

PATRICIA ANTONIOLLI

# SEGMENTAÇÃO AUTOMÁTICA DE PÁSSAROS EM IMAGENS DIGITAIS

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

CURITIBA  
2016

PATRICIA ANTONIOLLI

# SEGMENTAÇÃO AUTOMÁTICA DE PÁSSAROS EM IMAGENS DIGITAIS

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Jacques Facon

CURITIBA  
2016

Antoniolli, Patricia,

SEGMENTAÇÃO AUTOMÁTICA DE PÁSSAROS EM IMAGENS DIGITAIS. CURITIBA, 2016.

Dissertação - Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

1. Segmentação de Imagens 2. JND - Just noticeable difference 3. Agrupamento 4. Cálculo da Dimensão Fractal I. Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e Tecnologia. Programa de Pós-Graduação em Informática II - t



# Sumário

<b>Sumário</b>	i
<b>Lista de Figuras</b>	iv
<b>Lista de Tabelas</b>	vi
<b>Resumo</b>	vii
<b>Abstract</b>	viii
<b>Capítulo 1</b>	
<b>Introdução</b>	1
1.1 Descrição do Problema . . . . .	2
1.2 Objetivos . . . . .	3
1.2.1 Objetivo Geral . . . . .	3
1.2.2 Objetivos Específicos . . . . .	3
1.3 Hipótese do Trabalho . . . . .	4
1.4 Justificativa e Contribuições . . . . .	4
1.5 Organização do Trabalho . . . . .	6
<b>Capítulo 2</b>	
<b>Fundamentação Teórica</b>	7
2.1 Considerações Iniciais . . . . .	7
2.2 Cores e Espaço de Cores . . . . .	7
2.2.1 O Modelo RGB de Cores . . . . .	8
2.2.2 O Espaço de cor HSI - Hue, Saturation, Intensity . . . . .	8
2.2.3 O Espaço de cor Luv . . . . .	10
2.3 Segmentação de Imagens Coloridas . . . . .	10
2.4 Teoria Fractal . . . . .	12
2.4.1 Fractal Aplicado na Segmentação de Imagens . . . . .	14
2.4.1.1 Teoria Fractal de Durer . . . . .	14
2.4.2 Dimensão Fractal - Abordagem 2D Variation Procedure . . . . .	15
2.5 JND - Just Noticeable Difference . . . . .	17

2.6	Agrupamento K-Means . . . . .	18
2.7	Rotulação Binária - Labeling . . . . .	19
2.8	Multilimiarização - Multi Thresholding . . . . .	20
2.9	Morfologia Matemática . . . . .	21
2.9.1	Morfologia Matemática em níveis de cinza . . . . .	22
2.9.1.1	Erosão em níveis de cinza . . . . .	22
2.9.1.2	Dilatação em níveis de cinza . . . . .	23
2.9.1.3	Reconstrução em Níveis de Cinza . . . . .	24
2.9.2	Morfologia Matemática em cor . . . . .	24
2.9.2.1	Escolha do Espaço de Cor . . . . .	24
2.9.2.2	Erosão em Cor . . . . .	27
2.9.2.3	Dilatação em Cor . . . . .	27
2.10	Considerações Finais . . . . .	27

### Capítulo 3

<b>Estado da Arte</b>	28	
3.1	Considerações Iniciais . . . . .	28
3.2	Color Image Segmentation Based on Visual Perception . . . . .	28
3.3	Using Color for Object Recognition . . . . .	29
3.4	Automatic segmentation and indexing in a database of bird images . . . . .	30
3.5	GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts . . . . .	31
3.6	BiCoS: A Bi-level Co-Segmentation Method for Image Classification . . . . .	32
3.7	TriCoS: A Tri-level Co-Segmentation Method for Image Classification . . . . .	33
3.8	Bird Species Classification Based on Color Features . . . . .	33
3.9	Considerações Finais . . . . .	35

### Capítulo 4

<b>Método Proposto, Experimentos e Resultados Alcançados</b>	37	
4.1	Base de Imagens . . . . .	37
4.2	Visão Geral do Método Proposto . . . . .	39
4.3	Métricas para Avaliação de Desempenho . . . . .	42
4.4	Análise dos Resultados . . . . .	46
4.5	Considerações Finais . . . . .	47

### Capítulo 5

<b>Conclusão</b>	49
------------------	----



# Lista de Figuras

Figura 2.1	<i>Espaço de cor RGB (GOMES; VELHO, 1994).</i>	9
Figura 2.2	<i>Espaço de cor HSI (GOMES; VELHO, 1994).</i>	9
Figura 2.3	<i>Elipse do espaço de cor Luv (GOMES; VELHO, 1994)</i>	10
Figura 2.4	<i>Etapas fundamentais do PDI. (GONZALEZ; WOODS, 2000)</i>	11
Figura 2.5	<i>Pentágono.</i>	15
Figura 2.6	<i>Triângulo abc</i>	15
Figura 2.7	<i>Pentágono após 4 subdivisões</i>	15
Figura 2.8	<i>Boxes e/ou Janelas (KRAFT; KAUER, 1995).</i>	16
Figura 2.9	<i>Exemplo de redução de complexidade em imagens utilizando JND (BHOYAR; KAKDE, 2010).</i>	17
Figura 2.10	<i>Elemento estruturante: (a)cruz, (b)quadrado, (c)hexagonal (FACON, 1996).</i>	21
Figura 2.11	<i>Dilatação utilizando a ordenação marginal em cada canal RGB (CALIXTO, 2005).</i>	25
Figura 2.12	<i>Histograma das imagens: (a) imagem original e (b) imagem reconstruída com dilatação (CALIXTO, 2005).</i>	25
Figura 3.1	<i>Imagem Casa: (a) imagem original, (b) imagem quantizada, e (c), (d) e (e) são resultados de segmentação com valores para <math>n=2,3,4</math> respectivamente, (f) resultado da segmentação por peak-finding (CHANG; WANG; HUANG, 2012).</i>	29
Figura 3.2	<i>ROI (DAS; MANMATHA, 2001).</i>	31
Figura 3.3	<i>Imagem original, imagem com a extração do fundo e imagem segmentada (DAS; MANMATHA, 2001).</i>	31
Figura 3.4	<i>Três exemplos de segmentação com o GrabCut (ROTHER; KOLMOGOROV; BLAKE, 2004).</i>	31



Figura 3.5	<i>Imagens de pássaros segmentados usando BiCoS (CHAI; LEMPITSKY; ZISSERMAN, 2011).</i>	32
Figura 3.6	<i>Imagens de pássaros segmentados usando BiCoS (CHAI et al., 2012).</i>	33
Figura 3.7	<i>TabelaTRiCoS (CHAI et al., 2012).</i>	34
Figura 3.8	<i>Segmentação por cores (MARINI; FACON; KOERICH, 2013).</i>	34
Figura 3.9	<i>Visão geral do estado da arte (1/2).</i>	36
Figura 3.10	<i>Visão geral do estado da arte (2/2).</i>	36
Figura 4.1	<i>Exemplo de Bounding Box. (WAH et al., 2011).</i>	38
Figura 4.2	<i>Informações de segmentação de imagens (WAH et al., 2011).</i>	38
Figura 4.3	<i>Normalização de imagens.</i>	39
Figura 4.4	<i>Variações de imagens.</i>	39
Figura 4.5	<i>Exemplo de imagem ground truth após normalização.</i>	39
Figura 4.6	<i>Método Proposto.</i>	40
Figura 4.7	<i>Histograma das imagens processadas no JND.</i>	41
Figura 4.8	<i>Recuperação das cores originais da imagem segmentada.</i>	42
Figura 4.9	<i>Resultados para 4 espécies de pássaros.</i>	43
Figura 4.10	<i>Comparativo com os trabalhos correlatos.</i>	48

# Lista de Tabelas

Tabela 4.1	Matriz de confusão para avaliação da segmentação a partir da relação ground-truth e resultado da segmentação proposta (FAWCETT, 2006).	43
Tabela 4.2	Tabela de índices do desempenho alcançado pelo método proposto. ME (Misclassification Error), RAE (Relative Foreground Area Error), True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Precision (P), Accuracy (A), Recall (R), Error (E), F-measure (FM), Negative Rate Metric (NRM)	46

# Resumo

A identificação visual de espécies de pássaros procura ajudar os cientistas e biólogos no estudo de ecossistemas e coleta de dados sobre a biodiversidade. Atualmente este processo envolve a cooperação de pesquisadores no desenvolvimento de técnicas de segmentação, cuja finalidade é fornecer o reconhecimento automático de pássaros em imagens. O principal objetivo da pesquisa foi o desenvolvimento de um método de segmentação automática de pássaros em imagens digitais coloridas sem restrição de entrada e capturadas em ambientes naturais. Foram empregadas técnicas de redução de cor JND, algoritmo de agrupamento k-means e cálculo de dimensão fractal. Estas técnicas executam a segmentação automática de pássaros mediante a extração de características de cor e textura. Os resultados obtidos mostram a viabilidade da solução desenvolvida que pode chegar a uma média de 90,91% de acurácia na segmentação do conjunto formado por 1.200 imagens derivadas de 21 espécies de pássaros do conjunto de dados UCSD-Caltech Birds CUB-200-2011.

**Palavras-chave:** Segmentação de imagens, JND - just noticeable difference, agrupamento, dimensão fractal.

# Abstract

The visual identification of bird species seeks to assist scientists and biologists in the study of ecosystems and data collect on biodiversity. Currently this process involve the cooperation of researchers in the development of segmentation techniques whose purpose is to provide automatic recognition of birds in digital images. The main purpose of research was the develop a method of automatac segmentation of birds in colored digital images without input restriction and captures in narural environments. JND color reduction techniques were employed, k-means clustering algoritm and fractal dimension calculation. These techniques perform automatic segmentation of birds by extracting color and texture features. The results obtained show the feasibility of the solution developed that could reach an average of 90,91% of accuracy in segmentation of set formed by 1200 images derived from 21 species of birds from UCSD-Calthech Birds CUB-200-2011 dataset.

**Keywords:** Image segmentation, JND-just noticiable diference, clustering, fractal dimension.

# Capítulo 1

## Introdução

Embora os grandes avanços da tecnologia e o aumento da capacidade de processamento de dados e informações tenham ocorrido de forma exponencial, o reconhecimento visual por meio de sistemas computacionais ainda é considerado como um grande desafio para a atualidade.

A literatura tem mostrado que a tarefa de transferir para um computador a capacidade humana de distinguir um objeto com diferenças de escala, iluminação ou cenários, é uma tarefa árdua e complexa. O reconhecimento visual pode ser utilizado para alcançar objetivos, tais como: a identificação, localização, segmentação e a recuperação de objetos de interesse.

A visão computacional é a ciência que se dedica a desenvolver teorias e métodos voltados à extração automática das informações contidas em imagens digitais, que poderão ser utilizadas para a obtenção de conhecimento específico em aplicações computacionais.

Um sistema computacional que realiza a identificação de espécies de pássaros pode ser utilizado para auxiliar na contagem dos indivíduos de uma população que habita uma região e para o monitoramento de indivíduos de determinadas espécies. Esses dados podem influenciar na tomada de decisões por especialistas em relação a espécies ameaçadas de extinção, espécies com super população ou mesmo em desequilíbrio ecológico.

A realização dessa contagem pode ocorrer de forma manual por meio de armadilhas físicas para a captura e anilhamento do pássaro e também em forma de observação e com auxílio de imagens em formato digital capturadas por dispositivos imageadores, como câmeras fotográficas, câmera de vídeo, sistemas embarcados em veículos aéreos não tripulados (drones), entre outros.

Para Yao et al. aplicações computacionais com esse intuito tornam-se muito úteis e de fundamental importância para diversas aplicações reais, automatizando o processo de monitoramento (YAO, 2012).

A preservação da biodiversidade e o estudo do comportamento de pássaros são algumas das motivações que os pesquisadores da área de reconhecimento de padrões, processamento de imagens digitais e especialistas em pássaros possam implementar e empregar os sistemas de visão computacional como ferramentas no desenvolvimento de técnicas para a identificação automática de pássaros. Encontrando assim soluções para a preservação da biodiversidade através do uso de sistemas informatizados no auxílio aos ecologistas (CAI et al., 2007).

Para que seja realizada a identificação do pássaro, a tarefa de segmentação precisa ser executada de forma eficiente e com menor taxa de erro. A segmentação de pássaros a partir de imagens digitais coloridas é o objetivo deste trabalho. A utilização da técnica de redução de cores JND, algoritmo de agrupamento K-Means e Cálculo da Dimensão Fractal foram adotadas para definir um novo método para separação do pássaro em relação às demais informações do contexto total da imagem digital de forma automática. Proporcionando assim a possibilidade de desenvolvimento de sistemas informatizados de monitoramento de pássaros em ecossistemas através da identificação de espécies de pássaros.

Resultados obtidos nesta dissertação são animadores, e apresentam possibilidades para o sucesso na implementação do método proposto, visto que a redução da complexidade das imagens coloridas ocorreu de forma positiva e traz uma perspectiva de aumento nas taxas de acerto para a segmentação de pássaros em imagens digitais e consequentemente um aumento nas taxas de acerto na identificação de pássaros.

O presente Capítulo também apresenta a descrição do problema, os objetivos, a justificativa, bem como a contribuição e a organização deste documento.

## 1.1 Descrição do Problema

A área de pesquisa em processamento de imagens digitais possui grandes desafios, um deles é a capacidade de extração rápida de informações a partir das imagens para a realização de análise e identificação.

Devido à diversidade de uso e tipos distintos de imagens, não existe uma solução única para todas as aplicações que necessitam realizar a segmentação de imagens, o que dá origem a diversas técnicas e métodos (COMANICIU; MEER, 2002).

Para Gonzalez e Woods, um ponto fundamental na extração do objeto de interesse é a simplificação da imagem, reduzindo as informações em regiões homogêneas, permitindo uma representação mais apropriada e de melhor performance no processamento (GONZALEZ; WOODS, 2000).

O desafio da tarefa de segmentação desta dissertação se deve em grande parte à

variação do fundo, iluminação e posição do objeto de interesse, uma vez que a maioria das imagens de pássaros são coletadas no habitat natural dos mesmos. Assim, nestas imagens não é possível controlar rotação, escala e ângulo de visão no momento da aquisição.

Sendo assim as imagens que compõem a base CUB-200-2011 não possuem controles e padrões para a aquisição. Tornando a tarefa de segmentação um grande desafio a ser vencido uma vez que não existe restrição de entrada.

Para Kasten, McKinley, e Gage a revolução tecnológica têm permitido novas técnicas de coleta de imagens sobre o meio ambiente para a realização de uma importante tarefa: o monitoramento de ecossistemas (KASTEN; MCKINLEY; GAGE, 2010).

Recursos como satélites, redes de sensores, técnicas de reconhecimento de padrões e visão computacional, bem como o uso de veículos aéreos não tripulados (drones) são ferramentas úteis para a aquisição de imagens ambientais em larga escala. Para Acevedo et al. a compreensão acerca dos sistemas ecológicos ainda é limitada devido à necessidade de identificação de espécies por seres humanos (ACEVEDO et al., 2009).

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

O objetivo principal desta dissertação é o desenvolvimento de um método automático para a segmentação de pássaros em imagens digitais coloridas que tiveram sua aquisição em ambiente natural, sem restrições ou padrões de coleta. As imagens possuem grande complexidade na variação do fundo, da iluminação e posição do objeto de interesse.

### 1.2.2 Objetivos Específicos

O desenvolvimento do método para a segmentação de imagens de pássaros também incluirá os seguintes objetivos específicos:

- Estudar e avaliar as imagens base CUB-200-2011 disponibilizada pelo Instituto de Tecnologia da Califórnia analisando a compatibilidade e viabilidade para a sua utilização no método proposto;
- Utilizar as imagens segmentadas geradas pela interface Mechanical Turk para a criação de uma base Ground Truth, com o objetivo de validar o método e mensurar os resultados dos testes;
- Aplicar e analisar técnicas de redução de cores nas imagens por meio do estudo

bibliográfico e processamento do algoritmo JND-Just Noticeable Difference ;

- Estudar e aplicar os conceitos de agrupamento (clusterização) bem como o uso do algoritmo K-Means testando configurações mais eficientes para o método proposto visando extrair características de cores das imagens;
- Desenvolver a programação de algoritmos que realize a extração de objetos que tocam a borda da imagem, processe a rotulação e que recupere as cores originais de cada pixel da imagem;
- Pesquisar, aplicar e avaliar as técnicas de segmentação de imagens com a utilização do Cálculo da Dimensão Fractal, para a extração de características de textura da imagem;
- Codificar a abordagem escolhida de segmentação baseado nas técnicas estudadas.
- Realizar os testes e aplicar as métricas nos resultados obtidos pelo algoritmo do método proposto.

### **1.3 Hipótese do Trabalho**

A hipótese deste trabalho é a seguinte:

“É possível segmentar imagens coloridas de pássaros de forma automática ?”

### **1.4 Justificativa e Contribuições**

Nas aplicações de processamento de imagens há a necessidade de separar elementos visuais uns dos outros identificando-os como objetos individuais, com características e comportamentos próprios para que então o reconhecimento de objetos possa ser realizado. A separação ou o isolamento de regiões de interesse em uma imagem acontece durante a fase de segmentação, sendo esta fase responsável pela redução da quantidade de dados que serão analisados trazendo assim maior eficiência.

Segundo Chang, et al. a segmentação da imagem é a primeira etapa que constitui o processo de análise de imagens digitais. Esta é também uma etapa essencial e complexa do processamento de imagens, pois determina a qualidade e o êxito do trabalho de identificação de objetos e o seu reconhecimento aplicado à área de interesse que está inserido (CHANG; WANG; HUANG, 2012).



Pesquisadores de reconhecimento de padrões e de processamento digital de imagens estudam e desenvolvem algoritmos com a finalidade de realizar o reconhecimento automático de pássaros em imagens digitais. Buscando assim soluções para aplicações inovadoras que possam auxiliar cientistas e biólogos no estudo dos ecossistemas e coleta de dados sobre a biodiversidade.

Em muitos casos, é necessário estimar o número de pássaros em áreas ecologicamente sensíveis, como por exemplo, reservas naturais ou em áreas de difícil acesso.

Cientistas utilizam a observação de pássaros para analisar e compreender ecossistemas, por serem numerosos, sensíveis às mudanças ambientais, mais fáceis de controlar do que outras espécies, estão por toda parte e são relativamente fáceis de serem vistos. Segundo artigo publicado em 2012 pela revista *The Scientist*, diversas aplicações do mundo real podem utilizar as aves como coadjuvantes, tais como: monitoramento de poluição ambiental (COSSINS, 2012), avaliação da qualidade do meio ambiente e indicadores de sustentabilidade (BARDELI et al., 2010).

De acordo com Bardeli et al. os pássaros são um indicador muito eficaz de indícios de alterações na biodiversidade, porque são distribuídos em grande diversidade de paisagens, são fáceis de detectar em comparação com outros grupos de animais (BARDELI et al., 2010).

Para Sekercioglu, Daily e Ehrlich, dentre os grupos taxonômicos existentes, as aves possuem uma ampla distribuição mundial, ocorrendo em quase todos os tipos de habitat, oferecendo diversas funções atuando como: polinizadores, predadores, dispersores de sementes, necrófagos e engenheiros dos ecossistemas (SEKERCIOGLU; DAILY; EHRLICH, 2004).

Sistemas de monitoramento permitem a coleta de informações que auxiliam na conservação da avifauna local, indicando para os gestores ambientais quais espécies estão em maior risco de desaparecer, e quais as necessidades ambientais para que se mantenham populações viáveis em unidades de conservação.

Também pode-se citar que a utilização de sistemas de monitoramento de pássaros são de grande ajuda no que toca à segurança da aviação em aeroportos, podendo auxiliar na prevenção de colisões de pássaros e aeronaves evitando assim um possível incidente (SEKERCIOGLU; DAILY; EHRLICH, 2004).

Para Vas et al. o monitoramento de espécies pode ocorrer com a utilização de um sistema embarcado em veículos aéreos não tripulados, comumente chamados de drones, que estão sendo cada vez mais usados em pesquisa ecológica, em particular para abordar fauna sensível em áreas inacessíveis.

A revista inglesa *Biology Letters*, na sua edição de fevereiro de 2015, publicou um

artigo onde avalia o impacto da cor do drone, velocidade e ângulo de voo nas respostas comportamentais de patos selvagens (*Anas platyrhynchos*) em situação de semi-cativeiro, flamingos selvagens (*Phoenicopterus roseus*) em uma área de pantanal (VAS et al., 2015).

A importância do estudo de técnicas de segmentação de pássaros aplicadas a imagens digitais está diretamente ligada ao processo de reconhecimento de pássaros. Como forma de auxiliar no desenvolvimento sistemas automatizados para o monitoramento de espécies, a proposta deste trabalho é desenvolver uma abordagem para a segmentação automática de pássaros contribuindo assim nas pesquisas de especialistas em reconhecimento de padrões e visão computacional.

A adoção de métodos automatizados para identificar espécies de aves é uma forma interessante para avaliar a quantidade e diversidade das aves que aparecem em uma região e podem auxiliar em várias aplicações práticas. Goëau et al. enfatiza que este tipo de tarefa é essencial para um desenvolvimento sustentável e pertinente ao contexto atual (GOËAU et al., 2015).

Desta forma, as possibilidades de aplicações práticas dos sistemas computacionais já mencionadas justificam o estudo de mecanismos para a segmentação automática de pássaros em imagens digitais.

## 1.5 Organização do Trabalho

Este trabalho foi estruturado em cinco capítulos, conforme descrição a seguir: O Capítulo 1 aborda a descrição do problema, os objetivos, a hipótese, justificativa e contribuições. O Capítulo 2 traz a fundamentação teórica, sob o enfoque das principais técnicas relacionadas aos temas centrais deste trabalho. O Capítulo 3 trata o estado da arte com relação aos temas abordados nesta pesquisa por meio de estudos de trabalhos correlatos. O Capítulo 4 apresenta a método proposto para a dissertação de mestrado evidencia os experimentos e resultados alcançados, bem como comparações e análises realizadas. O Capítulo 5 expressa as conclusões, as principais contribuições desta dissertação, bem como sugestões de continuidade e trabalhos futuros.

# Capítulo 2

## Fundamentação Teórica

### 2.1 Considerações Iniciais

Neste capítulo são apresentados os conceitos relacionados na área de reconhecimento de padrões, visão computacional, processamento digital de imagens e segmentação. Bem como o referencial teórico necessário para a fundamentação do método apresentado nesta dissertação.

### 2.2 Cores e Espaço de Cores

No processamento digital de imagens coloridas, a cor é um poderoso descritor que simplifica a identificação e extração de objetos de uma cena. Milhares de cores podem ser identificadas aos olhos humanos, mas apenas duas dezenas de tonalidades de cinza (GONZALEZ; WOODS, 2000).

A cor é o resultado da percepção da luz que incide na retina em células fotorreceptoras, denominadas cones. A maioria das cores visíveis pelo olho humano pode ser representada pela combinação de luzes monocromáticas nos comprimentos de onda do azul, vermelho e verde.

O olho humano percebe cerca de 30 níveis de cinza e 7 milhões de cores. Ele é mais sensível ao verde, depois ao vermelho, e menos ao azul, porém percebe mais variações de azul, depois de vermelho e menos de verde. Uma cor pode ser decomposta em três componentes independentes: intensidade, matiz, e saturação. A intensidade é responsável pela sensação de brilho, a matiz pela sensação de cor, e a saturação pelo grau de pureza da cor em relação ao branco.

Imagens coloridas são armazenadas em três componentes primários formando um

modelo ou espaço de cor (GOMES; VELHO, 1994).

Segundo Gonzales e Woods, essencialmente um modelo de cor é uma especificação de um sistema de coordenadas tridimensionais e um subespaço dentro deste sistema onde cada cor é representada por um único ponto (GONZALEZ; WOODS, 2000).

Segundo Gomes e Velho, para a área de processamento de imagens, os espaços de cores RGB e o HSI podem ser utilizado com grande eficiência.

O modelo de cor RGB (Red, Green, Blue) tem como propósito principal a reprodução de cores em dispositivos eletrônicos como aparelhos de televisão, monitores de computador, projetores, scanners, câmeras digitais, assim como na fotografia tradicional (GOMES; VELHO, 1994).

O modelo HSI (Hue - matiz, Saturation - saturação e Intensity - intensidade ou brilho) é utilizado em sistemas de visão artificial, pois suas características proporcionam o desenvolvimento de algoritmos de processamento de imagens fortemente baseados no modelo de percepção de cor pelo ser humano (GOMES; VELHO, 1994). Existem vários modelos de espaço de cores, como por exemplo: HSI, HSV, YCrCb, Luv, Lav, XYZ entre outros.

### **2.2.1 O Modelo RGB de Cores**

O espaço de cor RGB- Red, Green, Blue, é modelado onde cada cor aparece nos seus componentes espectrais primários, vermelho, verde e azul.

Para Gomes e Velho em modelos RGB existe uma grande correlação entre seus canais. A alteração em apenas um canal reflete significativamente na imagem como um todo.

É um modelo de cor baseado na síntese aditiva, com o qual é possível representar uma cor com a adição das três cores primárias. As componentes do modelo RGB podem variar entre 0 e 255, definindo-se um cubo de cor, onde o valor (0,0,0) corresponde ao preto, e o valor (255,255,255) corresponde ao branco (GOMES; VELHO, 1994). A Figura 2.1 representa o cubo com as representações das cores no espaço RGB.

### **2.2.2 O Espaço de cor HSI - Hue, Saturation, Intensity**

O modelo HSI representa uma cor em torno dos valores de matiz (H - hue), saturação (S - saturation) e intensidade (I - intensity) (GOMES; VELHO, 1994).

A matiz é um atributo que descreve uma cor pura, enquanto que a saturação dá uma medida do grau de diluição de uma cor pura por luz branca e a intensidade ou brilho representa o nível de cinza na cor.

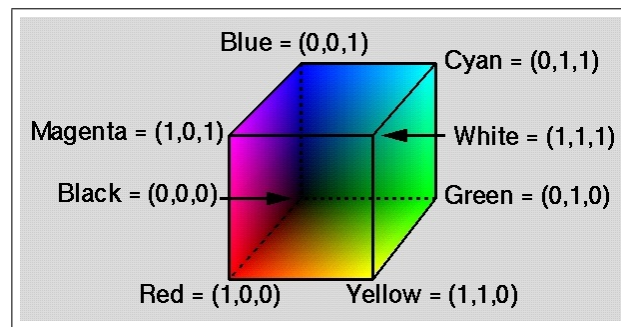


Figura 2.1: *Espaço de cor RGB (GOMES; VELHO, 1994).*

O modelo de cor HSI é muito importante e atrativo para aplicações de processamento de imagens. Apresenta canais mais decorrelacionados, sendo mais apropriado para processar canais separadamente.

Além disso, a cor é representada de uma forma que corresponde à forma da percepção pelo olho humano. A componente  $H$  descreve a cor propriamente dita na forma de um ângulo entre 0 e  $360^\circ$ .

A componente  $S$  sinaliza o quanto a cor está "poluída" com a cor branca, apresentando valores entre 0 e 1. A componente  $I$  define a caracterização da luz acromática presente na imagem, variando entre 0 (branco) e 1 (preto). O modelo HSI distribui as cores em um cone conforme a Figura 2.2:

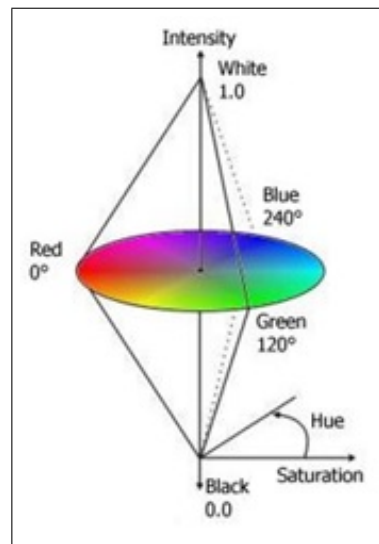


Figura 2.2: *Espaço de cor HSI (GOMES; VELHO, 1994).*

### 2.2.3 O Espaço de cor Luv

Segundo Gomes e Velho, um sistema de cor é perceptualmente uniforme se a métrica perceptual JND- Just Noticeable Difference, é a métrica euclidiana do sólido de cor. Diversos padrões de cor foram estabelecidos com o objetivo de se obter sistemas de cor perceptualmente uniformes. O sistema CIE-Luv é um destes padrões. Esse sistema utiliza uma decomposição crominância/luminância do espaço de cor. No entanto a luminância é substituída pela luminosidade (lightness). A luminosidade é uma medida de como a intensidade luminosa é percebida pelo olho humano (GOMES; VELHO, 1994). A Figura 2.3 ilustra a elipse do espaço de cor LUV. O sistema Luv faz uma transformação do sistema CIE-XYZ, de modo a obter uniformidade perceptual no plano de cromaticidade. Apesar de possuir a característica da uniformidade o sistema CIE-Luv requer grande recurso computacional para realizar os cálculos para a transição entre ele e o sistemas padrão CIE-RGB e CIE-XYZ (GOMES; VELHO, 1994).

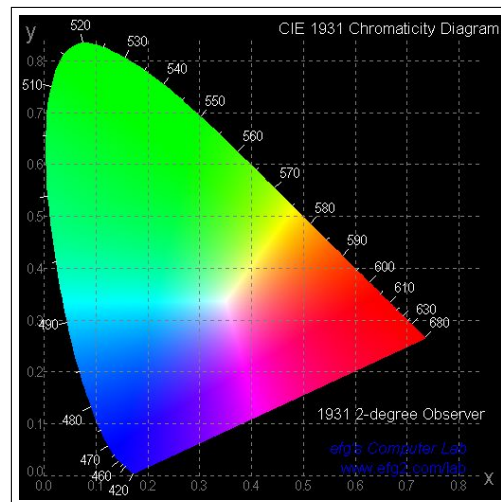


Figura 2.3: *Elipse do espaço de cor Luv (GOMES; VELHO, 1994)*

## 2.3 Segmentação de Imagens Coloridas

Para Gonzalez e Woods o processamento de imagens digitais aplicado ao reconhecimento e interpretação de dados está associado com aplicações de análise de imagens nas quais o objetivo é a extração automática, ou semi automática, de informação a partir de uma imagem, a Figura 2.4 ilustra as etapas do processamento digital de imagens.

Para isso uma questão fundamental é a redução das informações dessa imagem em

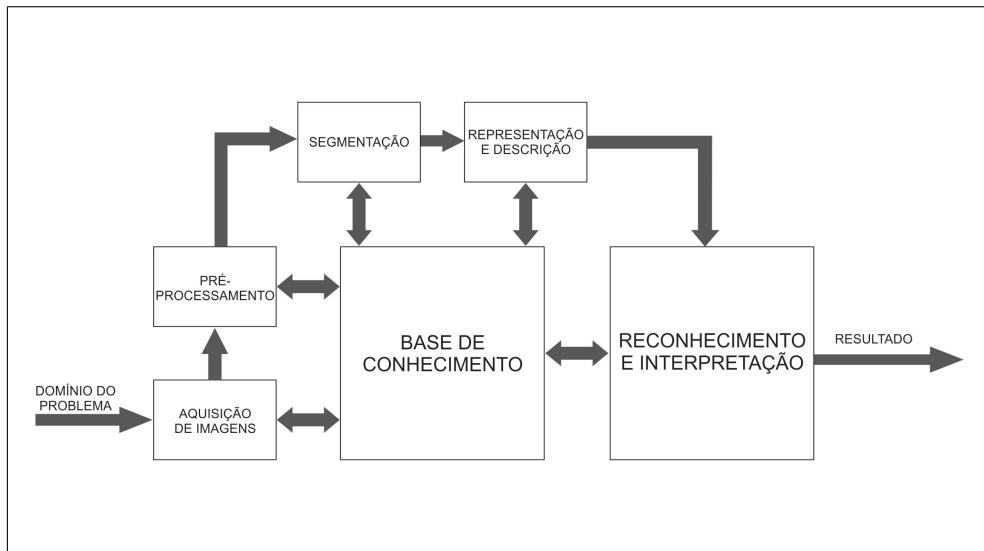


Figura 2.4: *Etapas fundamentais do PDI. (GONZALEZ; WOODS, 2000)*

regiões mais ou menos homogêneas, permitindo uma análise mais especializada da imagem.

Esse é o problema da segmentação, e têm dado origem ao desenvolvimento de um grande número de pesquisas. Sendo a segmentação uma etapa fundamental para o processamento de imagens digitais, e tendo a tarefa de dividir uma imagem em suas unidades significativas, ou seja, nos objetos de interesse. Esses objetos possuem duas características básicas: eles exibem alguma uniformidade interna em relação a uma propriedade da imagem, e eles contrastam em relação a sua vizinhança (GONZALEZ; WOODS, 2000).

A segmentação consiste em separar os pixels da imagem que fazem parte do objeto de interesse, dos pixels que representam o cenário ou o fundo em que o objeto encontra-se inserido (CHAI et al., 2012). No campo da visão computacional, a segmentação refere-se ao processo de decomposição de uma imagem digital em vários segmentos (regiões) que a formam. Esta tarefa, apesar de simples de descrever, é das mais difíceis de se implementar (FILHO; NETO, 1999). Pode-se dizer que a elaboração de algoritmos visando à automatização de processos de segmentação é uma das tarefas mais árduas na área de processamento de imagens (GONZALEZ; WOODS, 2000). O desenvolvimento dos algoritmos leva em consideração duas características básicas dos pixels que envolvem dois conceitos: descontinuidade e similaridade. Por meio do uso desses conceitos, conforme estruturado por Gonzalez e Woods, a segmentação é feita usando-se detecção de bordas e detecção de similaridades. Por sua vez, os métodos mais comuns empregados na segmentação podem ser divididos em três grupos principais:

- Métodos de Detecção de Descontinuidades: abrangem técnicas de detecção de pon-

tos, linhas e bordas;

- Métodos de Detecção de Similaridades: possuem técnicas de crescimento de regiões, detecção de pirâmides e de bacias;
- Métodos de Segmentação no Espaço de Atributos: faz a utilização de técnicas de limiarização, agrupamento (*clustering*), e do algoritmo k-means.

Cabe lembrar, que a qualidade de segmentação de uma imagem influencia diretamente no resultado da fase de análise e, portanto a escolha de uma técnica de segmentação adequada implica no conhecimento dos detalhes das técnicas existentes (GONZALEZ; WOODS, 2000).

## 2.4 Teoria Fractal

Pesquisas com aplicações baseadas na teoria do caos tiveram um aumento expressivo nas três últimas décadas. Segundo Gleick enquanto algumas teorias desafiavam os sistemas dinâmicos Newtonianos, o caos tem questionado a opinião tradicional da estrutura Newtoniana.

Onde começa o caos a ciência clássica pára. O crescimento do caos, sua fascinante combinação de ordem e desordem e a sua constante instabilidade estão relacionados com o rápido desenvolvimento de computadores com maior poder de processamento e armazenamento (GLEICK; DUTRA, 1998).

O uso do computador no desenvolvimento de simulações realistas e de técnicas de vídeo possibilitou aos cientistas verificar a grande evolução dos sistemas dinâmicos e os complexos efeitos caóticos das equações diferenciais básicas. O Caos ganhou interesse especial em varias áreas científicas, como a biologia, física, química, matemática, astronomia, economia, geografia e também a arte e design (LI; LIU; LO, 1997).

Gleick descreve que o matemático suíço Leonard Euler publicou um artigo em 1755 intitulado “Princípios gerais do movimento dos fluidos” , no qual mostrou um conjunto parcial de equações diferenciais para descrever o movimento de fluídos não viscosos.

Desde então houve grande evolução nas pesquisas sobre a teoria do caos, como a do engenheiro francês Claude Navier em 1821 que leva em consideração a viscosidade; em 1845 o físico matemático britânico George Stokes fez diferentes hipóteses das mesmas equações. Em 1920, o meteorologista Lew Fry Richardson ( que influenciou a historia do fractal através da medição da linha costeira) calculou soluções de uma equação especificando as condições iniciais e limítrofes, e usando métodos numéricos (GLEICK; DUTRA, 1998).



Após a segunda guerra John von Neumann notou que os métodos analíticos existentes dos matemáticos estavam inadequados para problemas não lineares. O problema de não linearidade é de notória dificuldade. Na dinâmica dos fluidos, por exemplo, o fluxo estável dos fluidos é perfeitamente compreendido. Mas os movimentos turbulentos, como o rápido movimento da correnteza das águas, são um problema de grande complexidade. O objetivo é a relação da turbulência na dinâmica dos fluidos com o caos (LI; LIU; LO, 1997).

A principal característica do caos é que um simples sistema determinístico pode gerar o que aparece ser um comportamento randômico. Caos pode ser observado em sistemas mecânicos básicos.

O pêndulo, no qual o eixo movimenta para cima e para baixo. Para muitas frequências do eixo o balanço do pêndulo é normal, mas em determinado ponto o movimento do pêndulo torna-se errático e imprevisível. Neste ponto o movimento periódico regular do pêndulo, se torna um movimento caótico (GLEICK; DUTRA, 1998).

Uma marca essencial do caos em sistemas não lineares é que eles são extremamente sensíveis às condições iniciais do sistema. Isto quer dizer que dois conjuntos de condições de um sistema que são inicialmente muito próximos, ao longo do tempo podem dar resultados extremamente diferentes.

De acordo com a lei de Newton, bolas soltas da mesma posição com a mesma velocidade inicial devem seguir o mesmo caminho para baixo. Na prática é improvável que as bolas serão soltas exatamente na mesma posição e com a mesma velocidade inicial. Quando as bolas batem nos pinos uma diferença mínima é ampliada e as bolas tomam caminhos diferentes.

Para Gleick a sensibilidade de sistemas não lineares a condições iniciais é popularmente conhecida como o efeito borboleta, por que um simples bater das asas da borboleta teoricamente alteraria as condições iniciais de um sistema meteorológico e poderia assim aumentar drasticamente padrões meteorológicos ao longo do tempo.

Os sistemas dinâmicos clássicos podem ser pensados como atratores que controlam uma porção do espaço. No caso do pêndulo, todas as trajetórias são atraídas para um simples ponto, o atrator, que é a posição quando ele está parado. Para o sistema descrito pelas equações de Lorenz existe um atrator, de qualquer maneira ele não é simplesmente um ponto atrator. A maneira que se faz a trajetória para o atrator neste caso é mostrado pela projeção bi-dimensional. Se as trajetórias das equações de Lorenz são plotadas no espaço de fase tri-dimensional, Lorenz vê que elas são atraídas em direção a uma delimitada elipsóide, a qual todas as trajetórias entram e saem na qual elas nunca emergem.

Uma propriedade matemática curiosa desta elipsóide é que possui volume zero. É de conhecimento que o atrator de Lorenz é uma estrutura de camadas aninhada infinitamente, e que o volume zero é consistente com o cruzamento de seções através das camadas sendo fractal nas estruturas (GLEICK; DUTRA, 1998).

Segundo Li et al. fractal não tem dimensão inteira como os mais comuns subconjuntos no espaço Euclidiano, mas tem dimensão fracionária. Também, quanto mais de perto você vê os fractais mais detalhes você pode observar e eles têm a propriedade básica da auto similaridade.

A teoria fractal com suas múltiplas aplicações tem demonstrado consistência nas aplicações de análise de imagens, demonstrando sua eficiência e robustez (LI; LIU; LO, 1997).

### 2.4.1 Fractal Aplicado na Segmentação de Imagens

O mundo fractal foi vislumbrado por Benoit Mandelbrot nos anos 70, mas os objetos agora considerados como sendo fractais existem desde muito tempo antes.

Segundo Mandelbrot, muitos dos objetos fractais ocorrem naturalmente como árvores, linha da costa ou nuvens, são agora considerados com propriedades fractais, e muito do interesse neste item decorre em simular os fenômenos naturais usando computação gráfica. Outra forma mais abstrata dos objetos fractais foi desenvolvida por artistas e matemáticos, e novamente as técnicas de computação gráfica tem dado novos discernimentos nas estruturas destes objetos (MANDELBROT, 1983).

A definição de Mandelbrot de fractal de um conjunto  $X$  é que, “ a dimensão Hausdorff  $h(X)$  não seja um inteiro”.

A propriedade essencial é a da auto-similaridade. Sub-conjuntos de um objeto fractal tem a mesma forma do objeto todo. Objetos fractais teóricos são infinitesimalmente sub divisíveis deste modo, cada sub-conjunto, embora menor que o conjunto do qual desce, no entanto contem os menos detalhes do que o conjunto completo (MANDELBROT, 1983).

#### 2.4.1.1 Teoria Fractal de Durer

Segundo Caliman um dos primeiros artistas que gerou objetos fractais baseados em pentágonos regulares foi Albrecht Durer( 1471-1528) (CALIMAN; IVANOVICI; RICHARD, 2011).

Tomando um pentágono regular de tamanho  $s$  e ao redor de cada lado outros cinco iguais, a forma criada ajusta quase exatamente a um pentágono grande de lado  $S$ .

A Figura 2.5 agora se assemelha a um pentágono grande com cinco incisões na forma de triângulos isósceles com centro em cada lado.

Os ângulos iguais destes triângulos são de  $72^\circ$  ( o ângulo externo de um pentágono regular é  $360^\circ / 5$  ), o terceiro ângulo sendo metade deste é  $36^\circ$  . A proporção dos lados deste triangulo é conhecida desde os dias de Pitágoras (MANDELBROT, 1983). Tudo o que se precisa saber é qual o menor lado do triangulo pode ser calculado como  $2s \cos 72^\circ$ . Isto quer dizer que a proporção dos tamanhos dos lados dos dois pentágonos é:

$$\frac{s}{S} = \frac{1}{2 + 2\cos 72} \quad (2.1)$$

Agora se começar o procedimento com o pentágono grande. Sabendo-se do meio de calcular  $s$  a partir de  $S$  pode-se desenhar seis pentágonos mais pequenos os quais se ajustam dentro do grande.

A subdivisão pode continuar com pentágonos cada vez menores sendo produzidos. Cada subpentágono é uma copia do todo. Se a subdivisão continuar *adinfinitum*, o pentágono de Durer forma realmente um objeto fractal (MANDELBROT, 1983).

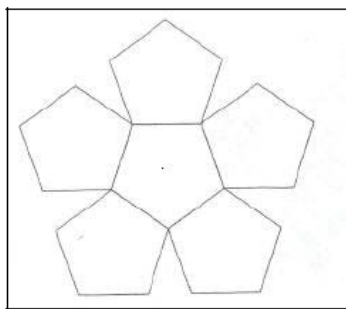


Figura 2.5: *Pentágono.*

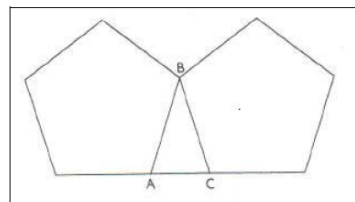


Figura 2.6: *Triângulo abc*

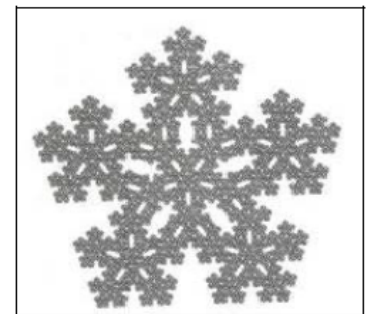


Figura 2.7: *Pentágono após 4 subdivisões*

#### 2.4.2 Dimensão Fractal - Abordagem 2D Variation Procedure

Segundo Zhu, Gao e Yu muitos métodos têm sido propostos para o cálculo da dimensão fractal de imagens em níveis de cinza. Um dos mais utilizados é o *Box-counting*, o qual baseia-se na cobertura da superfície cuja dimensão se deseja calcular por *boxes* ou também denominadas de *janelas* -  $r$  de tamanhos fixos determinados conforme a aplicação (ZHU; GAO; YU, 2010).

A abordagem *2D Variation Procedure* desenvolvida por Kraft e Kauer consiste na análise da vizinhança de um pixel para diferentes tamanhos de janelas ( $r$ ) centradas nele. As janelas utilizadas são quadradas de tamanhos diversos. A Figura 2.8 ilustra o

exemplo de *boxes* de tamanho 9,7 e 5.

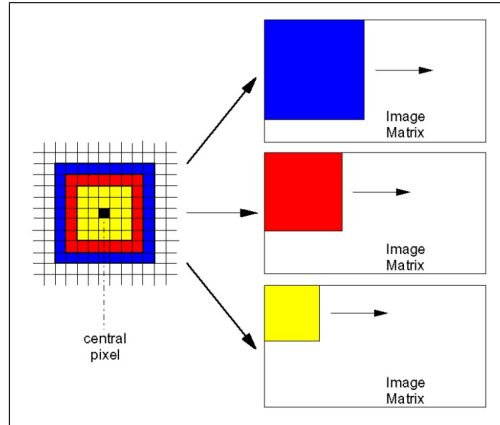


Figura 2.8: *Boxes e/ou Janelas* (KRAFT; KAUER, 1995).

Essa abordagem pode ser aplicada em imagens binárias e também em imagens em níveis de cinza (KRAFT; KAUER, 1995). Para estas, o algoritmo percorre pixel a pixel e procura os valores mínimo e máximo de níveis de cinza no *box* de tamanho  $r$  e os atribui ao pixel central. Definindo assim uma função bidimensional relacionando os valores máximo e mínimo para cada tamanho de *box* e assim a diferença volumétrica  $V(r)$ , para a imagem inteira, entre esses valores máximo e mínimo. No modelo de representação de Richardson-Mandelbrot, o volume  $V(r)$  mostra-se linearmente dependente do tamanho  $r$  do *box*, assim para calcular a Dimensão Fractal segue-se os seguintes passos:

Para cada pixel o tamanho de  $r$ : 3, 5, 7, 9 e 11.

O valor de  $V(r)$  é a diferença entre os valores máximo e mínimo de níveis de cinza de cada pixel.

$$V(r) = \max - \min \quad (2.2)$$

Para calcular a estimativa da inclinação  $s$ , utiliza-se o Método dos Mínimos Quadrados:

$$a = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \quad (2.3)$$

Onde

$$y_i = ax + b$$

e

$$x_i = \log r$$

e

$$y_i = \log V(r)$$

para  $r = 3, 5, 7, 9, 11$ . Assim a Estimativa da inclinação  $s = a$

Obtêm-se a dimensão fractal  $D2d$  avaliando a inclinação  $s$  da reta no gráfico de representação de Richardson-Mandelbrot:

$$D2d = 3 - s/2 \quad (2.4)$$

## 2.5 JND - Just Noticeable Difference

A sensibilidade visual para pequenas diferenças entre duas cores é de fundamental importância nos experimentos perceptuais com cor (GOMES; VELHO, 1994). Os estudos de Mac Adam na década de 40 mostraram que existe um conjunto formado por todas as cores que possuem a mesma diferença perceptual entre si que é uma elipse. Além disso a excentricidade e os comprimentos dos eixos dessa elipse variam para cores em diferentes regiões do diagrama de cromaticidade. Esse fato mostra que a distância perceptual no espaço de cor não pode ser medida usando a métrica euclidiana.

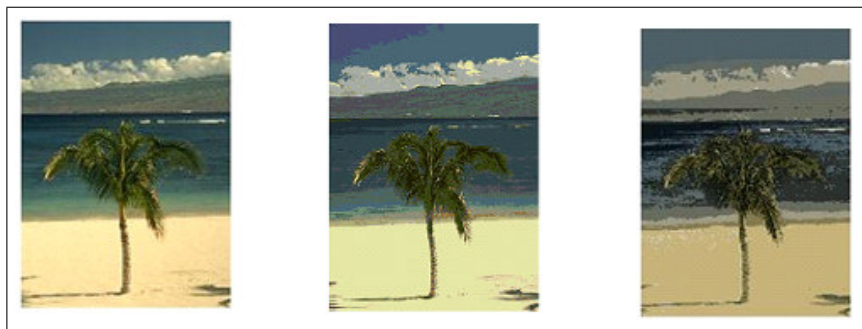


Figura 2.9: *Exemplo de redução de complexidade em imagens utilizando JND (BHOYAR; KAKDE, 2010).*

Segundo Gomes e Velho a métrica perceptual é conhecida na literatura como métrica JND (Just Noticeable Color Difference) (GOMES; VELHO, 1994). Nas imagens coloridas os atributos das cores podem ser representadas em histogramas.

Segundo os pesquisadores Bhoyar e Kakde os histogramas são fáceis de calcular, e não variam de acordo com a rotação e translação (espelhamento) do conteúdo da imagem. O potencial de uso dos histogramas para imagens coloridas com o objetivo de indexação da imagem é discutido por esses pesquisadores. No entanto histogramas de imagens coloridas,

possuem várias limitações inerentes para a realização tarefa de indexação e restauração de imagens.

Primeiramente, no histograma colorido convencional ( CCH ) duas cores serão consideradas totalmente diferente se elas se enquadram em duas caixas diferentes embora possam ser muito semelhantes entre si para a percepção humana. Isto é, CCH não considera a semelhança entre as cores em diferentes bin, nem a desigualdade de cor na mesma bin. Por isso, o CCH é sensível a interferências de ruídos, tais como alterações de iluminação e de erros de quantização.

Em segunda análise, a alta dimensão do CCH (ou seja, o número de barras do histograma ) requer grande recurso computacional em comparação a outros histogramas, a Figura 2.9 ilustra os resultados do estudo . Finalmente, histogramas de imagens coloridas não incluem informações espaciais e, portanto, não é adequado para suportar a indexação e recuperação de imagens , com base no local de conteúdo da imagem (BHOYAR; KAKDE, 2010).

## 2.6 Agrupamento K-Means

As técnicas de segmentação precisam encontrar e individualizar as regiões na imagem que tenham características similares. Sendo assim, padrões pertencentes a um determinado grupo devem ser mais similares entre si do que em relação a outros grupos.

Técnicas de agrupamento - *clustering*, encontradas na literatura de reconhecimento de padrões possuem objetivos similares e, sendo assim, de acordo com Jain, podem ser aplicadas para a segmentação de imagens (JAIN, 1989).

Nesta abordagem de segmentação, um conjunto de dados é substituído por grupos ou clusters. A definição de clusters parte da idéia que valores de atributos  $x$  são vistos como pontos no espaço  $l - dim$ , sendo assim, clusters são regiões contínuas deste espaço e que contém uma alta densidade de pontos, separadas de outras regiões densas, por regiões com baixa densidade de ponto (TAN et al., 2009).

Partindo do exposto, duas abordagens para fundamentação dos algoritmos podem ser encontradas. A primeira leva em consideração o conceito de agrupamento rígido (hard, crisp) no qual cada ponto pertence a um único grupo e a segunda considera o conceito de agrupamento nebuloso (fuzzy) onde cada ponto pertence a todos os grupos até certo grau.

O Segundo Tan et al., o k-means é um algoritmo não supervisionado usado em reconhecimento de padrões e segmentação de imagens, que gera clusters de  $n$  objetos baseados em atributos a partir de  $k$  partições, sendo  $k < n$ . Presumindo que os atributos

dos objetos formam um espaço vetorial, e tem por fundamento principal minimizar a variância intra-agrupamento total, ou seja, a função dada por:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (2.5)$$

onde existem  $k$  clusters  $S_i, i = 1, 2, \dots, k$ , e  $\mu_i$  é o centróide ou ponto médio de todos os pontos  $x_j \in S_i$ .

O k-means é iniciado pela partição dos pontos de entrada em  $k$  conjuntos iniciais, de forma randômica ou heurística. Calcula-se então o ponto médio, ou centróide, de cada agrupamento.

Constrói-se uma nova partição pela associação de cada ponto ao centróide mais próximo. Os centróides são re-calculados para os novos agrupamentos, e o algoritmo é repetido pela aplicação alternada desses dois passos até que a convergência seja obtida quando os pontos não mudarem mais de agrupamentos ou alternativamente se os centróides não mudarem de posição.

Mas também pode ser adotado como critério de parada para o algoritmo um número máximo de iterações a ser definido pelo usuário (TAN et al., 2009).

## 2.7 Rotulação Binária - Labeling

A rotulação binária - *Labeling* consiste em marcar cada região conexa com um rótulo através de características iguais (cor, brilho, forma, textura). Esta técnica busca um pixel com uma determinada característica, e lhe atribuir um rótulo na forma de um número. Em seguida utilizando-se dessa característica é possível localizar pixels vizinhos com a mesma característica e atribuir o mesmo valor de rótulo aos demais pixels da região (FILHO; NETO, 1999).

Antes de rotular uma imagem é preciso binarizá-la. Com a binarização da imagem, obtêm-se uma imagem composta de pixels pretos (valor 0) e de pixels brancos (valor 255).

Cada região branca ou preta em uma imagem é chamada de região conexa.

Segundo Gonzales e Woods o algoritmo de rotulação consiste em localizar numa imagem binarizada um pixel de posição  $(x, y)$  que contenha uma determinada característica. Por se tratar de uma imagem binarizada, ele localiza o pixel de cor preta (valor 0). Esse pixel de posição  $(x, y)$  recebe um valor numérico  $L$  (label) que representará o rótulo da região ao qual pertence. Após ser rotulado o pixel é armazenado em uma lista que está inicialmente vazia e que tem as seguintes operações:

- Inserção  $(s, t)$ : insere o pixel de posição  $(s, t)$  no fim da lista.

- Remoção  $(s, t)$ : remove o pixel da frente da lista e o salva para uso futuro como sendo o pixel  $(s, t)$ .

Depois de adicionado o pixel de posição  $(x, y)$  na lista, é verificado se a lista não está vazia. Caso a lista não esteja vazia, o algoritmo verificará os 4-vizinhos do pixel, ou seja, os pixels que contenham a mesma característica que o pixel rotulado, ou seja, os pixels de cor preta.

Para cada pixel de mesma semelhança é adicionado na lista a sua posição e em seguida retirado o pixel de posição  $(x, y)$  menos recentemente colocado. Se as características dos vizinhos forem iguais ao primeiro pixel rotulado, os vizinhos recebem o mesmo valor  $L$  e são colocados na lista. O mesmo ocorre com os vizinhos dos vizinhos que receberão o mesmo rotulo  $L$ .

Enquanto a lista estiver cheia, os vizinhos do pixel que estiverem na lista serão verificados. Quando a lista estiver vazia, seja uma região inteira foi rotulada, recomeça a varredura de cima para baixo, da esquerda para direita procurando por outro pixel sem rotulo. Se for encontrado outro pixel sem rotulo, recomeça o processo de rotulação (GONZALEZ; WOODS, 2000).

## 2.8 Multilimiarização - Multi Thresholding

Para Gonzalez e Woods, a multilimiarização é uma das abordagens mais importantes para a segmentação. A idéia que está por trás dessa técnica é a que um objeto pode ser entendido como uma região formada por pixels contíguos que tenham em comum uma faixa de intensidades. Dessa forma, a multilimiarização usa a intensidade dos pixels para distinguir as regiões.

Essa técnica é usada quando se tem vários objetos que se distinguem do fundo. A presença desses faz com que o histograma de distribuição de níveis de cinza seja multimodal. O limiar é encontrado pela localização dos vales que separam os objetos (GONZALEZ; WOODS, 2000).

Segundo Papamarkos e Gatos, a multilimiarização pode ser então entendida como sendo uma extensão dos métodos de segmentação, com a diferença que possibilita a segmentação em múltiplas classes.

Uma técnica de multilimiarização deve ser capaz de segmentar a imagem digitalizada para os diferentes objetos que apresentam propriedades similares (PAPAMARKOS; GATOS, 1994).



## 2.9 Morfologia Matemática

De acordo com Dougherty a morfologia matemática explora características geométricas das imagens, é uma teoria usada para análise, filtragem e segmentação da imagem.

Criada no meio de 1960 por Georges Matheron e Jean Serra da École de Mines de Paris, com o objetivo de quantificação de estruturas de acordo com sua geometria. Sua evolução teórica ao longo dos anos tem gerado uma ferramenta poderosa não só para lidar com a geometria das estruturas, mas também a análise e nos procedimentos de modelagem e simulação (DOUGHERTY, 1992).

O princípio básico da morfologia matemática, que pode ser binária, em níveis de cinza ou em cor, consiste em extrair informações relativas à geometria e à topologia do conteúdo de uma imagem.

Para Facon a morfologia matemática quantifica a intuição do pesquisador, analisando as estruturas geométricas da imagem a partir de um conjunto perfeitamente definido e conhecido (forma e tamanho) chamado de: "Elemento Estruturante", conforme ilustrado na Figura 2.10. O resultado dessa transformação permite gerar novas imagens e avaliar o conteúdo das mesmas.

A morfologia matemática disponibiliza uma grande gama de operadores de filtragem, segmentação, detecção de bordas, esqueletização, afinamento, análise de formas, compressão, etc (FACON, 1996). Foi originalmente desenvolvida para imagens binárias em meados dos anos 60, foi estendida para imagens em níveis de cinza. E as pesquisas atuais se concentram em imagens coloridas.

Os pilares da morfologia matemática são as duas operações básicas, a erosão e a dilatação, a partir das quais, por composição, é possível realizar muitos outros operadores poderosos (FACON, 1996). A morfologia matemática pode ser aplicada em imagens binárias, em níveis de cinza e também em imagens coloridas.

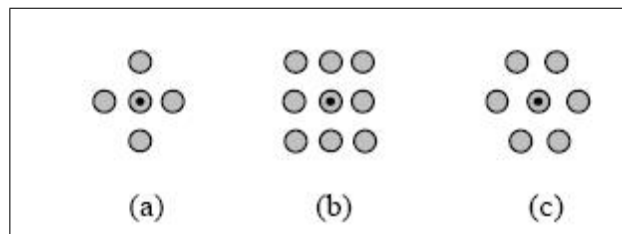


Figura 2.10: *Elemento estruturante: (a)cruz, (b)quadrado, (c)hexagonal (FACON, 1996).*

A morfologia matemática pode ser aplicada em imagens binárias, em níveis de

cinza e também em imagens coloridas.

### 2.9.1 Morfologia Matemática em níveis de cinza

A morfologia matemática possui uma vantagem que é a facilidade de implementação e o aproveitamento das implementações existentes, destacando-a das outras técnicas de processamento de imagens (NAJMAN; TALBOT, 2013).

Outro motivo para utilizar filtros morfológicos é a capacidade de preservar o formato original dos objetos da imagem, qualquer que seja seus tamanhos, enquanto que filtros lineares tem como desvantagens deteriorar a nitidez dos objetos. Uma terceira vantagem da morfologia matemática aplicada ao processamento de imagens é seu aspecto intuitivo, uma vez que trabalha diretamente sobre o domínio espacial.

Os elementos estruturantes são considerados como "tijolos fundamentais" para desempenhar o mesmo papel das frequências no caso de processamento de sinais (FACON, 1996).

#### 2.9.1.1 Erosão em níveis de cinza

Segundo Facon erosão em níveis de cinza pode ser definida através de um sinal  $f$  por um elemento estruturante  $g$  e pode ser descrita pela Equação 2.6:

$$\varepsilon^g(f(x)) = \text{Min}\{f(x) - g(x - y) : y \in E\} \quad (2.6)$$

onde:  $\text{Min}$  representa o mínimo entre vários valores e  $E$  representa o elemento estruturante.

Para o cálculo da erosão os seguintes passos são necessários:

1. Posicionar o elemento estruturante sobre o primeiro pixel da imagem.
2. Calcular a diferença de cada par correspondente de valores de pixels do elemento estruturante e da imagem.
3. Encontrar o valor mínimo de todas as diferenças e armazenar o pixel correspondente na imagem de saída para este valor.
4. Repetir este processo para cada pixel da imagem a ser erodida.

A erosão é uma transformação não comutativa apresentando várias propriedades interessantes como:

- Diminuir as partículas;
- Eliminar conjuntos de tonalidade clara de tamanho inferior ao tamanho do elemento estruturante;
- Aumentar os buracos;
- Separar conjuntos próximos;
- Ressalta as partes escuras da imagem, tornando-as mais espessas (FACON, 1996).

### 2.9.1.2 Dilatação em níveis de cinza

De acordo com Facon, a dilatação em níveis de cinza pode ser definida através de um sinal  $f$  por um elemento estruturante  $g$  e pode ser descrita pela Equação 2.7:

$$\varepsilon^g(f(x)) = \text{Max}\{f(y) + g(x - y) : y \in E\} \quad (2.7)$$

onde:  $\text{Max}$  representa o máximo entre vários valores e  $E$  representa o elemento estruturante (FACON, 1996).

Para o cálculo da dilatação os seguintes passos são necessários:

1. Posicionar o elemento estruturante sobre o primeiro pixel da imagem.
2. Calcular a soma de cada par correspondente de valores de pixels do elemento estruturante e da imagem.
3. Encontrar o valor máximo de todas essas somas e armazenar o pixel correspondente na imagem de saída para este valor.
4. Repetir este processo para cada pixel da imagem a ser dilatada.

Segundo Facon, a dilatação é uma transformação comutativa, cujos efeitos são:

- engordar partículas;
- eliminar conjuntos de tonalidade escura de tamanho inferior ao tamanho do elemento estruturante;
- preencher pequenos buracos;
- conectar conjuntos próximos (FACON, 1996).

### 2.9.1.3 Reconstrução em Níveis de Cinza

A reconstrução é uma técnica que permite recuperar componentes importantes da imagem, como por exemplo: informações de uma imagem danificada ou informações de uma imagem desgastada pelo tempo, ou tentar reconstruir uma imagem que sofreu modificações no decorrer do tratamento (FACON, 1996).

A reconstrução trabalha com duas imagens: a imagem marcadora (á reconstruir) e a imagem máscara (modelo), realiza-se uma dilatação condicional na imagem marcadora e faz-se uma comparação dos pixels da imagem máscara, conforme a Equação 2.8:

$$\forall P(x, y) \in D_s, \rho_s(Z)(P(x, y)) = \text{Max} \{k \in [0, n] : P(x, y) \in \rho_{-}\{L\}_{-}\{k\}_{-}\{(s)\}(L_{-}\{k\}Z)\} \quad (2.8)$$

onde:  $Z$  é a imagem marcadora e  $X$  é a imagem máscara.

## 2.9.2 Morfologia Matemática em cor

A morfologia matemática em cor representa atualmente uma linha de pesquisa nova e desafiadora da área de Processamento de Imagens Digitais (DOUGHERTY; LOTUFO, 2003). Os desafios matemáticos e teóricos residem na escolha apropriada do espaço de cor (modelo) e na definição da ordenação das cores.

### 2.9.2.1 Escolha do Espaço de Cor

Na morfologia em níveis de cinza acima apresentada, os operadores baseiam-se nos conceitos de mínimo e máximo de níveis de cinza. Conceitos que não apresentam nenhuma dificuldade teórica e/ou matemática. Por exemplo, procurar o mínimo ou o máximo entre dois níveis de cinza, 100 e 110 não apresenta nenhuma dificuldade computacional.

Mas como definir o mínimo ou o máximo entre, por exemplo, a cor verde e a cor azul? Essa pergunta representa o desafio matemático e teórico que os pesquisadores tentam solucionar (FACON, 1996).

Em geral os espaços de cores são divididos em três canais de cores, como o espaço RGB, mais conhecido computacionalmente e na indústria. Poderia parecer natural empregar esse espaço de cor para estender a morfologia matemática para cores.

Pesquisas recentes relatam da dificuldade de utilizar o espaço de cor RGB. O espaço de cor RGB apresenta canais R, G e B dependentes, o que deixa instável e ineficiente o uso desses canais em imagens coloridas complexas.

Os estudos de Calixto (CALIXTO, 2005), demonstram o problema da utilização da

morfologia matemática em cor com imagens RGB. Ele ilustra um exemplo da aplicação da dilatação em cada um dos canais R,G e B, onde estes são recombinados após a dilatação para gerar a imagem dilatada RGB. A Figura 2.11 exemplifica as distorções geradas pela aplicação direta da dilatação nos canais R, G e B.

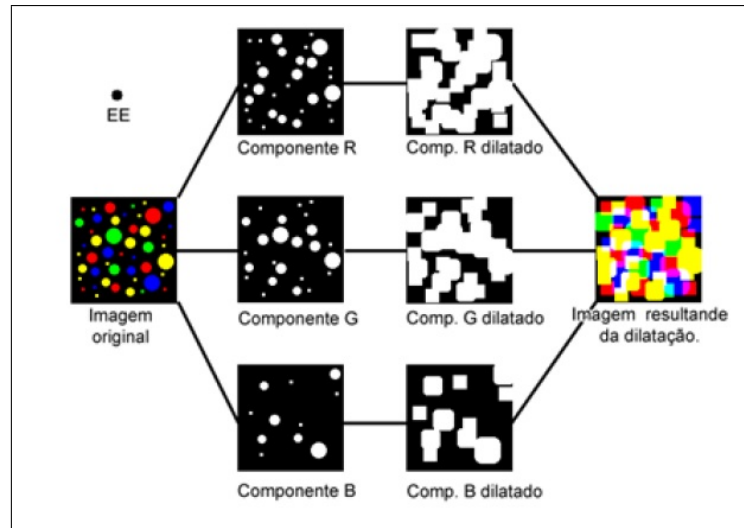


Figura 2.11: *Dilatação utilizando a ordenação marginal em cada canal RGB (CALIXTO, 2005).*

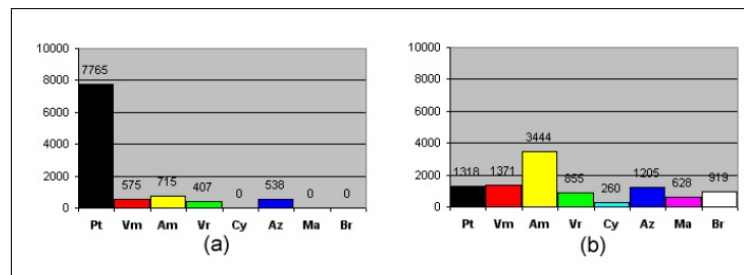


Figura 2.12: *Histograma das imagens: (a) imagem original e (b) imagem reconstruída com dilatação (CALIXTO, 2005).*

Nas Figuras 2.11 e 2.12, é possível perceber que aplicar diretamente a dilatação nos canais R, G e B gerou cores novas que não existiam na imagem original, como o branco. A geração de cores indesejáveis é um dos principais problemas do uso do espaço de cor RGB.

Existem uma série de maneiras diferentes para ordenar dados multivariados. Segundo Calixto algumas ordenações são propostas tais como: Ordenação Marginal; Ordenação Reduzida; Ordenação Parcial; Ordenação Condicional; Ordenação Lexicográfica;

Atualmente publicações científicas mostram que os aspectos da ordenação lexicográfica vem sendo mais comumente utilizado (CALIXTO, 2005).

Ordenação lexicográfica é uma técnica de ordenação vetorial que refere-se a ordem na qual são organizadas as palavras nos dicionários. Então, primeiro a ordem é decidida por um componente (letra), seguido de um segundo componente e depois de um terceiro componente. O algoritmo de ordenação lexicográfica requer uma ordenação interna em cada componente e outra ordenação entre os componentes (ORTIZ et al., 2002).

Esta ordenação é definida como a ordem que os canais de cores são organizados para processamento. Dado um espaço de cor representado por três canais, uma ordenação lexicográfica consiste basicamente em definir uma ordem para o processamento dos canais de cada pixel. Os três canais do pixel nunca serão recombinados/separados, uma vez que nesta ordenação não aparecem cores falsas, pois a escolha de um pixel é com os seus três canais juntos (ORTIZ et al., 2002).

De acordo com Talbot, a morfologia matemática em cor apresenta muitos desafios, o principal sendo a escolha do espaço de cor e da ordenação, pois quando a ordenação lexicográfica é utilizada, o aparecimento de falsas cores após o processamento é evitado (NAJMAN; TALBOT, 2013).

A ordenação lexicográfica é a definida como a ordem que os canais de cores são organizados para processamento. Dado um espaço de cor  $c$  representado por três canais  $C \subset \mathbb{R}^3$  uma ordenação lexicográfica consiste basicamente em definir uma ordem para o processamento de cada pixel. Seja um pixel  $P(x_1, y_1, z_1)$  de coordenadas  $(x_1, y_1, z_1)$  e um pixel vizinho  $P_{vizinho}(x_2, y_2, z_2)$  de coordenadas  $(x_2, y_2, z_2)$ .

A seguir a equação apresenta a ordenação geral proposta por Tobar, Platero, González e Asensio, no caso da busca do mínimo entre duas cores entre  $P(x_1, y_1, z_1)$  e  $P_{vizinho}(x_2, y_2, z_2)$  (TOBAR; P.; G, 2007):

$$\vee\{P(x_1, y_1, z_1) < P_{vizinho}(x_2, y_2, z_2)\} = \begin{cases} x_1 < x_2 \\ ou \\ y_1 < y_2 \text{ e } x_1 = x_2 \\ ou \\ z_1 < z_2 \text{ e } y_1 = y_2 \text{ e } x_1 = x_2 \end{cases} \quad (2.9)$$

De acordo com Calixto são propostos vários modos diferentes de ordenação de dados multivariáveis, desde então não há nenhum modo geralmente aceito para definir a ordenação em espaços de cor.

Uma alternativa, talvez mais natural, de modo a abordar o problema da morfologia

em cor é tratar a cor de cada pixel como um vetor. Esta técnica morfológica é chamada de vetor morfológico (CALIXTO, 2005).

### 2.9.2.2 Erosão em Cor

A operação de erosão colorida é definida como a seleção do valor da cor da imagem com menor distância absoluta (diferença de cor), ou seja, a erosão de uma imagem  $f$  por um elemento estruturante  $K$  é definida conforme a Equação 2.10.

$$\varepsilon_{vk}(f)(x, y) = \text{Min}_{olex} f(x + s, y + t) \text{ onde } (s, t) \in K \quad (2.10)$$

### 2.9.2.3 Dilatação em Cor

A operação de dilatação colorida é definida como a seleção do valor da cor da imagem com maior distância absoluta, ou seja, a dilatação de uma imagem  $f$  por um elemento estruturante  $K$  é definida pela Equação 4.1.

$$\delta_{vk}(f)(x, y) = \text{Max}_{olex} f(x - s, y - t) \text{ onde } (s, t) \in K \quad (2.11)$$

Uma vez definida as duas operações básicas da Morfologia Matemática, torna-se mais fácil a definição de outras operações morfológicas.

## 2.10 Considerações Finais

Este capítulo apresentou os fundamentos teóricos necessários para o entendimento sobre os principais elementos para a construção de um sistema de segmentação automático de pássaros. Foram estudadas as definições necessárias para estabelecer o referencial teórico que fundamenta este trabalho. No próximo Capítulo são descritas as principais práticas adotadas pelas metodologias atuais para a segmentação de pássaros em imagens digitais coloridas.

# Capítulo 3

## Estado da Arte

### 3.1 Considerações Iniciais

A pesquisa sobre trabalhos já publicados e sua posterior avaliação, contribuem para o entendimento do problema apresentado e a descoberta de possíveis estratégias de solução. Seu estudo contribui para a obtenção de novos olhares para a pesquisa. Esse é o objetivo do presente capítulo onde serão descritos trabalhos científicos relacionados aos diferentes métodos de segmentação de imagens digitais, avaliação do método e sua eficiência, métricas e resultados obtidos, bem como a análise de bases de imagens digitais e bases ground truth utilizadas.

### 3.2 Color Image Segmentation Based on Visual Perception

Os pesquisadores Chang, Wang e Huang (CHANG; WANG; HUANG, 2012) apresentam dois enfoques para a redução de cores: (I) Na extração de cores é utilizado um histograma e partir dele são obtidas a frequência e a intensidade das cores, representado por  $n$ . A escolha de  $n$  depende do tamanho da imagem. Deste modo, pode-se obter  $n$  aglomerações de picos que representam a intensidades no espaço de cor RGB (Red, Green, Blue); (II) Frequência e a intensidade das cores obtidas pelo histograma, representado por  $m$  é controlada por um limiar  $T$ , um valor inteiro maior do que 1.

Caso o algoritmo encontre intensidades  $x$  em vermelho,  $y$  em verde e  $z$  em azul satisfazem os critérios do limiar, então um conjunto de centros  $xyz$  é construído.

Além dos histogramas JND o algoritmo Fuzzy C-means também é utilizado para



definir agrupamentos de cores.

Existem dois métodos proposto na literatura: *peak-finding* e *vpc* que são comparados para diferentes valores de  $n$  e números de clusters. Os resultados demonstram que os mesmos valores nos histogramas são obtidos nas imagens originais e nas quantizadas e o tempo de processamento da segmentação foi reduzido. Entretanto, é necessário enfatizar que os experimentos do artigo foram realizados a partir de três imagens diferentes. A Figura 3.1 ilustra os experimentos.

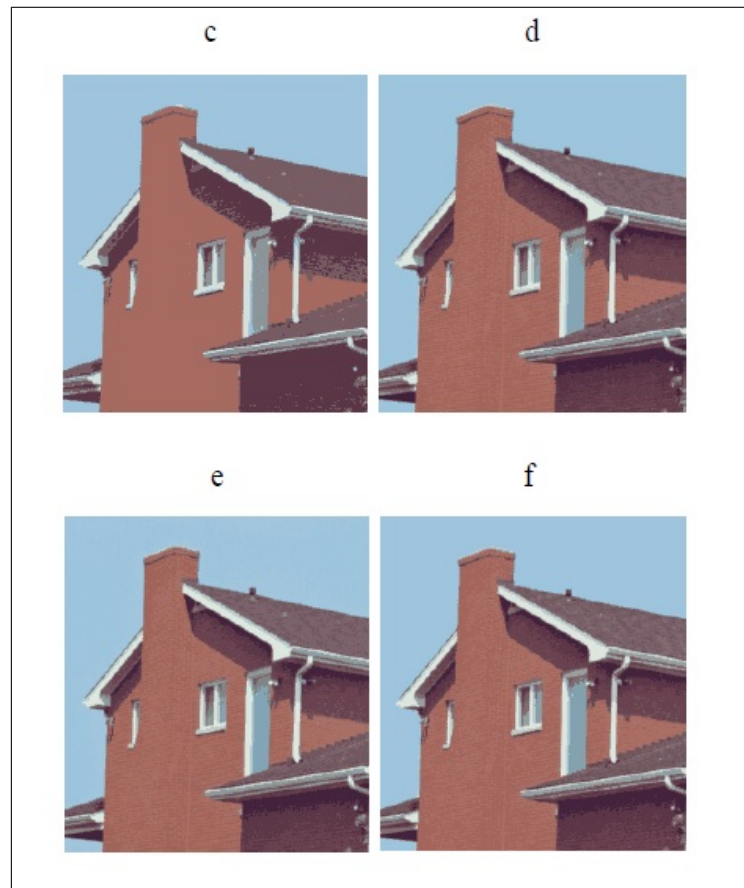


Figura 3.1: *Imagem Casa: (a) imagem original, (b) imagem quantizada, e (c), (d) e (e) são resultados de segmentação com valores para  $n=2,3,4$  respectivamente, (f) resultado da segmentação por peak-finding (CHANG; WANG; HUANG, 2012).*

### 3.3 Using Color for Object Recognition

O trabalho de Moghimi (MOGHIMI, 2011), investiga o uso das características das cores para o reconhecimento de pássaros.

São criados histogramas das cores de diferentes regiões de interesse da imagem para diferentes espaços de cores. A região de interesse ou ROI (Region of Interest) é definida como uma área de onde são extraídas as informações referentes as cores da imagem (a partir do histograma).

A representação em um espaço de cor facilita a especificação de cores de acordo com um determinado padrão.

Para realização dos experimentos foram utilizados os seguintes espaços de cores: RGB, HSV, YIQ, YCrCb, XYZ, LAB e LUV. A base de imagens utilizada nos experimentos foi a Caltech UCSD Birds 200, porém o artigo não cita a quantidade exata de imagens testadas. O autor enfatiza que a segmentação automática não conseguiu bons resultados (YIQ = 12,2%; YCbCr = 12,2%; RGB = 7,9%).

Os resultados para segmentação automática é semelhante aos obtidos quando aplicada a segmentação por particionamento da imagem.

Dependendo da quantidade informação que a imagem possui, a segmentação pode ser completa, pode ser em partes localizadas (demarcadas), ou apenas um ponto aleatório sobre o objeto a ser segmentado.

Outro fator que pode influenciar no desempenho do classificador é a escolha dos espaços de cores utilizados na representação das imagens a serem segmentadas. Todos os resultados são inferiores a 18,9% obtidos através do espaço de cor YIQ. A base de imagens que foi utilizada é a Caltech-UCSD Birds.

### **3.4 Automatic segmentation and indexing in a database of bird images**

O artigo desenvolvido por Das e Manmatha (DAS; MANMATHA, 2001), mostra uma estratégia para segmentar imagens de pássaros usando cores extraídas de parte da imagem a borda, ao invés de utilizar as cores que pertencem a imagem como um todo, conforme mostrado na Figura 3.2.

Este trabalho descreve uma segmentação automática da imagem. O espaço de cor utilizado foi o RGB.

A segmentação das regiões de interesse ROI (Region of Interest) eliminam o fundo a partir das cores da borda da imagem e deixam o apenas o pássaro em evidência, o resultado encontrado pelo autor utilizando uma base de 450 imagens foi de 87% de sucesso na segmentação do da imagens, conforme ilustrado na Figura 3.3.

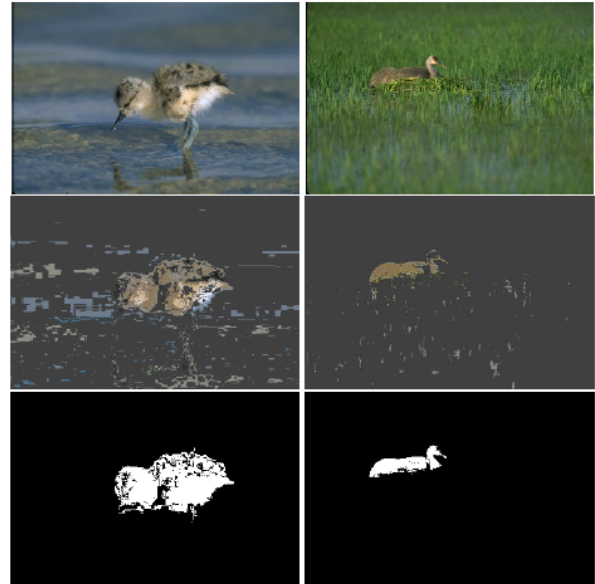
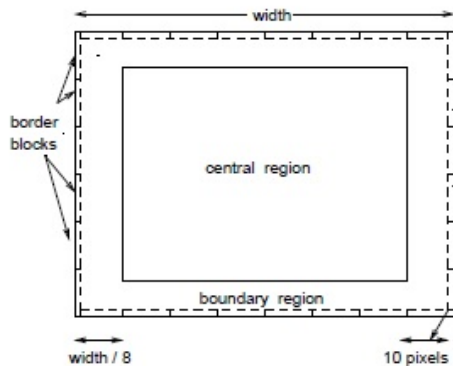


Figura 3.2: ROI (DAS; MANMATHA, 2001). (DAS; MANMATHA, 2001).

Figura 3.3: Imagem original, imagem com a extração do fundo e imagem segmentada

### 3.5 GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts

O trabalho de Rother, Kolmogorov e Blake (ROTHER; KOLMOGOROV; BLAKE, 2004), possui enfoque na segmentação interativa e exata de um objeto.



Figura 3.4: Três exemplos de segmentação com o GrabCut (ROTHER; KOLMOGOROV; BLAKE, 2004).

O método apresentado usou uma base de 50 imagens e quando comparado com outros propostos na literatura e apresenta resultados motivadores. O objetivo é atingir um alto desempenho para o a solução problema sem muito esforço por parte do usuário.

No método proposto pelo artigo, o usuário marca o objeto alvo dentro de um retângulo, que será considerado como marcação de fundo, porém pode ser necessário realizar mais de uma marcação para imagens mais complexas.

O espaço de cor utilizados nas imagens é o RGB. A Figura 3.4 ilustra imagens ori-

ginais com variações complexas de cor e em seguida apresenta a segmentação automática do objeto presente na imagem a partir da execução do algoritmo GrabCut para separação do objeto e o fundo da imagem.

### 3.6 BiCoS: A Bi-level Co-Segmentation Method for Image Classification

O artigo de Chai, Lempitsky e Zissermani apresenta dois algoritmos de segmentação de imagens: 1. Co-segmentação em dois níveis, BiCoS; 2. Evolução do primeiro, com a vantagem de ser multi-tarefa, BiCoS-MT (CHAI; LEMPITSKY; ZISSERMAN, 2011).

Ambos realizam segmentação em dois níveis em diferentes tipos de bases de dados. Foi utilizada uma base com 20 imagens para cada classe: pássaros, flores, cavalos, aviões, faces e motocicletas.



Figura 3.5: *Imagens de pássaros segmentados usando BiCoS (CHAI; LEMPITSKY; ZISSERMAN, 2011).*

Os resultados obtidos foram os seguintes: Pássaros (Caltech UCSD Birds 200) variação de 6,7% até 23,3%; Flores (Oxford Flowers) variação de 64,7% até 80%; Cavalos (Weizmann Horses) variação de 71% até 90%; Aviões, variação de 83,1% até 93,2%; faces, variação de 80% até 92,4%; motocicletas, variação de 71,6% até 90,3%.

A pesquisa de Chai et al. (2011) também traz a comparação dos resultados da classificação sem nenhuma segmentação, com a aplicação de alguns métodos apontados na literatura e também utiliza uma base ground truth. A Figura 3.5 apresenta os resultados obtidos pelos algoritmos apresentados pelo artigo.

Observa-se que os resultados globais do trabalho são bem atrativos, apesar dos resultados alcançados pelo classificador possuir um baixo desempenho com a base de imagens de pássaros Caltech UCSD Birds 200 em comparação com as demais bases que

o artigo contempla.

### 3.7 TriCoS: A Tri-level Co-Segmentation Method for Image Classification

Outro estudo de Chai et al denominado de TriCos (algoritmo de co-segmentação discriminativo de classes em três níveis para classificação de imagens) obteve uma taxa de classificação varia de 16,1%, com aplicações de diferentes versões do algoritmo BiCos ou aplicação do algoritmo proposto o TriCos que apresenta 25,5%.

Além disso, são apresentados resultados 26,7% obtidos com a utilização do ground truth e 22,7% utilizando apenas o recorte da caixa delimitadora (bounding box) (CHAI et al., 2012). As Figuras 3.6 e 3.7 trazem especificidades do estudo de Chai et al. com as imagens da base CUB.



Figura 3.6: *Imagens de pássaros segmentados usando BiCoS (CHAI et al., 2012).*

### 3.8 Bird Species Classification Based on Color Features

Os autores (MARINI; FACON; KOERICH, 2013) descrevem no artigo uma abordagem para identificação automática de pássaros em imagens digitais, onde as características visuais das imagens não são controladas.

O método é estruturado em três etapas: segmentação do pássaro no seu ambiente natural sem restrição de entrada, extração de características e classificação do pássaro

Coseg	Class-indep. seg.	Final classifier	Acc.	GT used?
TriCoS (this work)	ours	ours	<b>25.5</b>	no
BiCoS [6]	ours	ours	23.7	no
BiCoS-MT	ours	ours	24.2	no
-	GrabCut	ours	19.2	no
-	no segmentation	ours	14.1	no
BiCoS	Chai [6]	Chai	15.7	no
BiCoS-MT	Chai	Chai	16.1	no
<i>GT</i>	ours	ours	<b>26.7</b>	seg/trn
<i>GT</i>	ours	ours	22.7	bb/trn
-	-	Welinder et al. [25]	19.0	seg/trn+tst
-	-	Yao et al. CVPR'11 [26]	19.2	bb/trn+tst
-	-	Khan et al. NIPS'11 [14]	22.4	unsure

Figura 3.7: Tabela *TRiCoS* (CHAI et al., 2012).

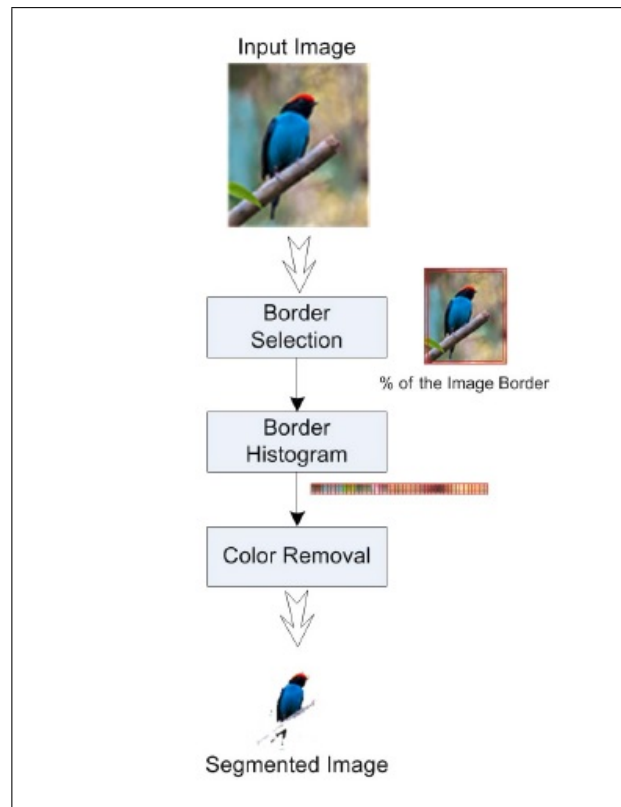


Figura 3.8: Segmentação por cores (MARINI; FACON; KOERICH, 2013).

conforme sua espécie.

Os experimentos foram realizados com 20 imagens de cada classe extraídas da base Caltech UCSD Birds 200 e mostraram que o espaço de cor RGB conseguiu segmentar 71% do conjunto de dados Caltech CUB 200 e o espaço de cor HSV 75%.

A estratégia de segmentação desta pesquisa é uma adaptação do trabalho de Das e Manmatha (DAS; MANMATHA, 2001), onde são consideradas exclusivamente as cores presentes na imagem para este processo.

Parte-se da premissa que considera o fato de que a maioria das imagens mostram os pássaros centralizados e afastados da borda. A extração prévia de informações do fundo colorido da imagem faz-se por um levantamento das cores presentes na borda da imagem.

Definido empiricamente um percentual de pixels de cada borda (superior, inferior, direita e esquerda), os pixels das bordas fornecem uma estatística prévia do fundo colorido.

A análise continua com os outros pixels encontrados fora da borda.

Caso esses pixels apresentem valores de cores pertencentes às faixas extraídas da borda, eles são considerados como fundo e são eliminados. Caso contrário, eles são considerados como pixels pertencentes ao pássaros.

Sendo assim possível isolar as cores do pássaro presentes na imagem e separá-lo do fundo, a Figura 3.8 apresenta a estratégia de segmentação.

### **3.9 Considerações Finais**

Os trabalhos selecionados para compor este capítulo, trazem uma grande contribuição para a atualização do estudo de trabalhos já realizados na área e auxiliam na compreensão sobre a complexidade do problema de que existe na segmentação de imagens de pássaros em abordagens interativas e também nas abordagens que trazem de alguma forma a segmentação automática.

Os resultados que os artigos analisados apresentam uma taxa de acerto na segmentação de imagens coloridas que podem ter o seu desempenho melhorado. Gerando motivação para a realização de pesquisas mediante a necessidade de melhoramento e aperfeiçoamento da segmentação, abrindo oportunidades para novos estudos na área da visão computacional.

As Figuras 3.9 e 3.10 apresentam esquema resumido dos trabalhos escolhidos e que serão utilizados para a comparação dos resultados obtidos com o método proposto.

Título	Autor(es)	Espaços de Cores	Método	Base	Resultados
Color Image Segmentation Based on Visual Perception	Chang, Wang e Huang, 2012	RGB	JND Fuzzy C-Means	03 imagens	Até 64%
Using Color for Object Recognition	Moghimi, 2011	RGB, HSV, YIQ, YCrCb, XYZ, LAB e LUV.	Segmentação por Região de Interesse (ROI)	04 imagens Caltech-UCSD Birds	YIQ = 12,2\%; YCbCr = 12,2\%; RGB = 7,9\%
Automatic segmentation and indexing in a database of bird images	Das e Manmatha, 2001	RGB	Segmentação por cores (borda)	450 imagens	87%
GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts	Rother, Kolmogorov e Blake, 2004	RGB	Alpha Graph Cut	50 imagens	13,6%

Figura 3.9: *Visão geral do estado da arte (1/2).*

Título	Autor(es)	Espaços de Cores	Método	Base	Resultados
BiCoS: A Bi-level Co-Segmentation Method for Image Classification	Chai, Lempitsky e Zissermani, 2011	RGB	GrabCut + Co-Segmentação em dois níveis BiCos-MT	20 imagens de cada classe Caltech-UCSD Birds	6,7 % até 23,3%
Bird Species Classification Based on Color Features	Marini, Facon e Koerich, 2013	RGB e HSV	Segmentação por Cores	20 imagens de cada classe Caltech-UCSD Birds	RGB = 71% HSV = 75%
Interactive Image Segmentation by Maximal Similarity Based Region Merging	Ning, Zhang e Chengke, 2010	HSI	Region Merging Mean Shift	8 imagens	87,49%
TriCoS: A Tri-level Class-Discriminative Co-Segmentation Method for Image Classification	Chai, Lempitsky e Zissermani, 2012	RGB	TriCoS	15 imagens de cada classe Caltech-UCSD Birds	26,7%

Figura 3.10: *Visão geral do estado da arte (2/2).*



## Capítulo 4

# Método Proposto, Experimentos e Resultados Alcançados

### 4.1 Base de Imagens

Nos últimos anos, tornaram-se disponíveis uma grande quantidade de conjuntos de dados, de imagens, dedicadas a assuntos específicos, tais como: imagens de plantas (flores), animais (aves, cachorros), objetos domésticos, veículos entre outros.

Para implementar um sistema de classificação, identificação e/ou segmentação automático viável, os conjuntos de imagens disponíveis, em especial imagens já rotuladas, são componentes críticos para sistemas de visão computacional.

Para que se obtenha uma validação científica, são necessárias algumas centenas ou milhares de imagens de alta resolução já catalogadas e rotuladas e com diversas informações adicionais implementadas por especialistas acerca do objeto de estudo.

De acordo com o exposto acima, a Caltech-UCSD Birds CUB 200-2011 é uma base de imagens de pássaros associadas a diversas informações complementares desenvolvida por Wah et al. no Instituto de Tecnologia da Califórnia (WAH et al., 2011). O conjunto contém 11.788 imagens de 200 espécies de pássaros.

Nesta base, cada espécie possui informações organizadas pela classificação científica (classe, ordem, família, gênero, espécie). Cada classe de pássaro possui em média 30 imagens. As classes com maior número de exemplos possuem 39 imagens e as com menor possuem 20 imagens.

A base de dados possui o *GroundTruth* - segmentação ideal, para cada imagem pertencente a base, criadas com a utilização da interface Mechanical Turk (WELINDER et al., 2010) conforme ilustra a Figura 4.1: dada uma imagem de um pássaro a interface Mechanical Turk mostra a separação ideal do fundo da imagem e do pássaro.

Possui também informações de *BoundingBox* definidas a partir da interface Mechanical Turk (WELINDER et al., 2010), como apresentado na Figura 4.2 onde as imagens de pássaros possuem uma caixa delimitadora retangular que é atribuída à imagem visando isolar informações de interesse e reduzir as informações de fundo da imagem.

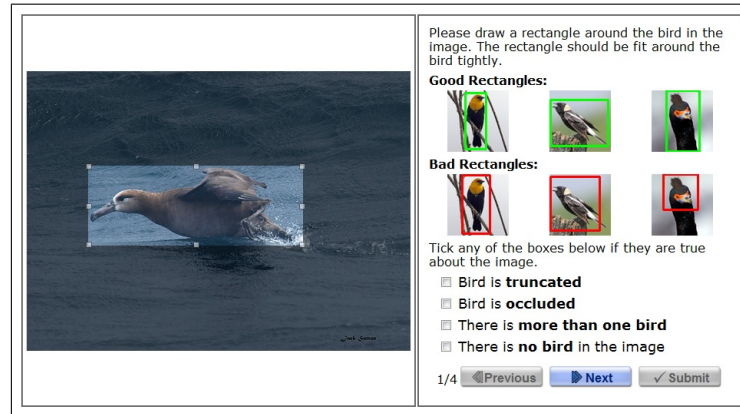


Figura 4.1: *Exemplo de Bounding Box.* (WAH et al., 2011).

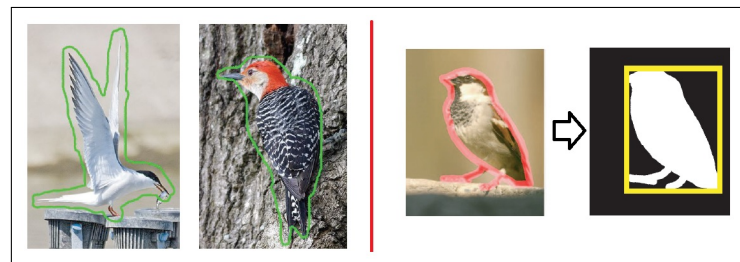


Figura 4.2: *Informações de segmentação de imagens* (WAH et al., 2011).

O desafio para tarefa de segmentação desta dissertação se deve em grande parte à variação do fundo, iluminação e posição do objeto de interesse, uma vez que a maioria das imagens de pássaros são coletadas no habitat natural dos mesmos. Assim, nestas imagens não é possível controlar rotação, escala e ângulo de visão no momento da aquisição, conforme ilustrado na Figura 4.4.

As imagens que foram utilizadas de ground truth, foram extraídas da base Caltech-UCSD Birds CUB 200-2011 e apresentam uma segmentação não perfeita, então foram binarizadas com Otsu no FePi. Conforme pode-se observar na Figura 4.3 a imagem (A) foi extraída da base e passou por um processo de normalização como mostra a imagem (B). A

geração de uma base ground truth no contexto desta dissertação, significa gerar a solução ideal para todas as imagens utilizadas para teste e avaliação do método desenvolvido.



Figura 4.3: *Normalização de imagens.*



Figura 4.4: *Variações de imagens.*

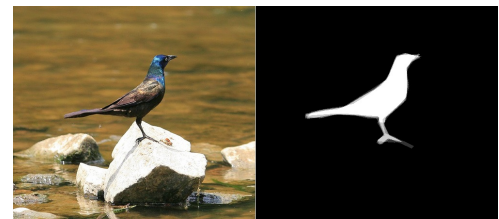


Figura 4.5: *Exemplo de imagem ground truth após normalização.*

## 4.2 Visão Geral do Método Proposto

O método proposto foi desenvolvido para realizar a segmentação automática de pássaros, isto é, sem que haja, em etapa alguma do processo, a interação do usuário com o intuito de efetuar marcações na imagem, para que então a segmentação seja realizada.

Conforme apresentado na Figura 4.6, o método foi particionado em 5 etapas de maneira a demonstrar seu comportamento.

Para a realização dos testes de validação do método proposto, foram selecionadas 1200 imagens de 21 espécies de pássaros que compõem a base Caltech-UCSD Birds CUB 200-2011, formando assim um conjunto de aproximadamente 55 imagens para cada espécie de pássaros selecionadas aleatoriamente. Não houve critério para a escolha destas imagens, porém observou-se que na maioria delas o pássaro não toca a borda da imagem, estando

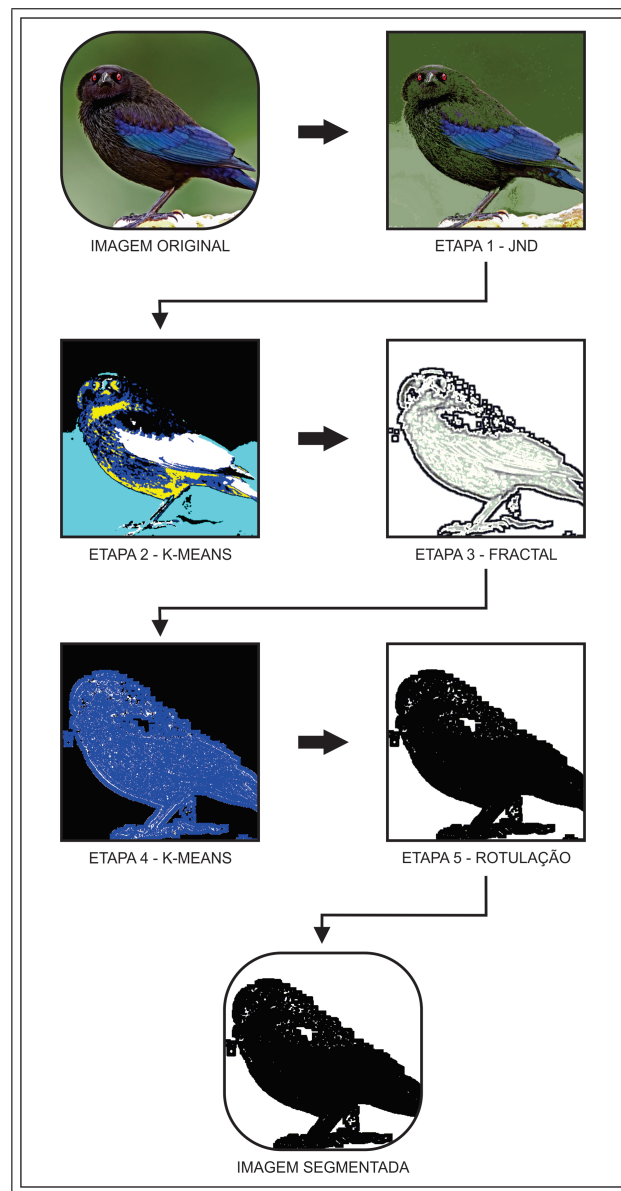


Figura 4.6: *Método Proposto.*

centralizado, pois o mesmo é o objeto principal das imagens que pertencem a base CUB 200-2011.

Cabe salientar que não foi utilizada a caixa delimitadora, *BoundingBox*, nem outro tipo de interação e/ou marcação, nas 1200 imagens processadas neste estudo.

Todas as imagens utilizam o espaço de cor RGB e foram convertidas de *.jpg* para *.bmp* preservando a qualidade das imagens.

Durante a 1<sup>a</sup> etapa as imagens foram processados pelo algoritmo JND que objetiva a redução de cores na imagem e sua complexidade, o que ocorreu de forma eficaz na diminuição da quantidade de informações visuais. Um exemplo da redução está ilustrado

na Figura 4.7, onde é possível comparar os histogramas das imagens antes e após a realização da etapa 1.

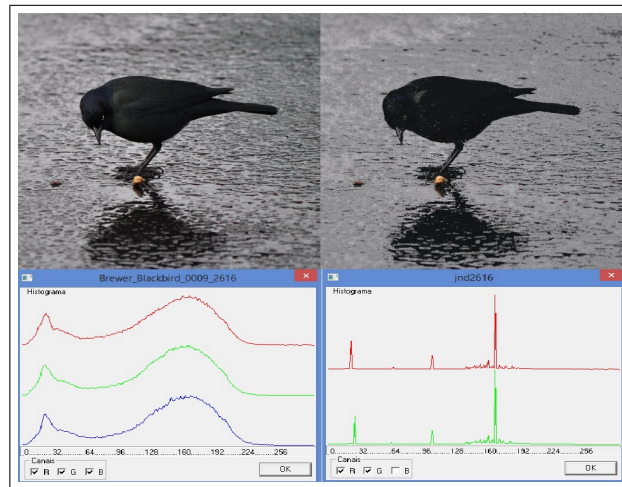


Figura 4.7: *Histograma das imagens processadas no JND.*

Na etapa seguinte as imagens passam pelo algoritmo K-Means, com o objetivo de extrair as características de cores da imagem buscando a segmentação da mesma. Neste processamento são utilizados 5 classes para o agrupamento das características de média das cores nos canais RGB.

Posteriormente as imagens têm extraídos os conjuntos de pixels que tocam a borda. Depois as imagens são submetidas a um processamento que retorna para cada pixel resultante a cor de origem inicial, baseado nas imagens originais. Então após a definição do pixel original, a imagem é transformada para níveis de cinza.

Na terceira etapa ocorre o cálculo da dimensão fractal objetivando extrair as características da textura da imagem. O cálculo é baseado no estudo de Chen et al.(2001), que utiliza uma técnica não morfológica do tipo *box – counting* ou janelas  $r$  que têm seu tamanho definido para cada pixel como sendo:  $r = \{3, 5, 7, 9, 11\}$ (CHEN et al., 2001). O valor da dimensão fractal atribuído a cada pixel é dado pela equação:

$$D2d = 3 - s/2 \quad (4.1)$$

Na sequência as imagens são novamente submetidas ao K-Means. Neste processo são utilizadas 3 classes para o agrupamento das características da curtose do canal R. Então ao final desta etapa as imagens são binarizadas.

Na quinta etapa as imagens binárias são processadas e um algoritmo de Rotulação - Labeling. E então passam por outro algoritmo que realiza a análise dos rótulos e preserva

apenas aquele com maior número de pixel, excluindo da imagem os rótulos menores. E a imagem resultante das etapas descritas é aquela que contem o pássaro segmentado.

Após a segmentação, cada pixel resultante na imagem recebe o valor do pixel original, conforme apresentado na Figura 4.8 .

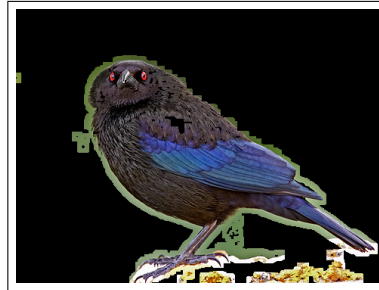


Figura 4.8: *Recuperação das cores originais da imagem segmentada.*

A Figura 4.9 ilustra 04 imagens de pássaros de diferentes espécies em que o método proposto foi testado.

### 4.3 Métricas para Avaliação de Desempenho

As métricas aplicadas para a validação do método proposto, foram baseadas no estudo de Fawcett publicado em 2006 e são demonstradas a seguir:

- um pixel verdadeiro positivo (TP - True Positive) é um pixel pertencente ao pássaro (foreground) na segmentação ideal, o qual também é considerado um pixel do pássaro (foreground) no método desenvolvido;
- um pixel verdadeiro negativo (TN - True Negative) é um pixel pertencente ao fundo da imagem (background) na segmentação ideal, o qual também é considerado um pixel de fundo da imagem (background) no método desenvolvido;
- um pixel falso negativo (FN - False Negative) é um pixel pertencente ao pássaro (foreground) na segmentação ideal, o qual na segmentação desenvolvida foi considerado como fundo da imagem (background);
- um pixel falso positivo (FP - False Positive) é um pixel pertencente ao fundo da imagem (background) na segmentação ideal, o qual na segmentação gerada pelo presente método foi considerado como pássaro (foreground) (FAWCETT, 2006).

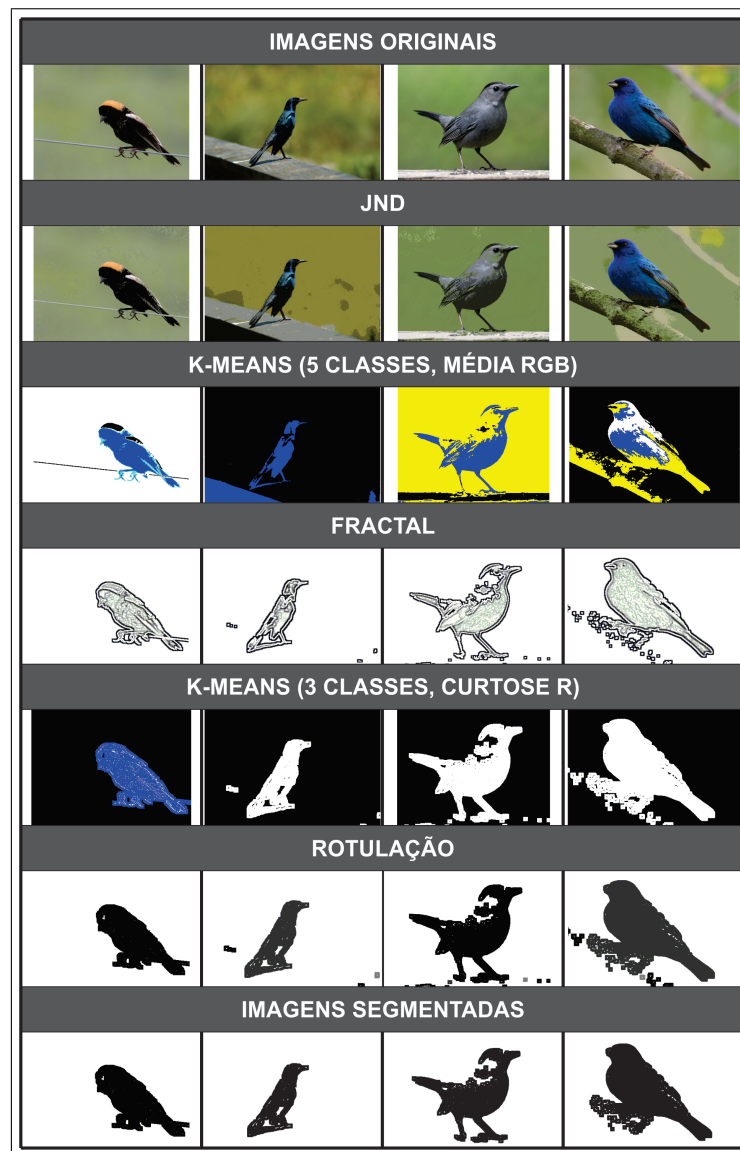


Figura 4.9: Resultados para 4 espécies de pássaros.

A matriz de confusão adotada para avaliação está descrita na Tabela 4.1.

Tabela 4.1: Matriz de confusão para avaliação da segmentação a partir da relação ground-truth e resultado da segmentação proposta (FAWCETT, 2006).

	<b>Segmentação Correta</b>	<b>Segmentação Incorreta</b>
Positivo	Verdadeiro Positivo (TP)	Falso Positivo (FP)
Negativo	Falso Negativo (FN)	Verdadeiro Negativo (TN)

Abaixo estão descritas as nove métricas aplicadas para a avaliação do método de segmentação desta pesquisa:

- *ME – Misclassification Error*: representa o erro de classificação que retorna 0 em caso de perfeita segmentação:

$$ME = 1 - \frac{|B_g \cap B_s| + |F_g \cap F_s|}{|B_g| + |F_g|} \quad (4.2)$$

onde  $(B_g, F_g)$  representam o fundo (background) e o primeiro plano da imagem (foreground) ground truth e  $(B_s, F_s)$  representam o fundo (background) e o primeiro plano (foreground) obtidos pelo método de segmentação avaliado, respectivamente, sendo  $||$  a cardinalidade.

- *RAE – Relative Foreground Area Error* representa o erro relativo ao primeiro plano (foreground) da imagem, valores mais baixos significam maior precisão do método, com o valor próximo de zero tem-se uma segmentação estatisticamente perfeita retornando 0 em caso de segmentação perfeita:

$$RAE = \begin{cases} \frac{|F_g| - |F_s|}{|F_g|} & \text{se } |F_g| > |F_s| \\ \text{ou} \\ \frac{|F_s| - |F_g|}{|F_s|} & \text{se } |F_g| < |F_s| \end{cases} \quad (4.3)$$

onde  $F_g$  representa o primeiro plano (foreground) da imagem ground truth e  $F_s$  representa o primeiro plano (foreground) obtido pelo método de segmentação que está sendo avaliado, respectivamente, sendo  $||$  a cardinalidade.

- *P – Precision* - A precisão representa o valor da predição positiva (número de pixels positivos por total de pixels cobertos), muito influenciada pela especificidade e pouco pela sensibilidade. Sensibilidade é o número de casos (pixels) positivos que são verdadeiramente positivos e especificidade é o número de casos (pixels) negativos que são verdadeiramente negativos. Deve retornar 1 em caso de segmentação perfeita:

$$P = \frac{TP}{TP + FP} \quad (4.4)$$

onde *TP* - é o pixel true positive (verdadeiro positivo) e *FP* - é o pixel false positive (falso positivo).

- *A – Accuracy* representa a métrica de acurácia que retorna 1 em caso de segmentação



perfeita:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.5)$$

onde  $TP$  - é o pixel true positive (verdadeiro positivo),  $TN$  - é o pixel true negative (verdadeiro negativo),  $FP$  - é o pixel false positive (falso positivo) e  $FN$  - é o pixel false negative (falso negativo).

- $R$  - *Recall* representa o valor da cobertura (abrangência) de casos muito influenciada pela sensibilidade e pouco pela especificidade. É calculada por número de casos cobertos pelo número total de casos aplicáveis., e que retorna 1 em caso de segmentação perfeita:

$$R = \frac{TP}{TP + FN} \quad (4.6)$$

onde  $TP$  - é o pixel true positive (verdadeiro positivo) e  $FN$  - é o pixel false negative (falso negativo).

- $E$  - *Error* é o erro total absoluto. Em todas as mensurações de erro, valores mais baixos significam maior precisão do modelo, com o valor próximo de zero temos o modelo estatisticamente perfeito e retorna 0 em caso de segmentação perfeita:

$$E = \frac{FP + FN}{TP + FP + TN + FN} \quad (4.7)$$

onde  $TP$  - é o pixel true positive (verdadeiro positivo),  $TN$  - é o pixel true negative (verdadeiro negativo),  $FP$  - é o pixel false positive (falso positivo) e  $FN$  - é o pixel false negative (falso negativo).

- $FM$  - *F - Measure* representa a média harmônica que considera o valor de Recall e Precision em uma única fórmula e retorna 1 em caso de segmentação perfeita:

$$FM = \frac{2 * P * R}{P + R} \quad (4.8)$$

Onde  $R$  é o valor de cobertura (recall) e  $P$  é o valor da precisão (precision).

- $NRM$  - *Negative Rate Metric* representa a métrica de taxa negativa que retorna 0 em caso de segmentação perfeita:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2}$$

e  $NR_{FN} = \frac{FN}{TP + FN}$

$$e \quad NR_{FP} = \frac{FP}{TN + FP}$$

onde  $TP$  - é o pixel true positive (verdadeiro positivo),  $TN$  - é o pixel true negative (verdadeiro negativo),  $FP$  - é o pixel false positive (falso positivo) e  $FN$  - é o pixel false negative (falso negativo).

De acordo com Fawcett, numa análise de resultados, pode-se dar ênfase as taxas de  $FM$  e  $NRM$  como sendo as mais representativas em termos de eficiência (FAWCETT, 2006).

## 4.4 Análise dos Resultados

Durante o desenvolvimento do método muitas foram as dificuldades encontradas para atingir o objetivo de realizar a segmentação automática. O fato de não ter sido utilizado uma técnica com caixa delimitadora, fez com que o processamento das imagens necessitasse de recursos computacionais mais robustos. Pois o algoritmo que implementa o método percorre a imagem inteira para realizar as tarefas de redução de cores, a extração de características de cor e textura.

Os resultados do desempenho do método proposto foram tabulados conforme os 12 índices (métricas) propostos por Fawcett descritas na seção anterior, conforme apresentadas na Tabela 4.2.

Tabela 4.2: Tabela de índices do desempenho alcançado pelo método proposto. ME (Misclassification Error), RAE (Relative Foreground Area Error), True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Precision (P), Accuracy (A), Recall (R), Error (E), F-measure (FM), Negative Rate Metric (NRM)

1200 Images	ME	RAE	TP	TN	FP	FN	P	A	R	E	FM	NRM
Média	3,05	3,69	91,62	81,82	18,18	8,38	85,47	90,91	88,65	9,09	86,62	9,09
Desvio Padrão	3,50	4,45	5,14	12,79	12,79	5,14	7,90	6,40	8,53	6,40	6,15	6,40

A seguir são descritos os valores da média dos índices obtidos pelo método proposto ao final do processamento de 1200 imagens:

- ME - Misclassification Error de  $(3,05 \pm 3,50)\%$  que representa o erro de classificação do pixel.
- RAE - Relative Foreground Area Error de  $(3,69 \pm 4,45)\%$  que representa o erro relativo ao primeiro plano (foreground) da imagem.
- P - Precision de  $(85,47 \pm 7,9)\%$  que representa o valor da predição positiva (número de pixels positivos por total de pixels cobertos).

- A - Accuracy de  $(90,91 \pm 6,40)\%$  que representa a métrica de acurácia.
- R - Recall de  $(88,65 \pm 8,53)\%$  que representa o valor da cobertura de casos.
- E - Error de  $(9,09 \pm 6,40)\%$  que é o erro total absoluto.
- FM - F-Measure de  $(86,62 \pm 6,15)\%$  que representa a média harmônica que considera o valor de Recall e Precision em uma única fórmula.
- NRM - Negative Rate Metric de  $(9,09 \pm 6,40)\%$  que representa a métrica de taxa negativa.

## 4.5 Considerações Finais

Verificou-se que os resultados do desempenho do método proposto demonstram sua eficiência na tarefa de segmentação automática de pássaros. Trabalhos relacionados no estado da arte que utilizaram a mesma base de imagens e modelo de cor RGB, foram empregados na comparação dos valores de média de acurácia alcançada por cada um deles. É pertinente fazer algumas considerações sobre estes trabalhos. Estes trabalhos integram o gráfico na Figura 4.10, a saber:

- O método de Das e Manmatha (2001) que obteve um valor de 87% de acurácia, faz uso de uma técnica de segmentação por cores que estão presentes na borda da imagem criando então uma região de interesse (boundary region), a partir de medidas de distância entre a borda e a região central da imagem definidas como parâmetros no início do processamento. Por esta particularidade é considerada uma segmentação automática.
- O trabalho de Marini, Facon e Koerich (2013) que atingiu um valor de 71% de acurácia, é uma adaptação no método de Das e Manmatha (2001) que utiliza uma caixa delimitadora para segmentar o pássaro, baseado nas cores presentes na borda da imagem.

Com base nas análises realizadas entre os resultados dos trabalhos correlatos e o método proposto, pode-se verificar que o objetivo geral desta pesquisa foi atingido com eficiência.

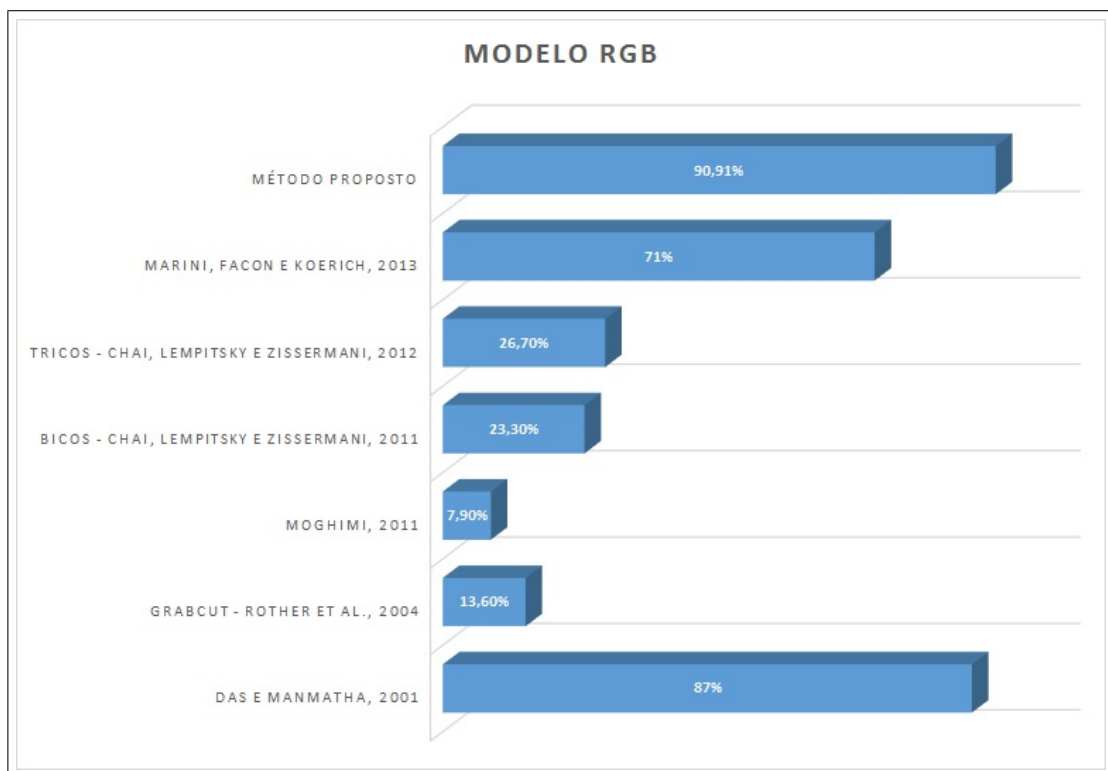


Figura 4.10: *Comparativo com os trabalhos correlatos.*

## Capítulo 5

### Conclusão

A segmentação de imagens coloridas tem sido estudada por muitos anos e possibilita o desenvolvimento de soluções para inúmeros problemas nas mais diversas áreas do conhecimento.

Neste trabalho, o foco foi segmentação de pássaros em imagens digitais coloridas, que por sua vez, foram capturadas em ambientes naturais sem restrições quanto a escala, luminosidade e localização do pássaro na imagem.

Os resultados obtidos neste trabalho atingiram os objetivos propostos inicialmente e buscaram responder a hipótese e estabelecer conclusões em relação as etapas pertencentes ao método proposto.

Demonstrando que um sistema de segmentação automatizado, isto é, sem a necessidade de interação do usuário para delimitar a área de interesse ou fazer marcações no objeto de interesse, pode ser viável em aplicações reais, auxiliando no desenvolvimento de softwares de reconhecimento de espécies de pássaros.

Durante os a execução das etapas do método, verificou-se que uma abordagem que realize a filtragem morfológica para imagens em níveis de cinza e para imagens coloridas, pode vir a ser uma alternativa promissora para a melhoria do método proposto, e conseqüentemente a melhora dos resultados obtidos. Portanto cabe citar que a filtragem morfológica poderá ser vista como uma possibilidade de continuação desta pesquisa para um futuro próximo, bem como a realização de testes com imagens em outros modelos de cores, como o HSI e o HSV.

Outra proposta de trabalho futuro permeia o conhecimento sobre o cálculo da dimensão fractal para imagens coloridas. E também é vista como uma pesquisa viável para uma implementação futura no método proposto nesta dissertação.

As contribuições apresentadas no desenvolvimento de um método automatizados de segmentação, se destacam como uma solução abrangente para a segmentação de imagens

complexas, com grande quantidade de informação de fundo (background), podendo ser aplicadas também para a segmentação de objetos de diferentes áreas de conhecimento.

Os resultados do presente trabalho foram obtidos a partir de uma avaliação metódica e criteriosa realizada, por meio de diferentes abordagens, o que promoveu as condições necessárias para conclusões e produção de conhecimento apropriado ao domínio da visão computacional na execução da tarefa de segmentação automática.

## Referências Bibliográficas

ACEVEDO, M. A. et al. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, v. 4, n. 4, p. 206 – 214, 2009. ISSN 1574-9541. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954109000351>>.

BARDELI, R. et al. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 31, n. 12, p. 1524–1534, set. 2010. ISSN 0167-8655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2009.09.014>>.

BHOYAR, K.; KAKDE, O. Color image segmentation based on jnd color histogram. *International Journal of Image Processing (IJIP)*, v. 3, n. 6, p. 283, 2010.

CAI, J. et al. Sensor network for the monitoring of ecosystem: Bird species recognition. In: *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*. [S.l.: s.n.], 2007. p. 293–298.

CALIMAN, A.; IVANOVICI, M.; RICHARD, N. Fractal feature-based color image segmentation for a healthcare application in dermatology. In: *E-Health and Bioengineering Conference (EHB), 2011*. [S.l.: s.n.], 2011. p. 1–4.

CALIXTO, E. *Granulometria morfológica em espaços de cores: estudo da ordenação espacial*. [S.l.]: Universidade Federal Fluminense - Pós Graduação - Depto de Ciência da Computação, 2005.

CHAI, Y.; LEMPITSKY, V.; ZISSERMAN, A. Bicos: A bi-level co-segmentation method for image classification. In: *IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2011.

CHAI, Y. et al. Tricos: A tri-level class-discriminative co-segmentation method for image classification. In: FITZGIBBON, A. et al. (Ed.). *Computer Vision ECCV 2012*. [S.l.]: Sprin-

ger Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, v. 7572). p. 794–807. ISBN 978-3-642-33717-8.

CHANG, P.; WANG, X.; HUANG, J. Color image segmentation based on visual perception. In: *Information Science and Technology (ICIST), 2012 International Conference on*. [S.l.: s.n.], 2012. p. 425–429.

CHEN, W.-S. et al. Algorithms to estimating fractal dimension of textured images. In: *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*. [S.l.: s.n.], 2001. v. 3, p. 1541–1544 vol.3. ISSN 1520-6149.

COMANICIU, D.; MEER, P. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 24, n. 5, p. 603–619, May 2002. ISSN 0162-8828.

COSSINS, D. Birds monitor pollution. *The Scientist*, 2012. Disponível em: <<http://www.the-scientist.com/?articles.view/articleNo/33329/title/Birds-Monitor-Pollution/>>.

DAS, M.; MANMATHA, R. Automatic segmentation and indexing in a database of bird images. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.: s.n.], 2001. v. 2, p. 351–358 vol.2.

DOUGHERTY, E. *An introduction to morphological image processing*. [S.l.]: SPIE - Optical Engineering Press, 1992. (Tutorial texts in optical engineering).

DOUGHERTY, E.; LOTUFO, R. *Hands-on Morphological Image Processing*. [S.l.]: SPIE Press, 2003. (Spie Press Series). ISBN 081944720X, 9780819447203.

FACON, J. *Morfologia Matemática: Teoria e Exemplos*. [S.l.]: Editora Champagnat, 1996.

FAWCETT, T. An introduction to roc analysis. *Pattern Recogn. Lett.*, Elsevier Science Inc., New York, NY, USA, v. 27, n. 8, p. 861–874, jun. 2006. ISSN 0167-8655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2005.10.010>>.

FILHO, O. M.; NETO, H. V. *Processamento Digital de Imagens*. Rio de Janeiro, Brazil: Editora Brasport, 1999.

GLEICK, J.; DUTRA, W. *Caos: a criação de uma nova ciência*. Campus, 1998. ISBN 9788570015945. Disponível em: <<https://books.google.com.br/books?id=3bZ8SwAACAAJ>>.



GOËAU, H. et al. Life clef 2015 bird task. *CLEF 2015 CEUR-WS Proceedings*, 2015. Disponível em: <<http://www.imageclef.org/lifeclef/2015/bird>>.

GOMES, J.; VELHO, L. *Computação Gráfica: Imagem*. [S.l.]: Impa/sbm, 1994. (Série de Computação Gráfica).

GONZALEZ, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. [S.l.]: Editora Blucher, 2000. ISBN 85-212-0264-4.

JAIN, A. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989. (Prentice-Hall information and system sciences series). ISBN 9780133325782. Disponível em: <<http://books.google.com.br/books?id=H0PLngEACAAJ>>.

KASTEN, E. P.; MCKINLEY, P. K.; GAGE, S. H. Ensemble extraction for classification and detection of bird species. *Ecological Informatics*, v. 5, p. 153–166, 2010. Disponível em: <<http://www.cse.msu.edu/6.953314e-310mckinley/Pubs/files/Kasten.Ensemble.2010.pdf>>.

KRAFT, R.; KAUER, J. *Estimating the Fractal Dimension from Digitized Images*. Germany, 1995. Disponível em: <<http://tka4.org/materials/lib/Articles-Books/DSP/Fractals/ALG.PDF>>.

LI, H.; LIU, K. J. R.; LO, S. C. B. Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms. *IEEE Transactions on Medical Imaging*, v. 16, n. 6, p. 785–798, Dec 1997. ISSN 0278-0062.

MANDELBROT, B. *The Fractal Geometry of Nature*. Henry Holt and Company, 1983. (Einaudi paperbacks). ISBN 9780716711865. Disponível em: <<https://books.google.co.uk/books?id=0R2LkE3N7-oC>>.

MARINI, A.; FACON, J.; KOERICH, A. Bird species classification based on color features. In: *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. [S.l.: s.n.], 2013. p. 4336–4341.

MOGHIMI, M. *Using Color for Object Recognition*. [S.l.], 2011.

NAJMAN, L.; TALBOT, H. *Mathematical Morphology*. [S.l.]: Wiley, 2013. (Iste Series). ISBN 9781118600900.

ORTIZ, F. et al. Colour mathematical morphology for neural image analysis. *Real-Time Imaging*, v. 8, n. 6, p. 455 – 465, 2002. ISSN 1077-2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1077201402902883>>.

PAPAMARKOS, N.; GATOS, B. A new approach for multilevel threshold selection. *CVGIP: Graphical Models and Image Processing*, v. 56, n. 5, p. 357 – 370, 1994. ISSN 1049-9652. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1049965284710339>>.

ROTHER, C.; KOLMOGOROV, V.; BLAKE, A. "grabcut": Interactive foreground extraction using iterated graph cuts. In: *ACM SIGGRAPH 2004 Papers*. New York, NY, USA: ACM, 2004. (SIGGRAPH '04), p. 309–314. Disponível em: <<http://doi.acm.org/10.1145/1186562.1015720>>.

SEKERCIOGLU, C. H.; DAILY, G. C.; EHRLICH, P. R. Ecosystem consequences of bird declines. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, v. 101, n. 52, p. 18042–18047, December 2004.

TAN, P. et al. *Introdução ao datamining: mineração de dados*. [S.l.]: Ciencia Moderna, 2009. ISBN 9788573937619.

TOBAR, M.; P., P. C. andGonzález; G, A. Teoria e implementacion de ordenes totales en los espacios de color y morfologia digital. *XXVIII Jornadas de Automática, Huelva*, 2007.

VAS, E. et al. Approaching birds with drones: first experiments and ethical guidelines. *Biology Letters*, The Royal Society, v. 11, n. 2, 2015. ISSN 1744-9561. Disponível em: <<http://rsbl.royalsocietypublishing.org/content/11/2/20140754>>.

WAH, C. et al. *The Caltech-UCSD Birds-200-2011 Dataset*. [S.l.], 2011.

WELINDER, P. et al. The multidimensional wisdom of crowds. In: LAFFERTY, J. D. et al. (Ed.). *Advances in Neural Information Processing Systems 23*. Curran Associates, Inc., 2010. p. 2424–2432. Disponível em: <<http://papers.nips.cc/paper/4074-the-multidimensional-wisdom-of-crowds.pdf>>.

YAO, B. A codebook-free and annotation-free approach for fine-grained image categorization. In: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Washington, DC, USA: IEEE Computer Society, 2012. (CVPR '12), p. 3466–3473. ISBN 978-1-4673-1226-4. Disponível em: <<http://dl.acm.org/citation.cfm?id=2354409.2355035>>.

ZHU, Z.; GAO, J.; YU, H. Face detection based on fractal and complexion model in the complex background. In: *Chaos-Fractals Theories and Applications (IWCFTA), 2010 International Workshop on*. [S.l.: s.n.], 2010. p. 491–495.