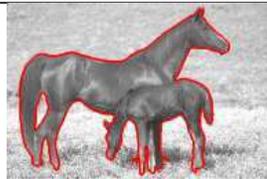
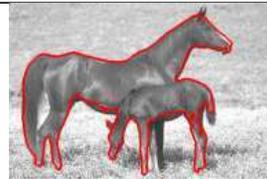
Utilizando os valores dos parâmetros fixos definidos, os melhores resultados encontrados para cada variante do método a partir da simulação das possíveis configurações geradas pelo usuário estão exibidos na Tabela 4.2.

Tabela 4.2: Melhor resultado por algoritmo

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|---|---|---|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 01 |  |  |  0,9872 |  0,9582 |  0,9147 |
| 02 |  |  |  0,9873 |  0,9807 |  0,9828 |
| 03 |  |  |  0,9324 |  0,6325 |  0,5571 |
| 04 |  |  |  0,9923 |  0,9909 |  0,9814 |

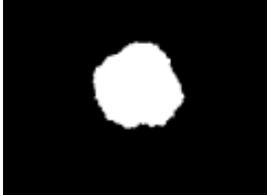
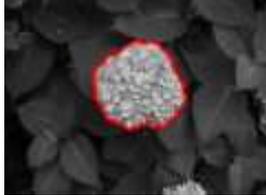
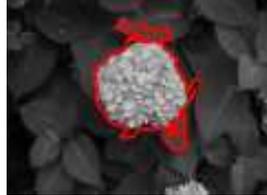
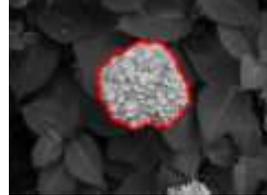
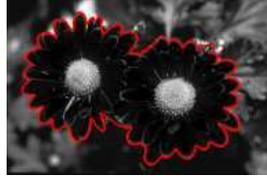
Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|--|---|---|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 05 |  |  |  0,9803 |  0,9773 |  0,9641 |
| 06 |  |  |  0,9622 |  0,9281 |  0,9563 |
| 07 |  |  |  0,9741 |  0,9667 |  0,8838 |
| 08 |  |  |  0,9890 |  0,9859 |  0,9305 |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|--|--|---|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 09 |  |  |  0,9978 |  0,9797 |  0,9973 |
| 10 |  |  |  0,9939 |  0,9944 |  0,9904 |
| 11 |  |  |  0,9847 |  0,9507 |  0,9546 |

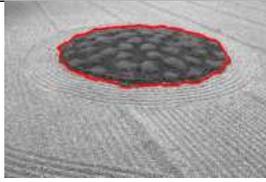
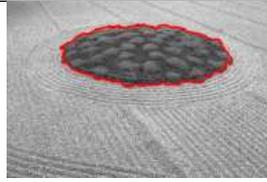
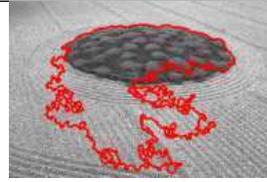
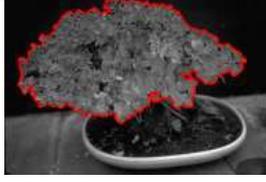
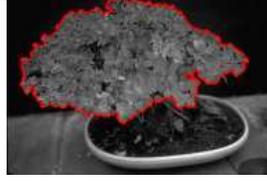
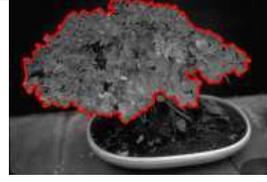
Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|--|--|---|--|---|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 12 |  |  |  0,9845 |  0,9813 |  0,9818 |
| 13 |  |  |  0,9503 |  0,9487 |  0,9673 |
| 14 |  |  |  0,9816 |  0,9625 |  0,9824 |

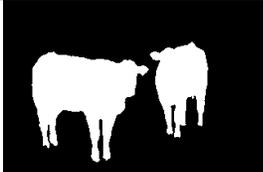
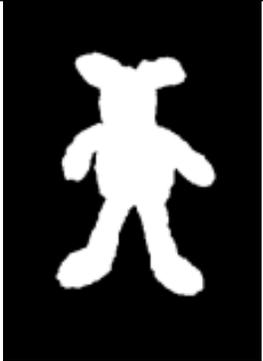
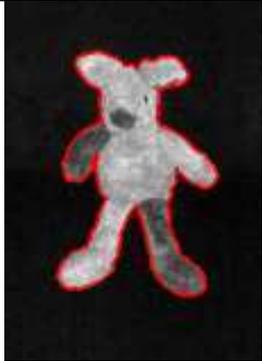
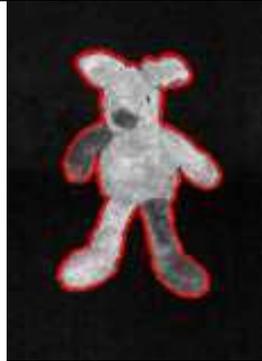
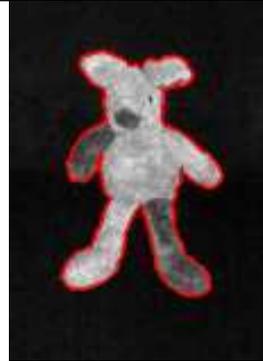
Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|---|---|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 15 |  |  |  0,9397 |  0,8866 |  0,8958 |
| 16 |  |  |  0,9836 |  0,9846 |  0,9877 |
| 17 |  |  |  0,9970 |  0,9930 |  0,8273 |
| 18 |  |  |  0,9947 |  0,9940 |  0,9935 |

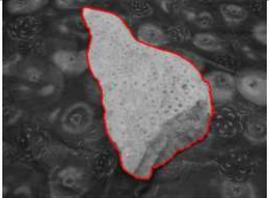
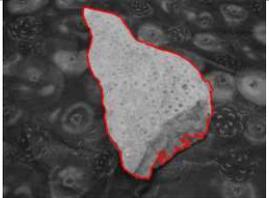
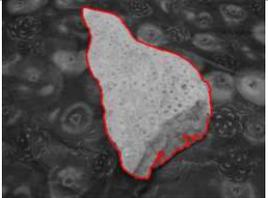
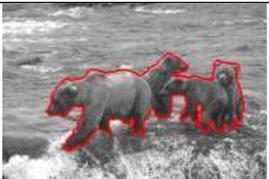
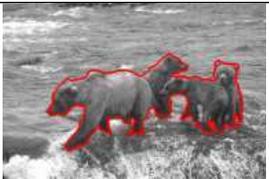
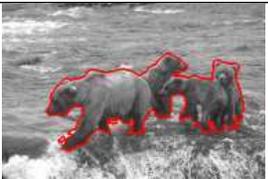
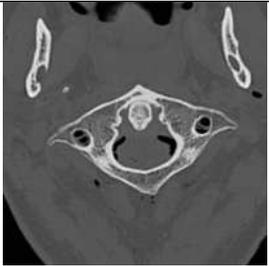
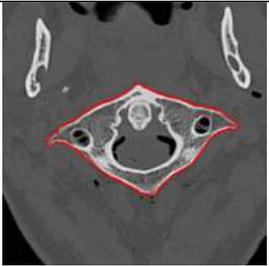
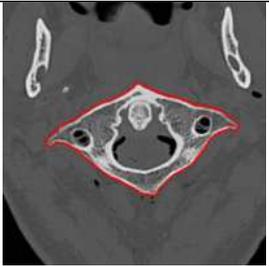
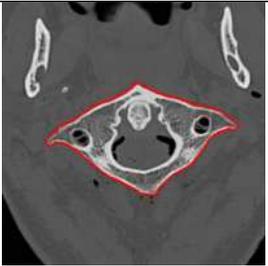
Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|--|--|--|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 19 |  |  |  0,9771 |  0,9522 |  0,9251 |
| 20 |  |  |  0,9899 |  0,9923 |  0,9965 |
| 21 |  |  |  0,9931 |  0,9916 |  0,6606 |
| 22 |  |  |  0,9943 |  0,9872 |  0,9937 |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

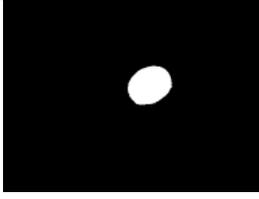
| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|---|---|---|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 23 |  |  |  0,9659 |  0,8901 |  0,8006 |
| 24 |  |  |  0,9964 |  0,9846 |  0,9884 |
| 25 |  |  |  0,9882 |  0,9815 |  0,9846 |
| 26 |  |  |  0,9983 |  0,9973 |  0,9978 |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|----------------------------|--|--|--|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 27 |  |  |  0,9921 |  0,9916 |  0,9721 |
| 28 |  |  |  0,9836 |  0,9910 |  0,9924 |
| 29 |  |  |  0,9792 |  0,9459 |  0,9503 |
| 30 |  |  |  0,9943 |  0,9823 |  0,9937 |
| Continua na próxima página | | | | | |

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|--|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 31 |  |  |  0,9957 |  0,9948 |  0,9939 |
| 32 |  |  |  0,9897 |  0,9751 |  0,9572 |
| 33 |  |  |  |  |  |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

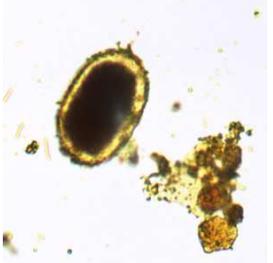
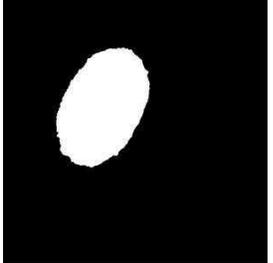
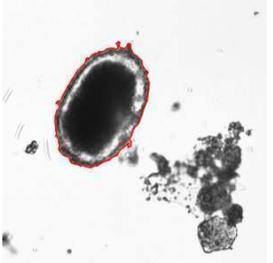
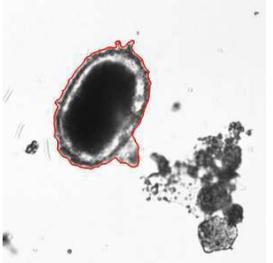
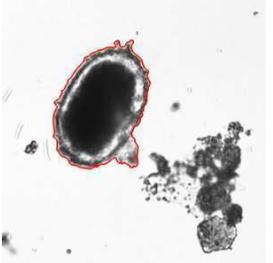
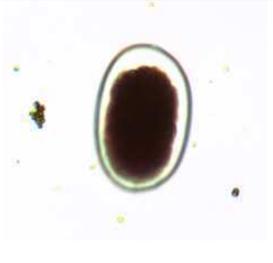
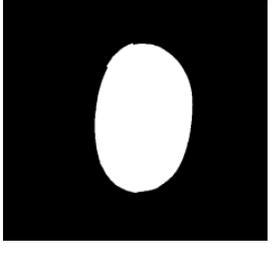
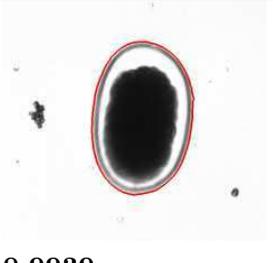
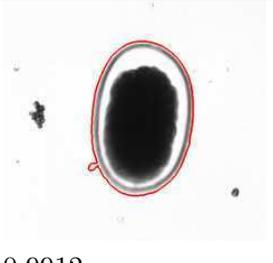
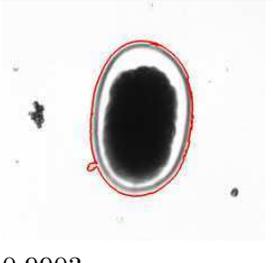
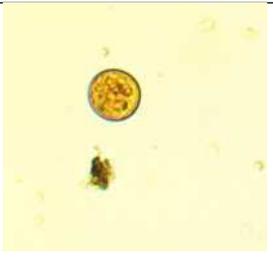
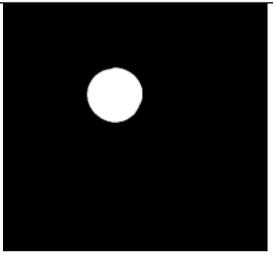
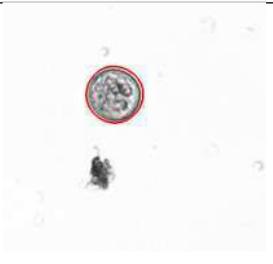
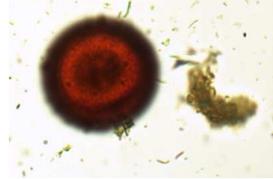
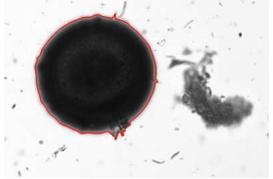
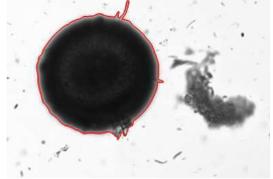
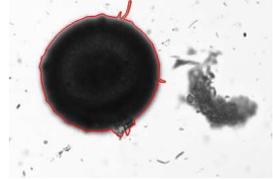
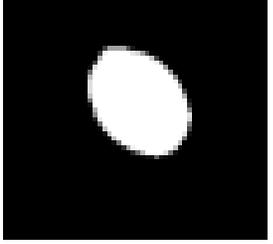
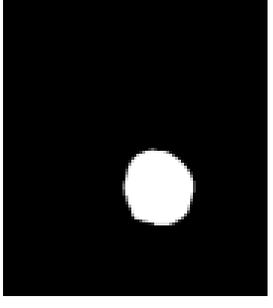
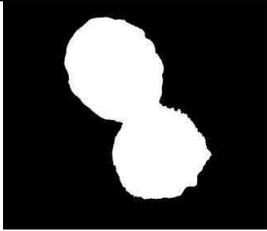
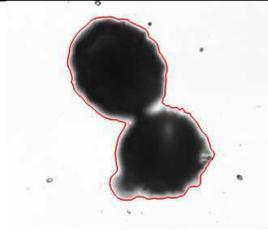
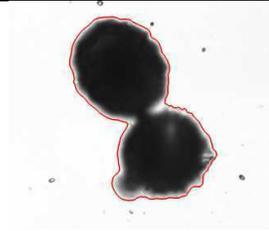
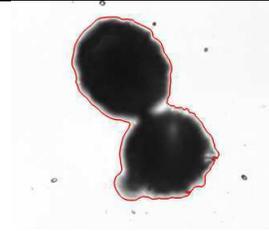
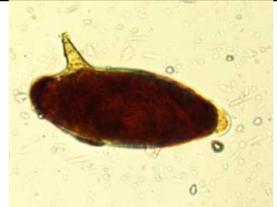
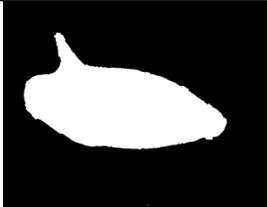
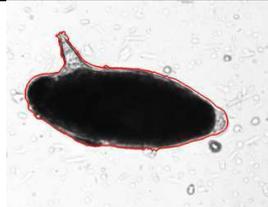
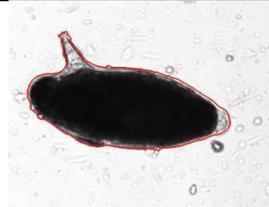
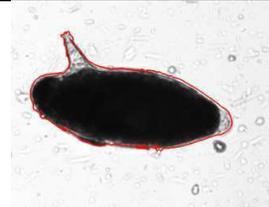
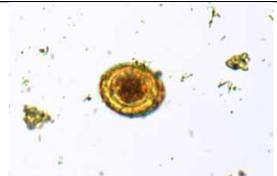
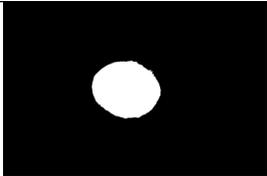
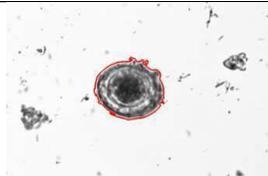
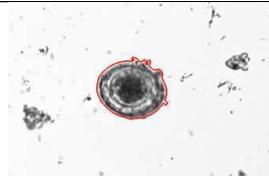
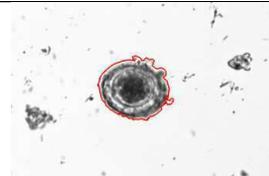
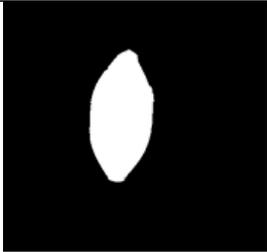
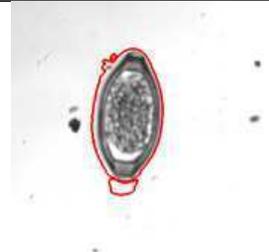
| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|----------------------------|--|--|--|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| | | | 0,9986 | 0,5161 | 0,9973 |
| 34 |  |  |  0,9977 |  0,9922 |  0,9916 |
| 35 |  |  |  0,9939 |  0,9912 |  0,9903 |
| 36 |  |  |  0,9990 |  0,9988 |  0,9973 |
| Continua na próxima página | | | | | |

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|--|--|---|---|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 37 |  |  |  0,9965 |  0,9955 |  0,9828 |
| 38 |  |  |  0,9955 |  0,7465 |  0,9933 |
| 39 |  |  |  0,9835 |  0,9906 |  0,9584 |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|---|--|---|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 40 |  |  |  0,9939 |  0,9938 |  0,9936 |
| 41 |  |  |  0,9905 |  0,9914 |  0,9828 |
| 42 |  |  |  0,9975 |  0,9978 |  0,9955 |
| 43 |  |  |  |  |  |

Continua na próxima página

Tabela 4.2 – Continuação da página anterior

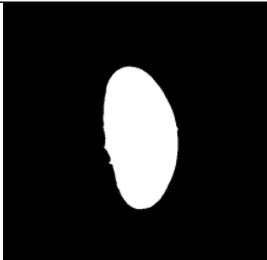
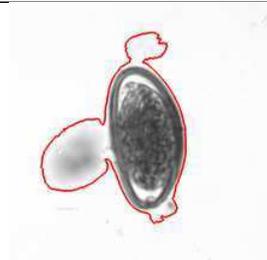
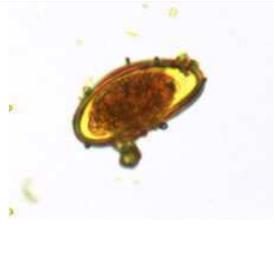
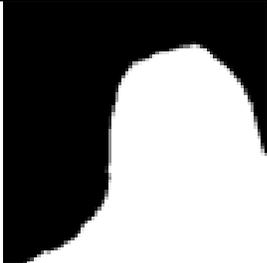
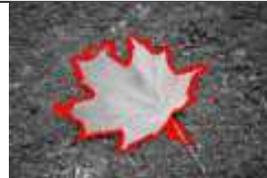
| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|----------------------------|--|--|--|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| | | | 0,9938 | 0,9926 | 0,9880 |
| 44 |  |  |  0,9977 |  0,9732 |  0,9502 |
| 45 |  |  |  0,9913 |  0,9895 |  0,9854 |
| 46 |  |  |  0,9709 |  0,9699 |  0,6340 |
| Continua na próxima página | | | | | |

Tabela 4.2 – Continuação da página anterior

| Índice | Imagens | | Resultados por algoritmos em acurácia | | |
|--------|---|---|---|--|--|
| | Original | Padrão Ouro | IFT | IFT + GMM | IFT + EM |
| 47 |  |  |  0,3956 |  0,8634 |  0,8630 |
| 48 |  |  |  0,9725 |  0,4665 |  0,9799 |
| 49 |  |  |  0,9923 |  0,9918 |  0,9907 |
| 50 |  |  |  0,9873 |  0,9869 |  0,9777 |

É importante lembrar que a acurácia mede acertos no objeto e no fundo, e não apenas no objeto. O entendimento inadequado desse conceito pode gerar confusão na análise dos resultados. Um exemplo é a imagem de índice 46, se apenas o objeto tivesse sido considerado, teria-se uma acurácia próxima de 33%. Entretanto, como o fundo e o objeto são considerados separadamente e equitativamente (ou seja, 50% do valor da acurácia provêm dos acertos no fundo e os 50% restantes provêm dos acertos no objeto), os acertos no objeto contribuíram com cerca de 16% ($0.33 * 50\%$) e os acertos no fundo contribuíram com cerca de 47% ($0.95 * 50\%$), o que dariam uma acurácia aproximada de 64%, portanto, próxima a acurácia exata encontrada de 63,4%.

O resultado médio por algoritmo está exibido na Tabela 4.3. Nela estão exibidos os valores média e desvio padrão para acurácia e tempo de processamento. Os experimentos foram realizados em um computador DELL Inspiron N4030, processador Intel Core, i3-380M, 2.53 GHz e memória RAM 4GB DDR3-1333MHZ.

Tabela 4.3: Resultado médio por algoritmo

| Mapa baseado em | acurácia média | tempo médio (desvio padrão) em segundos |
|-----------------|-------------------------|---|
| IFT | 0,9664 ($\pm 0,0972$) | 1,367 ($\pm 0,916$) |
| IFT + GMM | 0.9440 (± 0.1123) | 1,919 ($\pm 0,1358$) |
| IFT + EM | 0,9456 ($\pm 0,0939$) | 2,488 ($\pm 1,908$) |

4.4 Binarização do Mapa de pertinência

Para justificar o uso de um classificador binário no lugar de um algoritmo de limiarização ótima do mapa de pertinência, comparamos a classificação binária da imagem colorida com o algoritmo de Otsu [28], que é um dos algoritmos mais referenciados na literatura [33]. O algoritmo de Otsu seleciona um limiar ótimo tal que ele minimiza a soma ponderada da variância dos pixels de fundo e de objeto.

O fluxograma para a modificação introduzida pode ser visualizado na Figura 4.4. A binarização com Otsu foi feita diretamente no mapa de pertinência, gerado pela IFT apenas, pois os melhores resultados para classificação binária foram encontrados utilizando esta variante, como pode ser observado na Tabela 4.2. A acurácia para os melhores resultados e o tempo médio para executar em cada imagem estão exibidos na Tabela 4.4.

Mesmo que o algoritmo do Otsu tenha gerado um bom resultado na limiarização do mapa de pertinência, o classificador binário aplicado à imagem colorida obteve melhor acurácia. O algoritmo de Otsu é mais rápido que a classificação binária mas a melhora na acurácia justifica o uso do classificador no lugar do limiarizador. O bom resultado de acurácia obtido com o algoritmo de Otsu também revela que o método de criação dos mapas de pertinência gerou boas estimativas de densidade.

Outra abordagem utilizada para a etapa de binarização foi criar um método híbrido entre o Otsu e o agrupamento binário por IFT. Nesta abordagem, o Otsu seria utilizado para fazer uma binarização do mapa de pertinência. Considerado que o resultado é uma imagem binária, os pixels brancos estão associados aos objetos e os pixels pretos ao fundo.

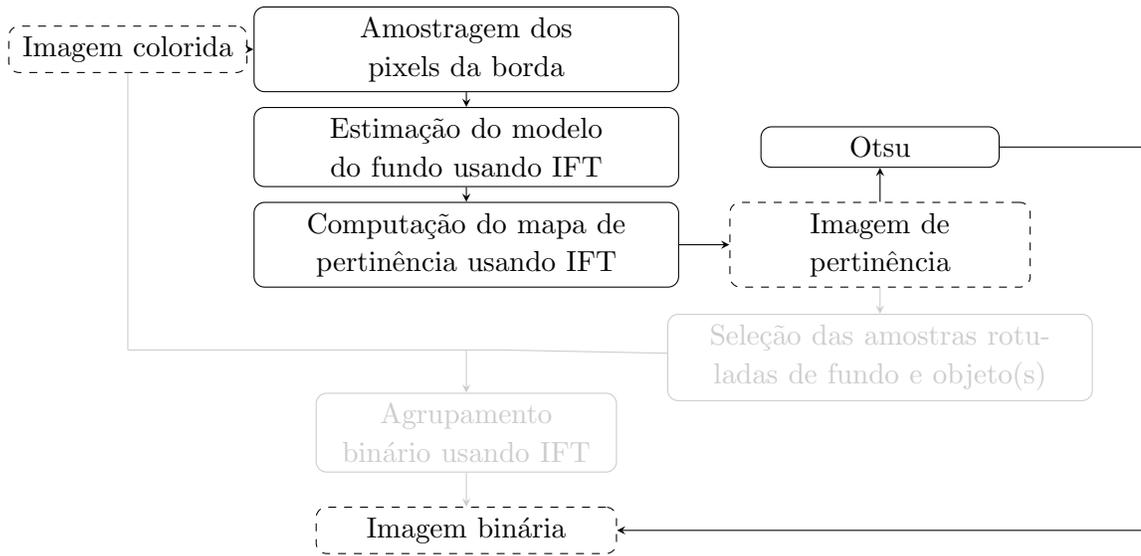


Figura 4.3: Modificação no fluxograma para abordagem em que o mapa de pertinência é binarizado pelo algoritmo de Otsu para gerar o resultado final.

Tabela 4.4: Resultado do teste com Otsu

| Algoritmo | acurácia média | tempo médio (desvio padrão) em segundos |
|--------------------|-------------------------|---|
| Classificador IFT | 0,9664 ($\pm 0,0972$) | 1,367 ($\pm 0,916$) |
| Binarizador Otsu | 0,9557 ($\pm 0,1010$) | 0,739 ($\pm 0,524$) |
| Híbrido Otsu + IFT | 0,9620 ($\pm 0,0928$) | 1,138 ($\pm 0,749$) |

Sendo assim, foi possível fazer um amostragem aleatória com distribuição de probabilidade uniforme da imagem binária e, selecionar amostras rotuladas de objeto e fundo. Essas amostras são utilizadas pelo agrupamento binário por IFT para gerar a segmentação final. A modificação do fluxograma adicionada para representar esta abordagem é ilustrada na Figura 4.4. A acurácia para os melhores resultados e o tempo médio para executar em cada imagem estão exibidos na Tabela 4.4.

4.5 Análise dos resultados

De uma forma geral, a distribuição de probabilidade proveniente dos grupos do fundo (borda) não é sempre uma mistura de Gaussianas. Isso pode ser constatado pelo fato de que poucas imagens tiveram seus valores de acurácia melhorados devido à transformação dos grupos com formas arbitrárias, gerados pelo agrupamento por IFT, para grupos com formatos Gaussianos, independente de terem sido geradas pelo mapa de pertinência utilizando GMM ou pelo mapa de pertinência utilizando EM, por exemplo, as imagens de índices.

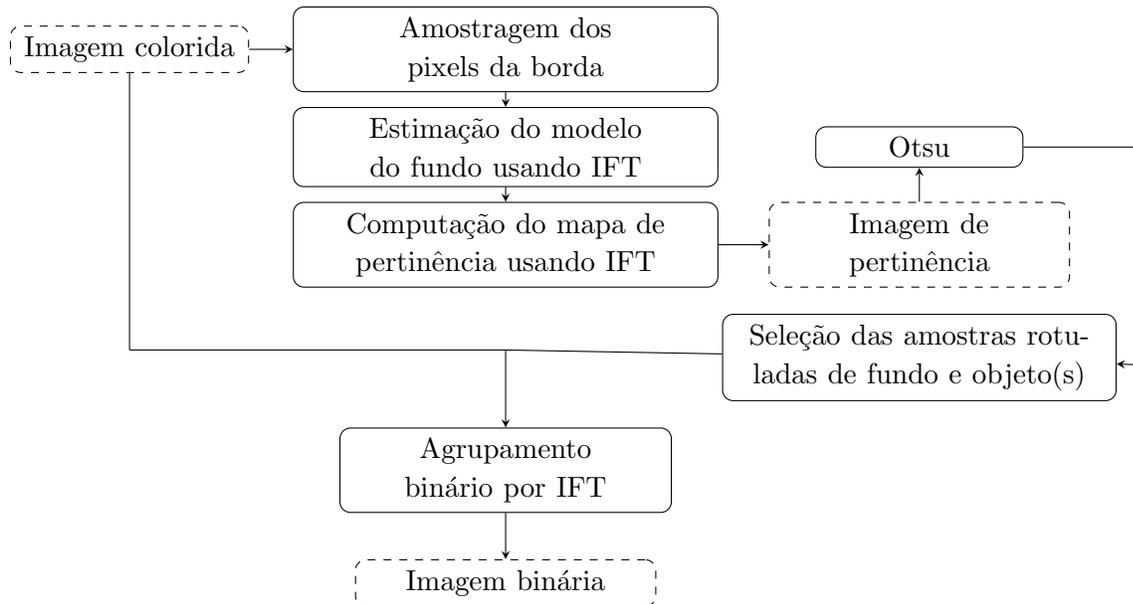


Figura 4.4: Modificação no fluxograma para abordagem híbrida entre o algoritmo de Otsu e o agrupamento binário por IFT

Para as imagens da base utilizada, 80% (40 de 50) dos melhores resultados foram gerados pelo mapa de pertinência inteiramente baseado em IFT, 8% (4 de 50) foram com a IFT + GMM e 12% (6 de 50) foram a IFT + EM.

Para imagens semelhantes à imagem de índice 10 da Tabela 4.2, cujas diferenças entre as acurácias para os três variantes são pequenas (em torno de 0.01), pode-se considerar que os métodos geram resultados equivalentes. Este cenário se repete para 58% das imagens da base.

Os resultados para os 42% das imagens restantes podem ser analisados nos seguintes termos:

- *IFT melhor do que GMM e EM*: os grupos não devem ser Gaussianos. 16% das imagens se enquadram nessa situação.

Um exemplo é a imagem da Figura 4.5. Nesta imagem os grupos iniciais do objeto se deslocam dos pixels do “urso” e se aproximam dos pixels da “vegetação”. Ao transformar os grupos em Gaussianas, além dos grupos de pixels que formam o “urso”, os grupos de pixels que foram a “vegetação” são considerados de objeto. Depois de 9 iterações, utilizando EM, os grupos que deveriam ser de objeto começam a conter grupos do fundo. É razoável atribuir essa mudança ao fato da sombra do urso tocar na borda, e da existência de uma região escura na parte superior esquerda da imagem. Pois ambas as regiões tocam na borda e são similares ao objeto.

- *IFT semelhante à GMM e melhor do que EM*: os grupos devem ser Gaussianos. A

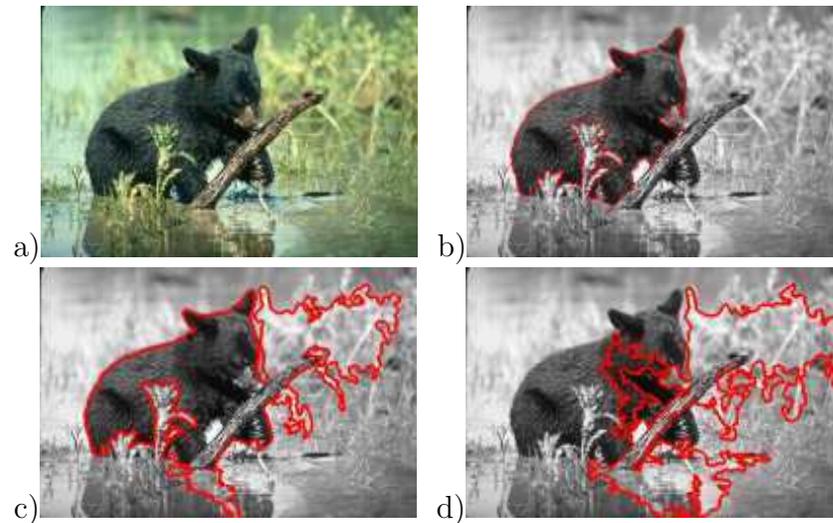


Figura 4.5: Exemplo de melhor resultado obtido pela IFT. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d)

IFT definiu os grupos de forma satisfatória, mas EM divergiu do resultado ótimo. 14% dos resultados se caracterizam desta forma. Um exemplo é a imagem da Figura 4.6.

- *IFT semelhante à EM e melhor do que GMM*: o resultado da IFT é um agrupamento satisfatório, que não é Gaussiano e, portanto, a transformação dos grupos para um modelo de GMM piora o resultado. Depois, o EM refina o GMM até encontrar um modelo de mistura de Gaussianas com parâmetros distintos, igualmente satisfatório ao da IFT. 10% dos resultados apresentam este comportamento. Exemplo exibido na Figura 4.7.
- *GMM melhor do que IFT e EM*: não houve ocorrência.
- *GMM semelhante à EM e melhor do que IFT*: os grupos devem ser Gaussianos, e o agrupamento encontrado pela IFT não foi satisfatório, mas serviu para inicializar o modelo de Gaussianas e EM.

A imagem da Figura 4.8 é a que mais se aproxima desse comportamento. Se as imagens geradas pelo GMM e EM forem comparadas entre si, elas não são semelhantes, mas se ambas forem comparadas com a gerada pela IFT, elas apresentam um comportamento similar.

O método baseado na IFT considerou que a parede era o objeto. Isso é razoável de aceitar considerando que a parede tem uma pequena representatividade de pixels na borda, e as regiões da mulher ocupam uma área razoavelmente grande da borda.

- *EM melhor que IFT e GMM*: não houve ocorrência.
- *IFT, GMM e EM totalmente diferentes*: não houve ocorrência.

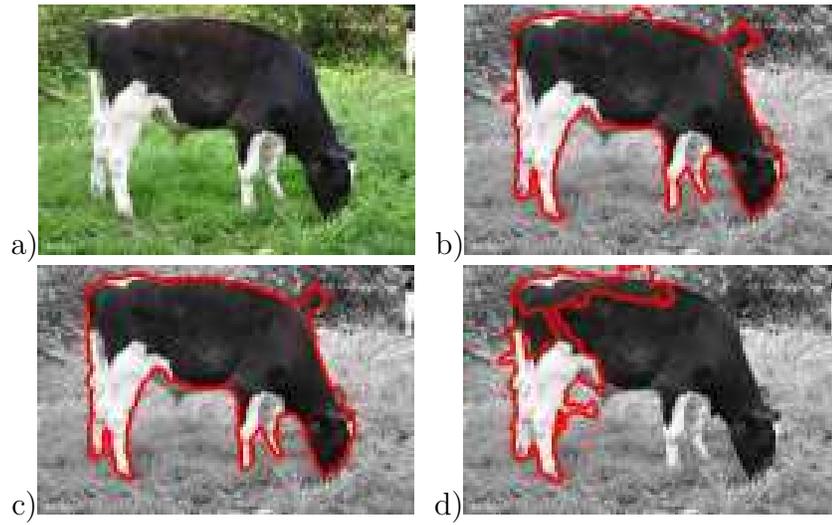


Figura 4.6: Exemplo do caso onde o EM diverge do resultado ótimo. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d)

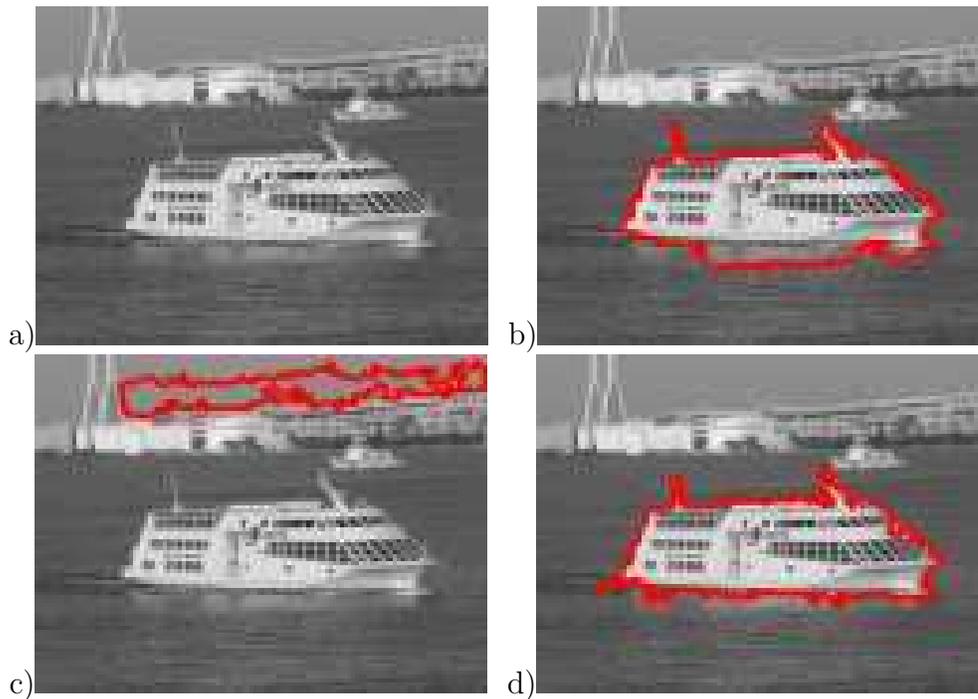


Figura 4.7: Exemplo de resultados semelhantes para IFT e EM. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d)

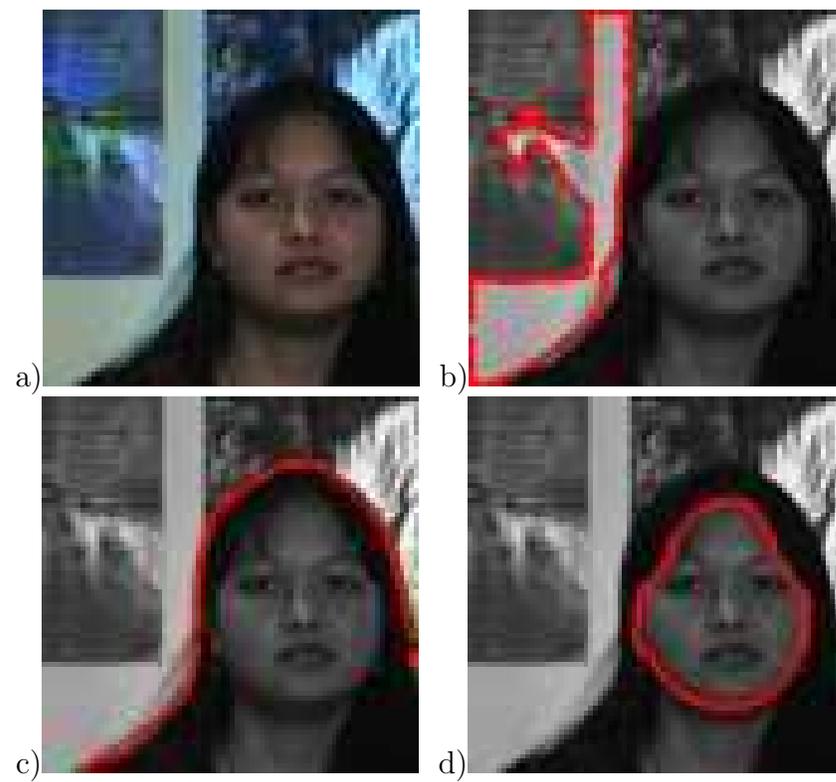


Figura 4.8: Exemplo de resultados para EM e GMM melhores que IFT. Imagem original a), resultado usando IFT b), resultado usando GMM c) e resultado usando EM d)

Capítulo 5

Considerações Finais

Este trabalho apresentou uma metodologia para segmentação automática de objetos de interesse de imagens naturais baseada em segmentação de fundo. Nos experimentos realizados definimos objeto de interesse como o maior (ou maiores, dependendo da aplicação) objeto que se destaca na imagem. E, diferentemente de outros métodos de segmentação de fundo, o objeto não precisa necessariamente estar centrado, mas deve possuir o mínimo possível de pixels na região assumida como fundo da imagem, que neste trabalho foi a borda desta com uma dada espessura.

A metodologia proposta para segmentação utiliza agrupamentos de pixels baseados em funções de densidade de probabilidade. Estudamos a viabilidade de utilização de alguns algoritmos existentes na literatura. Estudamos a transformada imagem floresta, o deslocamento médio e maximização da esperança. Percebemos que o deslocamento médio não apresentava vantagens teóricas em relação à transformada imagem floresta, pelo fato dele gerar supersegmentação. Contudo, a maximização da esperança nos permitia, a partir de grupos gerados pela IFT, melhorar a estimativa desses grupos.

Por isso, propomos uma abordagem híbrida entre a transformada imagem floresta e a maximização da esperança, além de uma abordagem híbrida entre a transformada imagem floresta e a mistura de Gaussianas, visto que esta é a etapa inicial da abordagem anterior. Além dessas duas abordagens, propomos outra que utiliza somente o resultado do agrupamento da transformada imagem floresta.

De forma geral, a metodologia foi dividida em três etapas: 1) criação do modelo do fundo 2) criação do mapa de pertinência e 3) binarização da imagem:

1. O modelo do fundo é uma função de densidade de probabilidade estimada para os pixels selecionados da borda da imagem. O agrupamento de múltiplos rótulos baseado na IFT foi utilizado para gerar este modelo.
2. O mapa de pertinência é um modelo nebuloso onde os pixels de objetos são mais claros que os pixels de fundo. Os métodos de agrupamento estudados foram utilizados para gerar o modelo. Sendo assim, três variantes do modelo foram criadas baseadas: na transformada imagem floresta, em mistura de Gaussianas e em maximização da esperança.

3. A binarização da imagem foi feita a partir de três abordagens:

- (a) binarização do mapa de pertinência pelo método do Otsu;
- (b) binarização da imagem colorida pelo agrupamento binário por IFT, utilizando um percentual das amostras mais claras e mais escuras selecionadas do mapa de pertinência;
- (c) método híbrido entre o algoritmo de Otsu e o agrupamento binário por IFT. O agrupamento binário utiliza um percentual de amostras de objeto e fundo rotuladas pelo método do Otsu;

Os experimentos foram realizados com uma base heterogênea composta por 50 imagens naturais, selecionadas de forma a satisfazerem a premissa de que o objeto (ou os objetos) não deve ter uma representatividade grande na borda e deve se distinguir dos outros objetos ao redor dele.

De uma forma geral, a metodologia apresentou bons resultados, onde as três variantes propostas tiveram acurácia superior a 94% em média. De forma mais específica, o mapa de pertinência que gerou melhores resultados foi aquele inteiramente baseado na Transformada Imagem Floresta, com acurácia em torno de 96%. Os mapas gerados pelos métodos híbridos apresentaram resultados equivalentes entre si, ambos com 94% de acurácia.

A necessidade de um método de agrupamento binário foi também avaliada. A conclusão foi que, em relação à binarização do mapa de pertinência gerada pelo algoritmo de Otsu, o agrupamento binário aumenta a acurácia dos resultados, visto que a limiarização por Otsu obteve 95% contra 96% de acurácia do agrupamento binário. No entanto, o método híbrido para binarização (Otsu + IFT) apresentou acurácia equivalente, 96%, além de apresentar a vantagem de não ser necessária a utilização de parâmetros para o algoritmo do Otsu, isso deixa o método mais robusto. O único parâmetro passa a ser o número de amostras utilizadas no agrupamento binário.

Após todas as análises realizadas, concluímos que a melhor configuração para o método de segmentação fundo de imagens naturais baseado na metodologia proposta seria: 1) o agrupamento binário por IFT para gerar a função de densidade de probabilidade (ou modelo do fundo) das amostras selecionadas da borda, 2) utilização do mapa de pertinência baseada na IFT e 3) o método híbrido entre Otsu e IFT para binarização da imagem colorida. O fluxograma para esta configuração está exibido na Figura 5.

De forma geral, os resultados encontrados contribuíram para o estado da arte nos seguintes aspectos:

- Proposta de um método simples e eficiente baseado em agrupamento por IFT para segmentação não supervisionada de fundo, gerando excelentes resultados, que atingem acurácia de 96% em média.
- Proposta das técnicas híbridas “transformada imagem floresta e mistura de Gaussianas” (IFT + GMM) e “transformada imagem floresta e maximização da esperança” (IFT + EM) para segmentação não supervisionada de fundo, atingindo acurácia de 94%; a IFT é utilizada para indicar o número de grupos e amostras

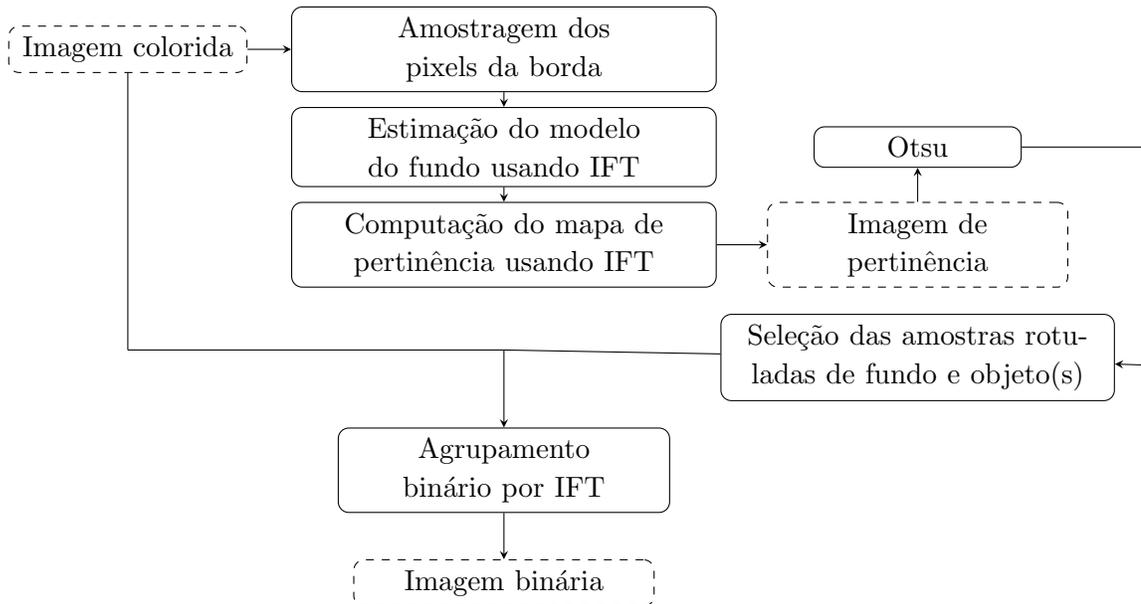


Figura 5.1: Melhor configuração da metodologia proposta

iniciais por grupo para as técnicas de GMM e EM, já que elas dependem desses parâmetros.

- Proposta de um método de agrupamento para binarizar imagens, ganhando em eficácia do binarizador Otsu, que minimiza otimamente a variância intra-grupo.
- Proposta de um método híbrido entre o algoritmo de Otsu e o agrupamento binário por IFT para binarização de imagens coloridas.

Os resultados do trabalho abrem oportunidade para outros trabalhos futuros, tais como:

- Utilizar histograma de cores da imagem para melhorar a determinação do número de amostras selecionadas da borda. Pode-se calcular a diferença entra vários histogramas, e o que possuir menor distância média em relação a todos os histogramas pode ser escolhido, e o conjunto de amostras que o gerou pode ser utilizados para criação do modelo do fundo.
- Fazer extensão do método aplicações específicas de segmentação de imagens. Uma possível aplicação é a etada de segmentação para análise de imagem parasitária no processamento de amostra fecal. Este projeto está sendo desenvolvido pelo LIV-IC/UNICAMP - Laboratório de Informática Visual, do Instituto de Computação da Universidade Estadual de Campinas.

- Fazer extensão do método para segmentação de fundo em vídeos. Nesse caso pode-se usar como fundo uma região de interesse em torno do objeto segmentado do quadro anterior, ou mesmo estimar movimentos e deslocar a região de interesse de acordo com a estimativa.
- Fazer extensão do método para Recuperação de Imagens por Conteúdo (CBIR). O descritor da imagem passa a ser o descritor do objeto ou fundo segmentado.
- Estudar outros métodos para estimação de densidade que possam ser usados para estimação do modelo do fundo.
- Fazer um estudo comparativo experimental mais amplo com outros métodos para agrupamento de pixels da literatura.

Referências Bibliográficas

- [1] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using em and its application to content-based image retrieval. In *International Conference on Computer Vision*, pages 675–682, 1998.
- [2] G.R. Bradski. Real time face and object tracking as a component of a perceptual user interface. *Proc.IEEE Workshop on Applications of Comp. Vis.*, pages 214–219, 1998.
- [3] F.A.M Cappabianco. *Segmentação de tecidos do cérebro humano em imagens de ressonância magnética e sua avaliação*. PhD thesis, Instituto de Computação, Universidade Estadual de Campinas, 2009.
- [4] F.A.M. Cappabianco, A.X. Falcão, C.L. Yasuda, and J.K. Udupa. MR-Image Segmentation of Brain Tissues based on Bias Correction and Optimum-Path Forest Clustering. Technical Report IC-10-07, Institute of Computing, University of Campinas, March 2010. In English, 23 pages.
- [5] R.M. Cesar, Jr., E. Bengoetxea, I. Bloch, and P. Larrañaga. Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms. *Pattern Recognition*, 38, November 2005.
- [6] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.
- [7] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):281–288, 2003.
- [8] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, May 2002.
- [9] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5):564–575, 2003.
- [10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, series B*, 39(1):1–38, 1977.

- [11] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.
- [12] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, 1973.
- [13] A.X. Falcão, J. Stolfi, and R.A. Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):19–29, Jan 2004.
- [14] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
- [15] K. Fukunaga and L Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21:32– 40, 1975.
- [16] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact Maximum A Posteriori Estimation for Binary Images. *Journal of the Royal Statistical Society*, 1989.
- [17] Jianping Han and Zhigeng Pan. Robust moving objects detection in dynamic scenes based on mean shift. *Cyberworlds, International Conference on*, pages 271–275, 2008.
- [18] Sonsoles Herrero and Jesús Bescós. Background subtraction techniques: Systematic evaluation and comparative analysis. In *ACIVS*, pages 33–42, 2009.
- [19] Zhi-Kai Huang and De-Hui Liu. Unsupervised image segmentation using em algorithm by histogram. In *Proceedings of the intelligent computing 3rd international conference on Advanced intelligent computing theories and applications*, ICIC'07, pages 1275–1282, Berlin, Heidelberg, 2007. Springer-Verlag.
- [20] Iasonas Kokkinos and Petros Maragos. Synergy between object recognition and image segmentation using the expectation-maximization algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1486–1501, 2009.
- [21] T. S. Korting, L. M. G. Fonseca, and F. L. Ba. Expectation-maximization x self-organizing maps for image classification. *Signal-Image Technologies and Internet-Based System, International IEEE Conference on*, 0:359–365, 2008.
- [22] D.S. Lee. Effective gaussian mixture learning for video background subtraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):827–832, 2005.
- [23] Tangwei Liu, Huiyu Zhou, Faquan Lin, Yusheng Pang, and Ji Wu. Improving image segmentation by gradient vector flow and mean shift. *Pattern Recogn. Lett.*, 29:90–95, January 2008.
- [24] Yi Lu and Hong Guo. Background removal in image indexing and retrieval. In *Proceedings of the 10th International Conference on Image Analysis and Processing*, ICIAP '99, pages 933–, Washington, DC, USA, 1999. IEEE Computer Society.

- [25] J. B. Macqueen. Some methods of classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- [26] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [27] P.A.V. Miranda, A.X. Falcão, and J. K. Udupa. Cloud Models: Their Construction and Employment in Automatic MRI Segmentation of the Brain. Technical Report IC-10-08, Institute of Computing, University of Campinas, March 2010. In English, 18 pages.
- [28] Nobuki Otsu. A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62–66, jan. 1979.
- [29] J.P. Papa. *Classificação Supervisionada de Padrões Utilizando Floresta de Caminhos Ótimos*. PhD thesis, Instituto de Computação, Universidade Estadual de Campinas, 2009.
- [30] Massimo Piccardi. Background subtraction techniques: a review. In *SMC (4)'04*, pages 3099–3104, 2004.
- [31] L.M. Rocha, F.A.M. Cappabianco, and A.X. Falcão. Data clustering as an optimum-path forest problem with applications in image analysis. *International Journal of Imaging Systems and Technology*, 19(2):50–68, June 2009.
- [32] Andres Sanin, Conrad Sanderson, and Brian C. Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. *Pattern Recognition, International Conference on*, 0:141–144, 2010.
- [33] Mehmet Sezgin and Bullent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, 2004.
- [34] Ronen Basri e Achi Brandt Sharon Alpert, Meirav Galun. "image segmentation by probabilistic bottom-up aggregation and cue integration.". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [35] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, Aug 2000.
- [36] T. V. Spina, Javier A. Montoya-Zegarra, A. X. Falcão, and P. A. V. Miranda. Fast interactive segmentation of natural images using the image foresting transform. In *Proceedings of the 16th international conference on Digital Signal Processing, DSP'09*, pages 998–1005, Piscataway, NJ, USA, 2009. IEEE Press.

- [37] T.V. Spina and A.X. Falcão. USIS: user-steered image segmentation framework, URL <http://code.google.com/p/ift-demo/> (data de acesso: março de 2011), 2010.
- [38] W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 37(5):1382–1389, October 2007.
- [39] S. Theodoris and K. Koutroumbas. *Pattern Recognition*. Elsevier Science, 4 edition, 2009.
- [40] C.F.J. Wu. On the convergence properties of the em algorithm. *Annals of Statistics*, 11(1):95–103, 1983.
- [41] R. Avros Z. Volkovich and M. Golani. On Initialization of the Expectation Maximization Clustering Algorithm. *Global conference on power control and optimization (PCO)*, 2(2):117–120, 2011.