

Luiz Carlos A. M. Cavalcanti

**Detecção de elementos antrópicos em imagens  
aéreas da floresta amazônica**

Manaus

2016

Luiz Carlos A. M. Cavalcanti

# **Detecção de elementos antrópicos em imagens aéreas da floresta amazônica**

Dissertação apresentada ao Programa de Pós-graduação em Informática da Universidade Federal do Amazonas como requisito para obtenção do grau de Mestre em Informática. Área de concentração: Ciência da Computação

PODER EXECUTIVO  
MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO AMAZONAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Orientadora: Prof<sup>ª</sup>. Dra. Eulanda Miranda dos Santos

Manaus

2016

## Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

C376d Cavalcanti, Luiz Carlos Amaral Mendonça  
Detecção de elementos antrópicos em imagens aéreas da  
floresta amazônica / Luiz Carlos Amaral Mendonça Cavalcanti.  
2016  
94 f.: il. color; 31 cm.

Orientadora: Eulanda Miranda dos Santos  
Dissertação (Mestrado em Informática) - Universidade Federal do  
Amazonas.

1. Aprendizado de máquina. 2. Segmentação. 3. Classificação. 4.  
Imagens aéreas. 5. Elementos antrópicos. I. Santos, Eulanda  
Miranda dos II. Universidade Federal do Amazonas III. Título

Luiz Carlos A. M. Cavalcanti

## **Detecção de elementos antrópicos em imagens aéreas da floresta amazônica**

Dissertação apresentada ao Programa de Pós-graduação em Informática da Universidade Federal do Amazonas como requisito para obtenção do grau de Mestre em Informática. Área de concentração: Ciência da Computação

Trabalho aprovado. Manaus, 1 de julho de 2016:

---

**Prof<sup>a</sup>. Dra. Eulanda Miranda dos Santos**  
Presidente  
PPGI/UFAM

---

**Prof. Dr. George Darmiton da Cunha Cavalcanti**  
PGCC/UFPE

---

**Prof. Dr. José Reginaldo Hughes Carvalho**  
PPGI/UFAM

---

**Prof. Dr. José Luiz de Souza Pio**  
PPGEE/UFAM

Manaus  
2016

À Isabel Wittmann

# Agradecimentos

À professora e orientadora Dra. Eulanda Santos, pela paciência, revisões, direcionamentos e mais paciência.

Ao Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENCI-PAM), especialmente à Dra. Solange Costa pela contribuição na avaliação e validação dos resultados preliminares, sem as quais este trabalho não seria possível.

Aos amigos do trabalho e academia, principalmente à Arthur Batista, que fora os dois durante todo o percurso.

Aos professores Edleno Moura, Marco Cristo, Altigran Soares, José Reginaldo Hughes e José Pio pelas aulas, cobranças e acompanhamento ao longo do curso.

À família, direta e indireta, pelo apoio vitalício.

*“As forças da natureza são máquinas infinitas, as máquinas são forças limitadas”*  
*Victor Hugo, Os Trabalhadores do Mar*

# Sumário

<b>Lista de ilustrações</b> . . . . .	<b>viii</b>
<b>Lista de tabelas</b> . . . . .	<b>ix</b>
<b>1 Introdução</b> . . . . .	<b>3</b>
1.1 Motivação . . . . .	3
1.2 Objetivos . . . . .	4
1.2.1 Geral . . . . .	4
1.2.2 Específicos . . . . .	5
1.3 Organização do Documento . . . . .	5
<b>2 Fundamentação Teórica</b> . . . . .	<b>6</b>
2.1 Processamento Digital de Imagens . . . . .	6
2.1.1 Características de baixo nível . . . . .	6
2.1.2 Características de alto nível . . . . .	8
2.1.3 Segmentação de Imagens . . . . .	11
2.1.4 Avaliação de segmentação . . . . .	12
2.2 Aprendizagem de máquina e reconhecimento de padrões . . . . .	13
2.2.1 Filtragem e Pré-processamento . . . . .	15
2.2.2 Extração e Seleção de Características . . . . .	15
2.2.3 Classificação . . . . .	16
2.2.4 Avaliação de aprendizagem . . . . .	22
<b>3 Trabalhos Relacionados</b> . . . . .	<b>25</b>
3.1 Segmentação de imagens . . . . .	26
3.2 Classificação de imagens aéreas . . . . .	32
3.3 Detecção de anomalias . . . . .	35
<b>4 Metodologia</b> . . . . .	<b>37</b>
4.1 Entrada . . . . .	38
4.2 Segmentador . . . . .	38
4.3 Extrator . . . . .	39
4.4 Classificador . . . . .	39
4.4.1 Classificador multi-classe . . . . .	40
4.4.2 Classificador binário . . . . .	40
4.4.3 Classificador unário . . . . .	41
4.4.4 Conjunto de classificadores unários . . . . .	41
4.5 Saída . . . . .	42
<b>5 Experimentos</b> . . . . .	<b>43</b>
5.1 Base de dados . . . . .	43
5.2 Segmentação . . . . .	44

5.2.1	Protocolo experimental . . . . .	44
5.2.2	Resultados . . . . .	45
5.3	Classificação . . . . .	48
5.3.1	Protocolo experimental . . . . .	49
5.3.2	Classificadores multi-classe . . . . .	51
5.3.3	Classificadores binários . . . . .	53
5.3.4	Classificadores unários . . . . .	56
5.3.5	Conjunto de classificadores unários . . . . .	57
5.4	Discussão . . . . .	60
<b>6</b>	<b>Conclusão e Trabalhos Futuros . . . . .</b>	<b>63</b>
6.1	Trabalhos futuros . . . . .	64
	<b>Referências . . . . .</b>	<b>65</b>
	<b>Apêndices . . . . .</b>	<b>69</b>
	<b>APÊNDICE A A comparison on supervised machine learning classification techniques for semantic segmentation of aerial images of rain forest regions . . . . .</b>	<b>70</b>
	<b>APÊNDICE B Instruções da ferramenta de segmentação manual . . . . .</b>	<b>77</b>

# Lista de ilustrações

Figura 1 – Exemplo de extração de texturas utilizando transformada de Fourier . . . . .	9
Figura 2 – Exemplos de uma mesma imagem segmentada de forma consistente, mas com diferentes granularidades de segmentação . . . . .	12
Figura 3 – Representação de classificador como uma função de bloco . . . . .	14
Figura 4 – Máquina de vetores de suporte . . . . .	19
Figura 5 – Exemplo de gráfico para análise de curva ROC. . . . .	24
Figura 6 – Exemplos de mapas de classes do algoritmo JSEG . . . . .	27
Figura 7 – Exemplo de segmentação mean-shift em uma imagem em tons de cinza . . . . .	29
Figura 8 – Arquitetura geral da solução desenvolvida para detecção de elementos antrópicos em imagens aéreas da floresta amazônica. . . . .	37
Figura 9 – Amostras da base de dados. . . . .	43
Figura 10 – Ferramenta para segmentação manual das imagens. . . . .	44
Figura 11 – Coeficiente de erro global das diferentes segmentações de mesma imagem. . . . .	46
Figura 12 – Coeficiente de erro local das diferentes segmentações de mesma imagem. . . . .	46
Figura 13 – Comparação visual de métodos de segmentação. . . . .	48

# Lista de tabelas

Tabela 1 – Modelo de matriz de confusão. . . . .	22
Tabela 2 – Comparação entre as técnicas de segmentação de imagens, ordenados por desempenho decrescente, conforme resultados em Yuan, Gleason e Cheriyyadat (2013). . . . .	32
Tabela 3 – Comparação entre os trabalhos sobre classificação de imagens aéreas. . . . .	34
Tabela 4 – Comparação entre os trabalhos sobre detecção de anomalias. . . . .	36
Tabela 5 – Comparação de métodos de segmentação em parte da base de imagens deste trabalho, em ordem alfabética. Os erros médios da segmentação manual foram calculados entre diferentes segmentações da mesma imagem. . . . .	47
Tabela 6 – Comparação de métodos de classificação para segmentação das imagens em uma única etapa, ordenados por acurácia. . . . .	48
Tabela 7 – Distribuição de classes na base de segmentos. . . . .	49
Tabela 8 – Atributos gerados a partir da base de segmentos. . . . .	50
Tabela 9 – Atributos selecionados pela técnica de CFS para a abordagem de classificação multi-classe. . . . .	52
Tabela 10 – Comparação de métodos de classificação para regiões segmentadas das imagens, ordenados por acurácia. . . . .	52
Tabela 11 – Comparação de métodos de classificação multi-classe em relação à classe de elementos antrópicos, ordenados pela medida F1. . . . .	52
Tabela 12 – Atributos selecionados pela técnica de CFS para a abordagem de classificação binária. . . . .	53
Tabela 13 – Distribuição de classes na base de segmentos para classificação binária. . . . .	54
Tabela 14 – Comparação de métodos de classificação binária para regiões segmentadas das imagens, ordenados por acurácia. . . . .	54
Tabela 15 – Comparação de métodos de classificação binária em relação à classe de elementos antrópicos, ordenados pela medida F1. . . . .	55
Tabela 16 – Distribuição de classes na base de segmentos para classificação unária. . . . .	56
Tabela 17 – Atributos selecionados pela técnica de CFS para a abordagem de classificação unária. . . . .	56
Tabela 18 – Comparação de métodos de classificação unária para regiões segmentadas das imagens, ordenados por acurácia. . . . .	57
Tabela 19 – Comparação de métodos de classificação unária em relação à classe de elementos antrópicos, ordenados pela medida F1. . . . .	57
Tabela 20 – Atributos selecionados pela técnica de CFS para a abordagem de classificação unária. . . . .	58

Tabela 21 – Acurácia, precisão, revocação e medida F1 para as classes e algoritmos do experimento de conjunto de classificadores unários, sem ordenação específica. . . . .	59
Tabela 22 – Comparação dos 6 melhores conjuntos de classificadores unários em relação à classe de elementos antrópicos, ordenados pela medida F1. . .	59
Tabela 23 – Comparação do desempenho das abordagens de aprendizagem de máquina para a classe de elementos antrópicos, ordenados pela medida F1. . . . .	62

# Lista de abreviaturas e siglas

BSD500	Berkeley Segmentation Data Set and Benchmarks 500
CENSIPAM	Centro Gestor e Operacional do Sistema de Proteção da Amazônia
CFS	Correlation-based Feature Subset Selection
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
FSEG	Factorisation-Based Segmentation
GCE	Global Consistency Error
HSV	Hue-Saturation-Value
JPEG	Joint Photographic Experts Group
JSEG	J Segmentation
KNN	K-Nearest Neighbours
LBP	Local Binary Pattern
LBP-HF	Local Binary Pattern Histogram Fourier
LCE	Local Consistency Error
LIDAR	Light Detection and Ranging
MoG	Mixture of Gaussians
MSEG	Multi-resolution Region Merging Segmentation
OCNN	One-Class Nearest Neighbours
OC-SVM	One-Class Support Vector Machines
PDI	Processamento Digital de Imagens
RADAR	Radio Detection and Ranging
ROC	Receiver Operating Characteristic
SCSVDD	Spatial-Contextual Support Vector Domain Description

SRM	Statistical Region Merging
SVDD	Support Vector Domain Description
SVM	Support Vector Machines
UAV	Unmanned Aerial Vehicle
VANT	Veículo Aéreo Não-Tripulado

# Resumo

Durante o patrulhamento de crimes ambientais, o tempo de resposta é um componente muito importante no sucesso das missões. Geralmente as infrações ocorrem em lugares ermos e de difícil acesso, características que dificultam tanto o patrulhamento quanto a ação de agentes de preservação ambiental. Para aumentar a taxa de sucesso das abordagens e reduzir o risco de vidas humanas, veículos aéreos não-tripulados (VANTs) podem ser usados para cobrir grandes áreas de floresta em pouco tempo, sem que sejam percebidos por infratores, permitindo que os órgãos de patrulhamento dessas áreas possam planejar e agir com mais eficiência na repressão a esses crimes. O novo problema gerado por essa abordagem é a enorme quantidade de dados gerada durante essas missões, que muitas vezes compreendem horas de vídeo. A inspeção manual de todo esse material em busca de elementos antrópicos é muito cansativa e propensa a erros. Este trabalho apresenta uma avaliação de técnicas de segmentação de imagens, inspeção de características a serem extraídas, seguido da classificação supervisionada destes segmentos para detecção de elementos antrópicos em imagens aéreas da floresta amazônica. Além da publicação de uma base de dados com cerca de 3.000 imagens e 10.000 segmentos devidamente rotulados e investiga diferentes estratégias para classificação de elementos antrópicos. Os experimentos realizados obtiveram taxas de erro de consistência inferiores a 8% na segmentação das imagens utilizando o algoritmo SRM e precisão acima de 94% na classificação dos objetos de interesse através de conjuntos de classificadores unários, utilizando os algoritmos One-Class SVM e REPTree.

**Palavras-chaves:** aprendizado de máquina, segmentação, classificação, imagens aéreas, elementos antrópicos

# Abstract

During environmental crimes patrolling, the response time is a very important component for the success of the missions. Generally, infractions occur in remote and hard-access places, characteristics that hinder both the patrolling as well the action of environmental protection agents. To increase the approaches' success rate and reduce the risk of human lives, unmanned aerial vehicles (UAVs) can be used to cover large areas of forest in a short time without being perceived by offenders, allowing the patrolling organs responsible for these areas to plan and act more efficiently in the repression of such crimes. The new problem generated by this approach is the huge amount of data generated during these missions, which often includes hours of video. The manual inspection of all this material in searching for anthropic elements is very tiring and error-prone. This work presents a evaluation of image segmentation techniques, inspections of features to be extracted, followed by a supervised classification of those segments for anthropic element detection in amazon's rain forest aerial images. Besides making publicly available a dataset with more than 3,000 images and 10,000 segments labeled accordingly, this work investigates different strategies for anthropic elements classification. The experiments obtained a consistency error rate inferior to 8% in image segmentation and a precision above 94% on target objects classification through one-class classifiers ensemble, using One-class SVM and REPTree algorithms.

**Keywords:** machine learning, segmentation, classification, aerial images, anthropic elements

# 1 Introdução

Os órgãos de segurança e de patrulhamento ambiental lidam diariamente com missões de alto risco. Essas missões têm grande custo de operação e representam um perigo significativo para os agentes humanos envolvidos. Especificamente na região da floresta amazônica, a área a ser patrulhada possui uma dimensão ímpar: 11 mil quilômetros de fronteiras e 22 mil quilômetros de rios, compondo uma área total de 5.217.423 quilômetros quadrados do que é conhecido como Amazônia Legal, o que corresponde à aproximadamente 61% do território brasileiro (IBGE, 2015).

Com o intuito de minimizar os custos, aumentar a eficiência e mitigar os riscos humanos de diversos tipos de atividades, são empregados Veículos Aéreos Não-Tripulados (VANTs). Trata-se de veículos autônomos ou controlados remotamente, que têm como objetivo a agilidade e a discrição no reconhecimento de áreas isoladas ou hostis, antes que uma equipe de agentes humanos precise intervir. Essa abordagem permite que as eventuais ameaças sejam previamente identificadas, e o comando tático possa tomar as melhores decisões sobre quando e como agir na região em questão.

Os VANTs têm sido utilizados em uma miríade de missões de diferentes naturezas, com destaque para o combate a crimes em áreas de fronteira. Na região da Floresta Amazônica, os veículos podem ser utilizados no combate ao tráfico de animais, à extração ilegal de madeira e outros crimes ambientais (SILVA, 2013).

Por acontecerem em regiões isoladas e de difícil acesso, os responsáveis por essas infrações costumam escapar antes que os agentes de segurança possam agir. A aproximação de humanos alerta os infratores, pois os veículos (barcos, helicópteros e veículos terrestres utilitários) usados na abordagem produzem sons bastante perceptíveis nesses ambientes. Por esses motivos, os VANTs representam uma boa alternativa para possibilitar flagrantes e o planejamento tático dessas missões.

Equipados com câmeras de diversos tipos e alcances, esses veículos são capazes de gerar muitas horas de vídeo a cada missão, tornando a análise simultânea ou posterior do material gerado um grande esforço para as pessoas responsáveis pelo planejamento tático, presentes em salas de controle para onde os vídeos são transferidos ao vivo, ou analisados posteriormente.

## 1.1 Motivação

A quantidade maciça de dados gerados pelos VANTs leva à necessidade de desenvolvimento de procedimentos capazes de avaliar essa grande quantidade de material e

identificar os prováveis pontos de interesse. Uma solução computacional capaz de fazer tal análise seria importante por possibilitar a rápida avaliação do material gerado por uma operação, ou mesmo para que uma quantidade menor de profissionais seja necessária para o acompanhamento em tempo real de diversos VANTs em uma mesma missão. Isso cria uma demanda para a filtragem e classificação desses dados em informações ou alertas relevantes aos agentes humanos que supervisionam esses equipamentos.

Ao provar a eficiência de uma implementação automatizada dessa classificação, um significativo aumento pode ser alcançado na qualidade e agilidade da detecção de elementos antrópicos em ambientes silvestres, bem como na redução da margem de erro dos agentes humanos envolvidos. A detecção de elementos antrópicos refere-se à identificação de objetos estranhos ao padrão normal do ambiente silvestre, oriundo de atividades humanas.

Através de processamento de imagens digitais e reconhecimento automático de padrões, é possível avaliar o material com maior rapidez, e sinalizar os pontos de possível interesse dos profissionais que acompanham as missões.

Baseando-se em elementos antrópicos encontrados nas imagens aéreas de regiões silvestres, áreas com relativa previsibilidade de paisagem, é possível buscar padrões e reconhecer ameaças ou interesses que possam passar despercebidos por olhos humanos, especialmente em um material com extensão de horas. Tais elementos antrópicos incluem acampamentos de caçadores, pistas clandestinas, estradas, embarcações, ou seja, indícios de presença humana em áreas de selva.

Até a conclusão desta dissertação, não foi possível encontrar na literatura trabalhos que abordassem a detecção de elementos antrópicos em regiões de floresta amazônica, tampouco que utilizassem imagens aéreas da região obtidas por aeronaves. Diante desse contexto, este projeto propõe o uso de técnicas de processamento digital de imagens e de aprendizagem de máquina para detectar elementos antrópicos em imagens obtidas por veículos aéreos, de uma região da Floresta Amazônica.

## 1.2 Objetivos

### 1.2.1 Geral

Este trabalho tem como objetivo propor uma abordagem utilizando processamento digital de imagens (PDI) e aprendizagem de máquina para o problema de detecção automática de elementos antrópicos em imagens aéreas da floresta amazônica. A intenção é demonstrar que a identificação automática destes elementos pode trazer benefícios, em especial agilidade, ao processo que hoje é completamente dependente de agentes humanos, e produzir elevadas taxas de acerto.

### 1.2.2 Específicos

- Organizar uma base de dados de imagens devidamente rotuladas que sirva de referência para futuros estudos desta problemática.
- Investigar e definir métodos para segmentação de imagens aéreas de floresta amazônica.
- Investigar e definir métodos para extração e seleção de características mais adequadas ao problema em questão.
- Apontar o melhor conjunto de técnicas para a detecção de elementos antrópicos em imagens aéreas da floresta amazônica.

## 1.3 Organização do Documento

O restante deste documento está organizado da seguinte forma: no capítulo 2 discorreremos sobre a fundamentação teórica necessária para o entendimento do trabalho proposto. No capítulo 3 os trabalhos relacionados são apresentados, com ordem e critérios descritos neste mesmo capítulo. Uma comparação das referências também é feita. No capítulo 4 é descrita a metodologia que será empregada no desenvolvimento deste trabalho de pesquisa. No capítulo 5 são apresentadas a base de dados, sua preparação e a descrição dos experimentos realizados, com os resultados alcançados. Por fim, no capítulo 6 uma conclusão para o trabalho é apresentada, consolidando os resultados dos experimentos e apontando possibilidades para trabalhos futuros.

## 2 Fundamentação Teórica

Neste capítulo são discutidos os conceitos básicos ao entendimento deste trabalho, abrangendo processamento digital de imagens, segmentação de imagens, aprendizagem de máquina, extração de características e avaliação de desempenho dos métodos utilizados.

### 2.1 Processamento Digital de Imagens

Uma imagem pode ser definida por uma função bidimensional,  $f(x, y)$ , onde  $x$  e  $y$  são coordenadas espaciais de um plano, e a amplitude de  $f$  para um par de coordenadas  $(x, y)$  é chamada de intensidade da imagem naquele ponto. Quando  $x$ ,  $y$  e a amplitude de valores de  $f$  representam quantidades finitas e discretas, podemos chamar a imagem de imagem digital. A linha de pesquisa de processamento digital de imagens se refere ao processamento digital dessas imagens através de um computador digital (GONZALEZ; WOODS, 2002).

Uma imagem digital é composta por um número finito de elementos, cada um com um valor ( $f$ ) e localização ( $x$  e  $y$ ) particulares. Estes elementos são chamados de elementos da imagem ou *pixels*, do inglês *picture elements*.

Ainda de acordo com Gonzalez e Woods (2002), o interesse em processamento digital de imagens advém de duas principais áreas de aplicação: melhoramento de informações pictóricas para interpretação humana; e processamento de imagens para armazenamento e transmissão de dados, e representação para percepção de máquinas autônomas.

Para praticamente qualquer aplicação de processamento de imagens, características destas mesmas imagens precisam ser extraídas. Segundo Yadav e Yadav (2009), as características de imagens digitais podem ser classificadas em baixo e alto nível. As características de baixo nível são extraídas diretamente das imagens, geralmente descrevendo atributos locais ou de pequenos agrupamentos de pixels. As características de alto nível são extraídas a partir de características de baixo nível, e compreendem informações sobre regiões maiores da imagem, ou mesmo da imagem completa. Na próxima seção, são descritas as características utilizadas neste trabalho.

#### 2.1.1 Características de baixo nível

As características de baixo nível de uma imagem compreendem aspectos técnicos da composição desta imagem. Informações sobre cores, intensidade e bordas de uma imagem são fundamentais para a extração de características de alto nível, mas também

são úteis diretamente em diversas tarefas de processamento digital de imagens, tais como segmentação e filtragem.

### Espaços de cores

O uso de cores em processamento de imagem é motivado por dois principais fatores. Primeiramente, cor é um poderoso descritor que frequentemente simplifica a identificação e extração de objetos de uma cena. Em segundo lugar, humanos podem discernir milhares de tons e intensidades de cores, comparado com apenas duas dúzias de tons de cinza (GONZALEZ; WOODS, 2002).

Os espaços de cores, também conhecidos como modelos de cores ou sistemas de cores, são maneiras padronizadas de representar a cor de pixels em uma imagem digital. A definição do modelo de cores persegue, geralmente, dois propósitos: o hardware ou meio pelo qual a cor será reproduzida (RGB para representação em telas, CMYK para impressão, por exemplo), ou a manipulação de certos elementos da cor, como tom, saturação, etc.

O espaço de cores RGB (*Red, Green, Blue*) é baseado numa representação de coordenadas cartesianas de três canais de cores: vermelho (R), verde (G) e azul (B). Os 3 canais possuem a mesma amplitude, portanto, em uma imagem RGB de 24 bits, cada canal é representado por 8 bits, que pode ter seu valor de 0 a 255.

O espaço de cores RGB é ideal para a representação de hardware que combine cores através da mistura de luzes, tais como monitores e projetores. Seus canais são cores, e a combinação deles é intuitiva à composição ou geração de cores, mas é muito pouco intuitiva para seres humanos.

Com esta finalidade, outros espaços de cores podem ser menos intuitivos a seres humanos, mas são definidos de forma que outras características das cores sejam facilmente extraídas. O HSI (*Hue, Saturation, Intensity*), também conhecido como HSV (*Hue, Saturation, Value*) representa uma cor através de seu tom, saturação e intensidade. Este espaço de cores facilita a extração de algumas características como saturação ou intensidade. Seus canais também se parecem mais com a forma como seres humanos descrevem cores, já que um dos canais representa a informação de cor pura, enquanto os outros canais tratam da saturação (força) e intensidade (brilho) da cor. Todos os canais têm seus valores representados por percentuais, com valores decimais de 0 a 1.

### Intensidade

É comum que se precise extrair a intensidade dos pixels de uma imagem, seja para análise de textura, para composição de histograma ou qualquer outro propósito. Quando a imagem está no espaço de cores HSI, basta que o canal de intensidade (I) seja coletado. Quando se está trabalhando no espaço de cores RGB, uma conversão precisa ser feita:

$$I = \frac{R + G + B}{3} \quad (2.1)$$

onde  $I$  é a intensidade do pixel e  $R$ ,  $G$  e  $B$  são, respectivamente, os canais vermelho, verde e azul do pixel. Esta fórmula também pode ser utilizada para converter imagens coloridas RGB em imagens em tons de cinza, embora, por características intrínsecas ao sistema de visão humano, normalmente alguns pesos diferentes são atribuídos aos canais quando o objetivo é a visualização humana da imagem, ao invés da média aritmética simples da equação 2.1, que é considerada pela literatura uma aproximação suficientemente adequada para sistemas de visão computacional.

### 2.1.2 Características de alto nível

Muitas vezes, características de baixo nível não são suficientes para descrever aspectos da imagem a ser processada. Especialmente em tarefas de visão computacional, é muito comum que seja necessário extrair características ou conceitos a partir de características de baixo nível. Informações de textura, morfologia e pontos de interesse são algumas destas características.

#### Textura

De acordo com Nixon e Aguado (2008), não há uma definição formal de textura. Gonzalez e Woods (2002), por sua vez, definem textura como uma medida de propriedades sensoriais como maciez, coesão e regularidade. Um campo de pesquisa que constantemente apresenta novidades substanciais, as soluções para descrição e extração de texturas em imagens digitais partem de duas principais abordagens: estatística e espectral.

A abordagem estatística consiste no uso de modelos estatísticos do histograma de intensidade de uma imagem ou de uma porção dela. Seja  $z$  uma variável que representa uma intensidade e  $p(z_i)$ ,  $i = 0, 1, 2, \dots, L$ , onde  $L$  é o número de intensidades distintas na região investigada, o  $n$ -ésimo momento de  $z$  pode ser definido por:

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i) \quad (2.2)$$

onde  $m$  é o valor médio de  $z$ :

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (2.3)$$

A partir disto, medidas de maciez relativa podem ser extraídas, medindo a variância entre diferentes momentos estatísticos do histograma ( $z$ ,  $z + 1$ ,  $z + 2$ ...). A literatura está

repleta de exemplos desta abordagem, se utilizando de atributos como a entropia média (equação 2.4) ou uniformidade (equação 2.5).

$$e(z) = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \quad (2.4)$$

$$U(z) = \sum_{i=0}^{L-1} p^2(z_i) \quad (2.5)$$

A descrição e extração de texturas seguindo uma abordagem espectral consiste na transformação da imagem para o domínio de frequência, através da transformada de Fourier, e o posterior agrupamento das informações de frequência a fim de obter algumas métricas ou descritores das texturas. A principal desta abordagem é sua invariância à aspectos espaciais. Desta forma, padrões de variação de intensidade nas imagens podem ser visualizados e extraídos (figura 1).

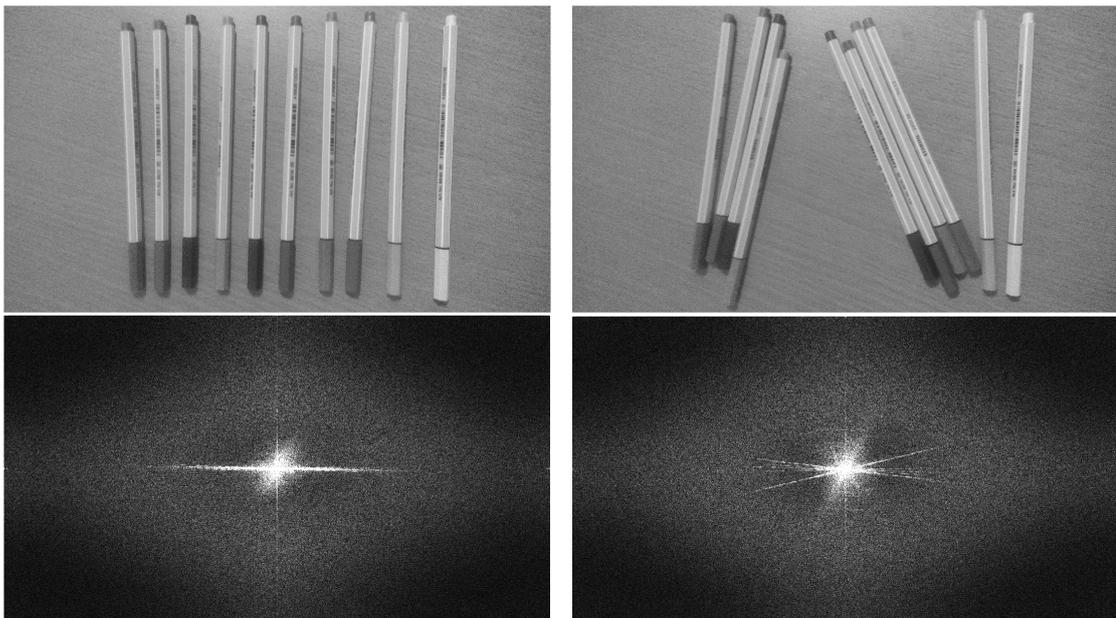


Figura 1 – Exemplo de extração de texturas utilizando transformada de Fourier. As imagens inferiores são representações gráficas do espectro de frequência de intensidade das imagens acima

Um dos modelos mais populares e descritivos de características de textura com abordagem espectral é o Local Binary Pattern (LBP), introduzido por Ojala, Pietikainen e Harwood (1994). Em primeiro lugar, este modelo divide a região examinada em células, normalmente cada célula corresponde a um pixel. Em seguida, cada célula é comparada com suas oito células vizinhas, e a diferença de intensidade (um valor binário que simboliza "maior" ou "menor") é utilizada para construir uma máscara de 8 bits, que serve como uma espécie de impressão digital daquela célula na região examinada. A junção da máscara de

todas as células compõe o que é chamado de Padrão Binário Local (LBP) da região, uma característica de alta dimensionalidade, mas bastante descritiva.

Diversas variações de LBP existem na literatura, algumas visando melhorar a descrição de certos aspectos da imagem, outras almejando a invariância de rotação ou escala da imagem. Como será visto no capítulo 3, este tipo de descritor é comumente usado em trabalhos de classificação de imagens.

## Morfologia

Segundo Nixon e Aguado (2008), parte importante da extração de características de alto nível em imagens é a busca por formas e padrões. As abordagens para extrair formas de imagens vão desde as mais simples, como a limiarização, que faz análise pixel a pixel da imagem para tentar separar o componente principal do fundo da imagem (todo o resto que não interessa à análise), passando pelo casamento de padrões, que se utiliza de uma amostra da forma desejada para realizar a busca pelo mesmo padrão na imagem, até poderosos extratores de formas geométricas como a transformada de Hough.

A transformada de Hough, introduzida por Hough (1962) é uma técnica criada para extrair formas geométricas de imagens. O objetivo desta técnica está na descrição da imagem ou região através de retas, círculos e elipses presentes na mesma. Inicialmente concebida para detectar apenas linhas em diversos ângulos, atualmente todas as implementações e variações da transformada Hough derivam do trabalho de Duda e Hart (1972), responsáveis pela criação da transformada de Hough generalizada, capaz de detectar os demais círculos e elipses. Desde então, diversas melhorias têm sido feitas para acelerar o algoritmo, bem como suportar a detecção de formas geométricas arbitrárias.

Para detectar linhas em imagens utilizando a transformada Hough, é preciso, primeiramente, compreender que em uma abordagem cartesiana, pontos colineares em uma imagem estão relacionados pela inclinação  $m$  e um ponto de interceptação  $c$ , de forma que:

$$y = mx + c \quad (2.6)$$

ou, reescrevendo a equação de forma linear homogênea:

$$Ay + Bx + 1 = 0 \quad (2.7)$$

onde  $A = -1/c$  e  $B = m/c$ .

Dado que um par de coordenadas  $(x, y)$  define uma reta no espaço com parâmetros  $(A, B)$ , podemos dizer que a equação é simétrica. Se  $(x_i, y_i)$  é um conjunto de pontos lineares que definem a reta  $(A, B)$ , temos que:

$$Ay_i + Bx_i + 1 = 0 \quad (2.8)$$

Esta equação pode ser vista como um sistema de equações e pode ser reescrita em termos de parametrização cartesiana:

$$c = -x_i m + y_i \quad (2.9)$$

Com a finalidade de encontrar linhas retas na imagem, a transformada de Hough busca satisfazer as incógnitas desta equação, utilizando também parâmetros externos para determinar tolerância de inclinação e comprimento mínimo das retas.

De forma similar, a transformada de Hough busca por círculos que satisfaçam:

$$(x - x_0)^2 + (y - y_0)^2 = r^2 \quad (2.10)$$

que, convertida para forma paramétrica, temos:

$$x = x_0 + r \cos(\theta) \quad (2.11)$$

$$y = y_0 + r \sin(\theta) \quad (2.12)$$

Diversas características de baixo e alto nível são aplicadas em regiões da imagem, ao invés da imagem completa. Isto acontece por conta da necessidade de reconhecer porções da imagem, e separá-las umas das outras. Em muitos problemas, como o de interpretação de imagens por uma máquina, é comum que algum processo de segmentação seja realizado, e as imagens sejam divididas em regiões para processamento posterior.

### 2.1.3 Segmentação de Imagens

O processo de segmentação subdivide uma imagem em suas várias regiões ou objetos. O nível em que a subdivisão é feita depende do problema a ser resolvido, ou seja, a segmentação deve parar quando os objetos de interesse de uma aplicação forem isolados. Por exemplo, na inspeção automática de uma linha de montagem de produtos eletrônicos, o interesse reside em analisar imagens dos produtos com o objetivo de determinar a presença ou ausência de anomalias específicas, como a falta de componentes ou conexões quebradas. Não há sentido em prosseguir com a segmentação além do nível de detalhes necessário para a identificação destes elementos.

Segmentação de imagens é um dos problemas mais difíceis em processamento digital de imagens (GONZALEZ; WOODS, 2002). Algoritmos de segmentação geralmente

se baseiam em uma das duas propriedades básicas de valores de intensidade dos pixels: descontinuidade e similaridade. Na primeira propriedade, a abordagem é particionar a imagem baseando-se em mudanças abruptas na intensidade dos pixels, como as bordas. A segunda categoria se baseia no particionamento de uma imagem em regiões que são similares de acordo com um critério em particular, que pode ser coloração, textura, entre outros.

Neste trabalho, conforme será descrito no capítulo 4, testamos uma série de atributos para determinar a melhor forma de segmentar as regiões das imagens aéreas da floresta amazônica, como uma das etapas da pesquisa de detecção de elementos antrópicos. Muitas vezes, técnicas de segmentação de imagens não são suficientes para isolar ou detectar componentes específicos na coleção de imagens. Trabalhos recentes têm utilizado, com frequência, algoritmos de aprendizagem de máquina para resolver este problema.

#### 2.1.4 Avaliação de segmentação

Para avaliar a qualidade de uma segmentação, em primeiro lugar é preciso aferir a validade da base de dados de segmentações manuais (*ground-truth*), demonstrando que segmentações de uma mesma imagem feitas por diferentes indivíduos são consistentes. Em um segundo momento, é preciso avaliar os algoritmos de segmentação de forma objetiva.

O problema em mensurar a consistência entre duas segmentações é que não há uma única segmentação possível para uma imagem. Duas pessoas podem segmentar uma imagem de forma significativamente diferente por terem uma percepção diferente da cena, ou ainda, por usar diferentes granularidades na segmentação, caso exemplificado na figura 2.



Figura 2 – Exemplos de uma mesma imagem segmentada de forma consistente, mas com diferentes granularidades de segmentação

Quando a diferença entre duas segmentações advém de uma diferença de percepção da cena, é esperado que o erro seja elevado e que as segmentações sejam consideradas inconsistentes. Quando a diferença é de granularidade, uma segmentação pode ser con-

siderada apenas um refinamento da outra, portanto, o erro deve ser baixo ou até zero (MARTIN et al., 2001).

Segmentação de imagens é simplesmente a divisão de pixels de uma imagem em conjuntos (segmentos). Uma possível medida de erro tem como entrada duas segmentações,  $S_1$  e  $S_2$ , e tem como saída um valor real no intervalo  $[0...1]$ , onde 0 significa ausência de erros de segmentação.

Ainda de acordo com o trabalho de Martin et al. (2001), é necessário utilizar medidas que sejam lenientes com o refinamento de granularidade entre duas segmentações de uma mesma imagem. Entende-se que se os pixels de um segmento podem ser considerados um subconjunto adequado de um outro segmento, trata-se de um refinamento e o erro deve ser zero. Se não existe relação de subconjunto entre os segmentos, as regiões devem ser consideradas inconsistentes e o erro deve ser significativo.

Se  $R(S, p_i)$  é o conjunto de pixels correspondentes a uma região na segmentação  $S$  que contém o pixel  $p_i$ , onde  $\setminus$  denota o conjunto complementar, o erro de refinamento local é definido por:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \setminus R(S_2, p_i)|}{|R(S_1, p_i)|} \quad (2.13)$$

A partir desta relação, Martin et al. (2001) definem duas métricas de erro de segmentação: Erro de Consistência Global (Global Consistency Error, ou GCE) e Erro de Consistência Local (Local Consistency Error, ou LCE):

$$GCE(S_1, S_2) = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\} \quad (2.14)$$

$$LCE(S_1, S_2) = \frac{1}{n} \sum_i \min \left\{ E(S_1, S_2, p_i), E(S_2, S_1, p_i) \right\} \quad (2.15)$$

Uma vez que  $LCE \leq GCE$  entre segmentações de uma mesma imagem, é correto afirmar que  $GCE$  é uma medida mais rígida que  $LCE$ . Além dos casos em que uma segmentação é um refinamento de outra, há ainda dois casos em que o erro de segmentação pode ser zero: quando os segmentos são compostos por apenas um pixel cada ou quando toda a imagem é composta de apenas um segmento.

## 2.2 Aprendizagem de máquina e reconhecimento de padrões

Conforme Alpaydin (2010), aprendizagem de máquina é uma área da inteligência artificial que estuda métodos computacionais, a fim de obter um determinado conhecimento específico através de experiências. A aplicação prática de aprendizado de máquina inclui o

processamento de linguagem natural, diagnósticos médicos, detecção de intrusos, entre outros. Um sistema de aprendizado tem a função de analisar informações e generalizá-las, para a extração de novos conhecimentos.

Segundo Russell e Norvig (2010), os tipos de aprendizagem podem ser classificados de acordo com o tipo de *feedback* que recebem do ambiente:

- Aprendizagem não-supervisionada: o agente aprende padrões na entrada, embora não seja fornecido nenhum *feedback* explícito. A tarefa mais comum de aprendizagem não-supervisionada é o agrupamento, ou seja, a detecção de grupos de exemplos de entrada potencialmente úteis.
- Aprendizagem por reforço: também conhecida como aprendizagem semi-supervisionada. O agente aprende a partir de uma série de reforços - recompensas ou punições.
- Aprendizagem supervisionada: o agente observa alguns exemplos de pares de entrada e saída, e aprende uma função que faz o mapeamento da entrada para a saída.

Os problemas de aprendizagem podem ainda ser divididos de acordo com o tipo de saída que demandam:

- Problemas de classificação: quando a saída esperada para o problema é uma classe ou categoria, ou seja, um valor discreto;
- Problemas de regressão: quando a saída esperada para o problema é um valor numérico, normalmente contínuo.

Um problema de classificação consiste na determinação de regras e posterior classificação desses exemplos. Este conjunto de regras é criado por um classificador, que recebe como entrada um vetor de características e oferece como saída uma classe resultante para a instância que as características descrevem, conforme pode ser visto na figura 3.



Figura 3 – Representação de classificador como uma função de bloco

Os tipos de classificadores utilizados neste trabalho serão discutidos com mais detalhes na seção 2.2.3.

Para a composição do modelo de aprendizagem, uma base de dados de treinamento é utilizada. Esta base deve possuir uma quantidade significativa e com boa representatividade

das classes envolvidas no problema. Normalmente se usa uma parte da base de dados de treinamento para validação do modelo de aprendizado (validação cruzada) ou mesmo uma base de dados diferente (base de testes ou validação), para que indicadores de qualidade do modelo possam ser avaliados. A seção 2.2.4 discorre sobre os métodos de avaliação utilizados neste trabalho.

Técnicas de aprendizagem de máquina podem ser utilizadas para encontrar padrões em diversos domínios, inclusive em imagens. É neste ponto que a linha de pesquisa de aprendizagem de máquina, advinda da área de inteligência artificial, se encontra com a linha de pesquisa de reconhecimento de padrões, advinda da área de processamento de sinais. Segundo Jain (1989), o fluxo padrão para soluções de reconhecimento de padrões consiste em três etapas:

1. Filtragem e pré-processamento da entrada;
2. Extração e seleção de características;
3. Classificação.

### 2.2.1 Filtragem e Pré-processamento

A etapa de filtragem e pré-processamento é responsável pela escolha e montagem da base de dados que será usada no processo de aprendizagem. A base deve conter uma quantidade significativa de exemplos de todas as classes envolvidas no problema.

Em aprendizado relacionado a imagens, essa etapa é comumente a responsável por normalizar e salientar as características desejadas nas amostras (realce de imagens, filtragem, etc). Exemplos irrelevantes, distorcidos ou repetidos também são eliminados durante a filtragem. O objetivo principal desta etapa é preparar a base de dados para as etapas seguintes.

### 2.2.2 Extração e Seleção de Características

A extração de características é feita selecionando os atributos oriundos dos dados (imagens, no trabalho em questão), a fim de encontrar as características úteis para o processo de reconhecimento. Essa etapa é crítica ao sucesso do aprendizado, uma vez que bons algoritmos de aprendizado só obtêm êxito com um bom conjunto de características relevantes ao problema.

Em projetos que envolvem classificação de imagens, uma gama de atributos pode ser extraída, e podem ser descritos pelo nível da informação que representam. Nesta etapa há uma forte contribuição da linha de pesquisa de processamento digital de imagens

(GONZALEZ; WOODS, 2002), que descreve filtragens, transformações e outras técnicas capazes de extrair informações sobre uma imagem ou pedaços dela.

Segundo Nixon e Aguado (2008), características de baixo nível como bordas, histogramas de intensidade e coloração, são úteis para o reconhecimento de padrões em imagens, assim como características de níveis mais altos, como texturas, transformadas de Hough e extração de regiões conectadas.

Em um estágio inicial de uma solução de aprendizagem de máquina, é natural que sejam utilizadas grandes quantidades de características para descrever as amostras. Com o avanço da pesquisa e análise de resultados parciais, é esperado que a dimensionalidade deste vetor diminua. Isto é desejado por dois motivos: a alta dimensionalidade do vetor de características tem impacto significativo no desempenho computacional de vários algoritmos de aprendizagem de máquina e a presença de atributos que possuam pequena contribuição para a classificação podem causar ruído no aprendizado, dificultando sua generalização. Estes dois problemas serão abordados com mais detalhes nas seções 2.2.3 e 2.2.4.

Uma técnica bastante difundida para a seleção de características em problemas de aprendizagem de máquina supervisionada é a *Correlation-based Feature Selection* (CFS), introduzida por Hall (1998). A idéia geral desta técnica é a avaliação de quais atributos são bons ou ruins para um determinado conjunto de dados supervisionados. Bons atributos possuem alta correlação com a classe da amostra, e ao mesmo tempo, baixa correlação com os demais atributos da mesma amostra. Formalmente, um atributo  $V_i$  é considerado relevante se, e somente se, existe um  $v_i$  e  $c$  para os quais  $p(V_i = v_i) > 0$  de tal maneira que:

$$p(C = c|V_i = v_i) \neq p(C = c) \quad (2.16)$$

onde  $p(C = c|V_i = v_i)$  é a probabilidade condicional de que a classe  $C$  de uma amostra é  $c$  dado que o valor do atributo  $V_i$  é  $v_i$ .

O produto desta etapa é a representação de cada exemplo da base de dados em um vetor de características, de forma que possa ser usado por um ou mais classificadores em uma etapa posterior.

### 2.2.3 Classificação

Nesta etapa, um modelo de aprendizado é gerado. Pode-se também usar múltiplos classificadores, ao invés de apenas um. Esta abordagem é chamada de sistemas com múltiplos classificadores (do inglês *multiple classifier systems*), também chamados de *ensemble* de classificadores, os quais podem ser compostos por classificadores do mesmo tipo ou de diferentes algoritmos de classificação.

Existe uma grande variedade de algoritmos de aprendizagem de máquina propostos

na literatura. Alguns dos mais utilizados são: classificadores estatísticos, redes neurais artificiais, árvores de decisão, máquinas de vetores de suporte (SVM), k vizinhos mais próximos (KNN), etc. A seguir, descrevemos sucintamente os algoritmos de aprendizagem de máquina utilizados neste trabalho.

### Árvores de decisão

Amplamente utilizadas em algoritmos de classificação, as árvores de decisão são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados. As árvores de decisão consistem de nodos que representam os atributos; de arcos, provenientes destes nodos e que recebem os valores possíveis para estes atributos; e de nodos folha, que representam as diferentes classes de um conjunto de treinamento. Classificação, neste caso, é a construção de uma estrutura de árvore, que pode ser usada para classificar corretamente todos os objetos do conjunto de dados da entrada.

A partir de uma árvore de decisão é possível derivar regras. As regras são escritas considerando o trajeto do nodo raiz até uma folha da árvore. Estes dois métodos são geralmente utilizados em conjunto. Devido ao fato das árvores de decisão tenderem a crescer muito, de acordo com algumas aplicações, elas são muitas vezes substituídas pelas regras. Isto acontece em virtude das regras poderem ser facilmente modularizadas. Uma regra pode ser compreendida sem que haja a necessidade de se referenciar outras regras.

Uma árvore de decisão tem a função de particionar recursivamente um conjunto de treinamento, até que cada subconjunto obtido deste particionamento contenha casos de uma única classe. Para atingir esta meta, a técnica de árvores de decisão examina e compara a distribuição de classes durante a construção da árvore. O resultado obtido, após a construção de uma árvore de decisão, são dados organizados de maneira compacta, que são utilizados para classificar novos casos. A árvore de decisão não presume nenhum modelo estatístico a priori, sendo a divisão do espaço de atributos feita de acordo com as amostras provenientes do treinamento.

### KNN

O algoritmo KNN (K-Nearest Neighbours, ou K vizinhos mais próximos) (COVER; HART, 1967) é um método de classificação baseado na proximidade de amostras de treino no espaço de características. É considerado um dos mais simples algoritmos de aprendizagem de máquina.

O processo de treinamento para esse algoritmo consiste em armazenar o vetor de característica e rótulos (classes) de cada amostra de treinamento em um espaço n-dimensional, onde n é o número de características de cada amostra. No processo de

classificação de amostras não-rotuladas, a amostra é simplesmente projetada no espaço e é classificada de acordo com as  $k$  amostras mais próximas.

Quando  $k = 1$ , a amostra é simplesmente classificada de acordo com o rótulo de seu vizinho mais próximo no espaço de características. Quando  $k$  é maior que 1, a classificação se dá por um esquema de votação, onde a classe com as amostras vizinhas mais numerosas é considerada como a classe da amostra. Por essa razão, em problemas bi-classe como o apresentado nesta dissertação, deve ser utilizado um valor ímpar para  $k$ , com a finalidade de evitar empates. Em problemas multi-classe, ou seja, com mais de duas classes possíveis, empates podem acontecer mesmo considerando um número ímpar de vizinhos, de maneira que uma forma de desempate deve ser definida na implementação do algoritmo.

Diversas formas de calcular a distância entre duas amostras  $a$  e  $b$  em um espaço  $n$ -dimensional podem ser utilizadas no  $k$ NN, dentre elas a distância Euclidiana, representada pela fórmula 2.17, onde  $n$  é o número de dimensões.

$$d(a, b) = \|a - b\| = \sqrt{(a - b) * (a - b)} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2.17)$$

## Support Vector Machines

As máquinas de vetores de suporte, ou SVM (*Support Vector Machines*) são classificadores baseados na teoria de aprendizagem estatística proposta por (VAPNIK; CORTES, 1995). A teoria é baseada em uma forte fundamentação matemática para estimação de dependências e previsão do aprendizado a partir de conjuntos de dados finitos.

Um modelo SVM é a representação das amostras como pontos em um espaço  $n$ -dimensional (onde  $n$  é o tamanho do vetor de características) de tal forma que as amostras de diferentes classes sejam divididas por um hiperplano de separação que maximiza a distância entre essas classes. Isso se deve ao fato de que o hiperplano de separação possui a maior distância possível às amostras mais próximas entre as classes, o que ajuda a diminuir o erro de generalização do classificador.

A principal vantagem do classificador SVM é seu bom desempenho em conjuntos de dados que possuem muitos atributos, mesmo quando há poucas amostras de treino. No entanto, suas desvantagens são a baixa velocidade e alto consumo de recursos durante as fases de treinamento, assim como a complexidade de parametrização de suas funções *kernel*. As amostras desconhecidas são classificadas ao serem posicionadas no espaço de características e avaliadas em que lado da superfície de separação elas se encontram, como mostrado na figura 4.

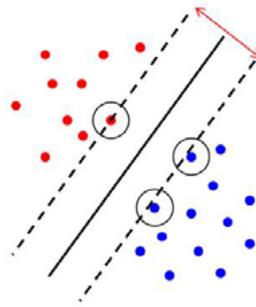


Figura 4 – Amostras em um espaço bidimensional (pontos coloridos) separadas por um hiperplano apoiado por vetores de suporte (pontos circulados), maximizando a distância entre as amostras mais próximas do limiar das classes.

### Naive Bayes

Há um grupo de algoritmos de aprendizado de máquina que se baseiam em raciocínio probabilístico, especialmente sob o viés bayseano. De acordo com Mitchell (1997), o raciocínio bayesiano é baseado no pressuposto de que as quantidades de interesse são governadas pela distribuição das probabilidade e que decisões ótimas podem ser tomadas a partir da análise destas probabilidades em conjunto com os dados observados.

O teorema de Bayes estabelece a relação entre uma probabilidade condicional e sua probabilidade inversa. A equação 2.18, conhecida como regra de Bayes, permite calcular a probabilidade de um evento ( $A$ ) dado que outro evento aconteceu ( $B$ ).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.18)$$

A probabilidade *a priori*, ou seja, a probabilidade isolada de que um evento  $A$  aconteça é expressa por  $P(A)$ . A probabilidade *a posteriori* de um evento  $A$  dado que ocorreu também um evento  $B$  é expressa por  $P(A|B)$ .

Dentre as técnicas de inspiração probabilística, *Naive Bayes* se destaca por sua simplicidade. Apresentada por John e Langley (1995), esta família de algoritmos presume uma forte independência entre as variáveis do vetor de características, por isso é chamado de Bayes Ingênuo (do inglês *Naive Bayes*).

Segundo Russell e Norvig (2010), os valores de máxima probabilidade de cada classe de acordo com as variáveis do vetor de características são encontrados. Uma vez que um modelo é gerado desta maneira, as futuras amostras podem ser classificadas de acordo com as mesmas probabilidades. Dado que o vetor de características das amostras é denotado por  $x_1, x_2, \dots, x_n$ , a probabilidade de cada classe  $C$  é expressa pela equação 2.19. A família de algoritmos *Naive Bayes* tende a ser bastante robusta em face a dados ruidosos ou ausentes.

$$P(C|x_1, x_2, \dots, x_n) = P(C) \prod_i P(x_i|C) \quad (2.19)$$

## Random Forest

O algoritmo *Random Forest*, introduzido por Breiman (2001), é um *ensemble* de classificadores de árvore de decisão. Na fase de treinamento, um número determinado de árvores é gerado utilizando um subconjunto pseudo-aleatório do vetor de características completo utilizado no problema. Tanto o número de árvores quanto a quantidade de atributos utilizados por cada árvore pode ser determinado pelo usuário do algoritmo.

Uma vez que a base de treinamento foi utilizada para gerar os modelos das árvores de decisão, estes modelos podem ser utilizados para classificar novas amostras do problema. Preferencialmente, todas as árvores geradas são envolvidas na classificação, cada uma chegando à sua própria conclusão sobre a classe da amostra apresentada. Cada árvore tem um “voto”, que é contabilizado para a definição da classe mais votada, que é então escolhida como a classe da amostra. Esta técnica de soma de votos é conhecida como *polling*.

A principal vantagem deste algoritmo é a eliminação de *overfitting*, problema bastante comum quando se utiliza árvores de decisão de forma tradicional. Uma exposição mais detalhada acerca de *overfitting* é feita na seção 2.2.4. O trabalho de Breiman (2001) ainda afirma que a taxa de erro de um modelo de aprendizado criado com *Random Forest* está relacionada a dois fatores: a correlação entre quaisquer duas árvores geradas no modelo e a “força” de cada árvore gerada, ou seja, o quão precisa cada árvore é em relação ao modelo geral. À medida que a correlação entre árvores cresce, também cresce a taxa de erro do modelo final. Quanto menor for a taxa de erro de uma árvore individual, mais “forte” é considerado o classificador, e por consequência, menor é a taxa de erro do modelo.

Reduzir o número de variáveis do vetor de características a ser utilizado em cada árvore reduz a correlação entre árvores e também reduz a “força” da árvore. Incrementar este número de variáveis incrementa a correlação e a “força” da árvore. A parametrização do algoritmo deve se preocupar em encontrar o número de variáveis do vetor de características que melhor minimize a correlação e maximize a “força” das árvores do modelo.

Os algoritmos descritos até então podem ser utilizados em qualquer problema de classificação binária ou multi-classe. Entretanto, há um grupo de algoritmos em aprendizagem de máquina que tem como objetivo a detecção de anomalias a partir de um modelo formado por apenas uma classe. A próxima seção trata desta classe de algoritmos.

## Classificadores unários

Algoritmos conhecidos como classificadores unários são utilizados, normalmente, em problemas onde há apenas uma classe conhecida, e há interesse em se encontrar anomalias ou novidades, ou seja, amostras que não pertencem à classe conhecida.

Neste tipo de problema, a base é normalmente composta por uma única classe (conhecida como classe alvo, ou *target*) e o modelo gerado é supostamente capaz de determinar se novas amostras do problema pertencem à classe alvo ou se são anomalias ou novidades (*outliers*). Estes algoritmos também são chamados de detectores de novidades (*novelty detection*) ou detectores de anomalias (*outlier detection*).

Apesar de serem concebidos para problemas de classificação de apenas uma classe, algoritmos de detecção de anomalias podem ser utilizados para resolução de problemas com múltiplas classes. A estratégia consiste em treinar um classificador deste tipo para cada classe do problema. Diversas técnicas para apuração dos resultados podem ser utilizadas, tais como votação (*voting* ou *polling*), onde cada classificador tem um voto que é computado com um peso e levado em consideração na classificação final; ou bagging (abreviação do termo inglês *Bootstrap aggregating*), que se utiliza da combinação do resultado de classificação de conjuntos de dados de treinamento selecionados aleatoriamente, com o objetivo de reduzir a variância e o *overfitting*, discutidos com mais detalhes na seção 2.2.4.

Uma variante do SVM denominada OC-SVM (acrônimo em inglês para *One Class Support Vector Machines*) foi introduzida por Schölkopf et al. (1999) com a finalidade de adaptar o robusto algoritmo de Vapnik e Cortes (1995) para problemas de detecção de anomalias. Este é, na verdade, uma expansão da aplicação da técnica de SVM para dados não-rotulados.

A ideia central é que uma função *kernel* de base radial é utilizada para comportar as amostras da classe alvo, de forma a medir a possibilidade de anomalias futuras de acordo com a distância destas mesmas amostras em relação à hipersfera de separação descrita pela função *kernel*. Em um problema com múltiplas classes e classificadores OC-SVM, costuma-se usar uma estratégia de votação com pesos para determinação da classe da amostra.

Outros métodos tradicionais em problemas multi-classe também podem ser usados como classificadores para detecção de anomalias. Variações de KNN como o OCNN (*One-Class Nearest Neighbour*) ou de árvores de decisão como o REPTree são comumente utilizados nesta abordagem, utilizando diversas estratégias para determinação final das classes das novas amostras.

Há ainda na literatura trabalhos que descrevem modelos para utilização de diversos algoritmos unários diferentes em um mesmo problema, definindo também estratégias para a determinação da classificação resultante da miríade de classificadores. Isto possibilita

um grau de liberdade elevado na escolha desses algoritmos e estratégias para problemas multi-classe.

Tanto durante o desenvolvimento de uma solução, quanto após sua execução em ambiente de produção, é preciso aferir e quantificar o desempenho da técnica desenvolvida ou utilizada, conforme discutido na próxima seção.

## 2.2.4 Avaliação de aprendizagem

Avaliar o desempenho de uma técnica de aprendizagem de máquina é útil para determinar a qualidade do modelo criado, aferir se o modelo continua adequado e inspecionar se os atributos escolhidos são relevantes para a classificação das amostras.

Comumente, o percentual de acerto obtido na classificação das amostras é um importante parâmetro para medir o desempenho do modelo. Este parâmetro é conhecido como acurácia ou taxa de reconhecimento. O oposto da acurácia é conhecido como taxa de erro.

De grande importância também é a composição da matriz de confusão (tabela 1). Nela pode-se avaliar como um modelo está se comportando em termos de falsos positivos (um exemplo é classificado como pertencente à classe C, mas não é) e falsos negativos (um exemplo é atribuído a outra classe, mas deveria ser da classe C). A principal função desta matriz é dar possibilidade de pensar sobre o custo dos erros, ou seja, mesmo que a taxa de acerto para o problema seja alta, uma ou mais classes do problema pode ter uma taxa de acerto bem abaixo do esperado.

<b>Resultado obtido</b>			<b>Resultado esperado</b>
Classe A	Classe B		
Verdadeiro positivo	Falso negativo	Classe A	
Falso positivo	Verdadeiro negativo	Classe B	

Tabela 1 – Modelo de matriz de confusão.

Alguns valores podem ser obtidos através desta matriz. A própria acurácia do modelo pode ser obtida com a equação 2.20, onde TP representa o número de verdadeiros positivos, FP representa os falsos positivos, FN representa os falsos negativos e TN representa os verdadeiros negativos.

$$Acurácia = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.20)$$

Ainda é possível obter a precisão e a revocação. A precisão é o número de elementos relevantes recuperados dividido pelo número total de elementos recuperados (equação 2.21) enquanto a revocação é definida como o número de elementos relevantes recuperados

dividido pelo número total de elementos relevantes existentes, que deveriam ter sido recuperados (equação 2.22).

$$Precisão = \frac{TP}{TP + FP} \quad (2.21)$$

$$Revocação = \frac{TP}{TP + FN} \quad (2.22)$$

Como forma de medir o equilíbrio entre precisão e revocação, a média harmônica entre estas duas medidas é calculada. Esta medida é chamada de  $F_1$  ( $F_1$  score, em inglês), e é obtida a partir da equação 2.23.

$$F_1 = 2 \cdot \frac{precisão \cdot revocação}{precisão + revocação} \quad (2.23)$$

Embora a maximização das métricas apresentadas seja desejável para o ajuste de um algoritmo de aprendizado, é necessário evitar o sobreajuste (*overfitting*) que pode ser gerado a partir disto. O *overfitting* é o termo utilizado quando o modelo criado se ajusta em excesso às amostras de treinamento, tendo altas taxas de acurácia para esta base, mas falhando em classificar com boa taxa de acurácia as demais amostras de teste ou amostras reais. Em resumo, o problema é ter um modelo que possui bom desempenho na etapa de treinamento mas não é uma boa representação das amostras reais.

Um método bastante utilizado para medir a generalização de um modelo de aprendizado é a validação cruzada. O conceito principal deste método é o particionamento do conjunto de amostras em um número de subconjuntos mutuamente exclusivos, dividindo estes subconjuntos entre base de treinamento e teste.

A técnica de validação cruzada conhecida como *k-fold* divide a base em  $k$  subconjuntos de mesmo tamanho, escolhe um dos subconjuntos como base de testes e utiliza os demais subconjuntos como bases de treinamento do modelo. A razão para esta distribuição é a premissa de que é preciso de mais amostras para treinamento que para testes, a fim de criar um modelo robusto. O procedimento é repetido  $k$  vezes, de forma que todos os subconjuntos possam ser utilizados para construir uma versão de modelo de aprendizado e ser validado pelos demais subconjuntos.

Uma outra avaliação importante a ser feita é a da relação entre a taxa de acerto (verdadeiros positivos) e a taxa de falsos positivos. Para tal análise, pode ser feito o estudo da curva ROC (do inglês *Receiver Operating Characteristic*), um gráfico bidimensional que exhibe no eixo y o percentual de verdadeiros positivos na amostra e, no eixo x, o percentual de falsos positivos da mesma amostra (figura 5).

Originalmente utilizada em detecção de sinais para mostrar a relação entre sensibilidade (taxa de acerto) e o inverso da especificidade (alarme falso) em canais com ruído,

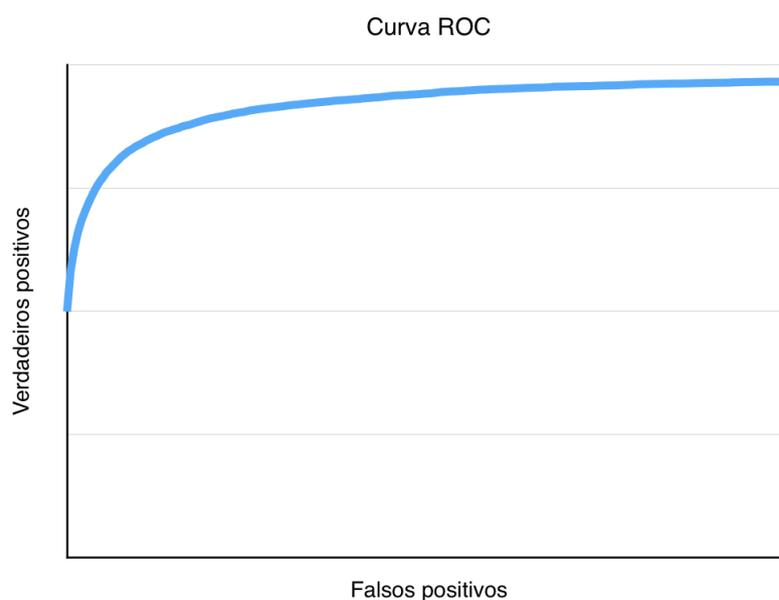


Figura 5 – Exemplo de gráfico para análise de curva ROC.

pode-se utilizar este gráfico na avaliação de modelos de aprendizagem de máquina. Em geral, o cálculo da área abaixo da linha do gráfico é realizado para determinar a relação entre verdadeiros positivos e falsos negativos para um conjunto de amostras. Quanto maior a área abaixo da curva, melhor é considerado o aprendizado sob esta avaliação.

Uma vez que a fundamentação teórica necessária foi coberta, é necessário fazer um levantamento na literatura pelos trabalhos considerados estado da arte em suas respectivas áreas. O capítulo a seguir trata deste levantamento.

## 3 Trabalhos Relacionados

A floresta amazônica é tema recorrente em diversas áreas de pesquisa, incluindo a área de sensoriamento remoto. Trabalhos vindos desta área costumam utilizar técnicas computacionais que pertencem ao estado da arte para analisar imagens que cobrem grandes áreas, sejam essas de espectro visível ou mesmo de RADAR ou LIDAR, e inferir as mais diversas informações sobre as regiões estudadas.

Trabalhos como os de Filho, Nascimento e Batista (1998), Vasconcelos e Novo (2001), Lu et al. (2012) e Ferreira (2012) se utilizam de imagens com 6 bandas multiespectrais e 1 banda termal do satélite LANDSAT da região amazônica para determinar a cobertura de terreno das áreas inspecionadas. Estes trabalhos têm foco nos tipos de vegetação que cobrem o terreno, classes tais como floresta, capoeira, vegetação secundária, ou até mesmo espécies e grupos de plantas. Há trabalhos mais elaborados, como o de Espírito-Santo e Shimabukuro (2005), que se utiliza das mesmas imagens do satélite LANDSAT, mas analisa amostras multitemporais, ou seja, imagens da mesma região em diversos sobrevoos do satélite, com o objetivo de conseguir informações sobre mudanças na cobertura vegetal do terreno.

Há ainda trabalhos como o de Azevedo et al. (2014), que utiliza como fonte de dados o satélite COSMO-SkyMed, que disponibiliza imagens duais multitemporais da região amazônica. Estudos como o de Latorre et al. (2007) vão além e mesclam os dados de diversas fontes diferentes, utilizando técnicas de fusão de imagens.

Apesar dos resultados alcançados, os trabalhos sobre a classificação de imagens da floresta amazônica geralmente avaliam a cobertura vegetal da região, sem qualquer menção a elementos antrópicos, utilizando para isso imagens termais ou multiespectrais de diversos satélites. Dada a natureza relativamente lenta das mudanças de cobertura vegetal de terreno e o tamanho das áreas estudadas, é compreensível que satélites sejam utilizados.

Para o trabalho aqui apresentado, no entanto, os objetos de interesse mudam rapidamente, e a grande cobertura de área proporcionada por imagens de satélite deve dar lugar a imagens feitas por veículos aéreos, com maior proximidade e disponibilizadas com mais rapidez em uma missão real de patrulhamento ambiental. A base de imagens disponível para o trabalho é composta exclusivamente por imagens de espectro visível, por isso, muitas das técnicas e conclusões dos trabalhos citados acima não são prontamente transferíveis para este trabalho.

Os trabalhos relacionados apresentados nesta seção foram divididos primeiramente por assunto: de início, nos concentramos nos trabalhos publicados em segmentação de

imagens que são considerados estado da arte, uma vez que esta é o primeiro desafio técnico a ser enfrentado pelo trabalho proposto. Depois, fazemos o levantamento de trabalhos relevantes sobre classificação de imagens aéreas. Em uma etapa posterior, investigamos trabalhos relacionados à área de detecção de anomalias.

As seções a seguir apresentam os trabalhos relacionados em ordem própria. A seção 3.1 apresenta os trabalhos sobre segmentação de imagens de acordo com a ordem apresentada no trabalho de comparação realizado por Arbelaez et al. (2011), explanado com mais detalhes em seguida. As seções 3.2 e 3.3 apresentam os trabalhos relacionados sobre classificação de imagens aéreas e detecção de anomalias, ambas em ordem cronológica de publicação.

Cada artigo é descrito com suas características, vantagens e desvantagens. No fim de cada seção, é apresentada uma tabela de comparação entre os trabalhos levantados.

### 3.1 Segmentação de imagens

Os trabalhos relacionados nesta seção foram escolhidos por terem obtido os maiores índices de precisão e revocação em uma base de dados de imagens naturais bastante conhecida pela comunidade de processamento de imagens, *Berkeley Segmentation Data Set and Benchmarks 500*, também conhecida como BSDS500. Uma comparação entre os mais promissores algoritmos de segmentação de imagens foi realizada por Arbelaez et al. (2011), consistindo na segmentação das imagens presentes no BSDS500 manualmente por seres humanos, e depois segmentadas pelos diversos algoritmos testados. A precisão e revocação de cada algoritmo são obtidas através de comparação com a segmentação manual.

Para reduzir ainda mais o número de algoritmos a serem inspecionados para este trabalho, a revisão de Yuan, Gleason e Cheriadat (2013) compara os mesmos algoritmos que Arbelaez et al. (2011) investigaram, mas dessa vez em uma base de imagens aéreas. A metodologia é a mesma usada no BSDS500, mas com uma base especializada. Os algoritmos com os melhores resultados foram selecionados para serem utilizados na construção deste trabalho de pesquisa, e coincidem com os resultados do trabalho de Arbelaez et al. (2011). Os seis algoritmos selecionados serão discutidos a seguir.

O trabalho de Deng e Manjunath (2001) trata de um algoritmo chamado JSEG (acrônimo para *J Segmentation*), que obtém a segmentação da imagem em duas etapas. A primeira é a quantização das cores da imagem em um número bem menor que o número original, cerca de 16 milhões, em uma imagem de 24 bits, para um número entre 10 e 20 cores na imagem resultante. Esta etapa é responsável por criar uma espécie de mapa de classes, baseada nas cores quantizadas e ignorando a distribuição espacial destas mesmas cores (figura 6).

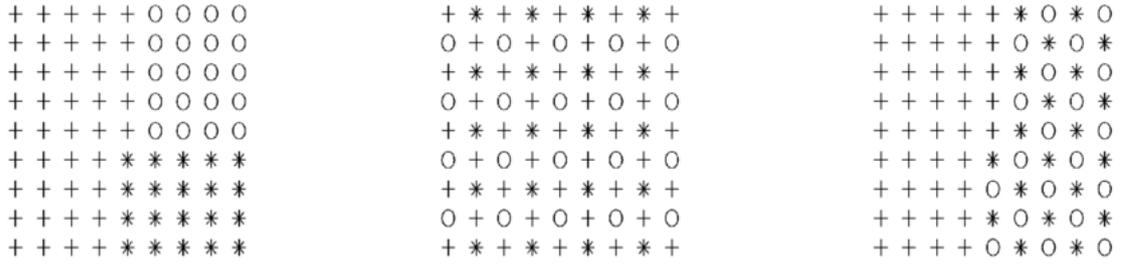


Figura 6 – Exemplos de mapas de classes do algoritmo JSEG, criado a partir da quantização de cores. Fonte: (DENG; MANJUNATH, 2001).

A segunda etapa consiste na segmentação espacial do mapa de classes sem considerar a similaridade cromática dos pixels. Para que isto seja possível, primeiramente é encontrado o valor para uma variável  $J$ . Seja  $Z$  o conjunto de todos os  $N$  pontos do mapa de classes, temos que  $z = (x, y)$ ,  $z \in Z$  e  $m$  seja a média, dada por

$$m = \frac{1}{N} \sum_{z \in Z} z \quad (3.1)$$

Supondo que  $Z$  é dividido entre  $C$  classes,  $Z_i$ ,  $i = 1, \dots, C$ . A média  $m_i$  dos  $N_i$  pontos da classe  $Z_i$  pode ser definida por

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} z \quad (3.2)$$

Podemos definir  $S_T$  como:

$$S_T = \sum_{z \in Z} \|z - m\|^2 \quad (3.3)$$

e  $S_W$ , a variância total de pontos pertencentes à mesma classe, como:

$$S_W = \sum_{i=1}^C \sum_{z \in Z} \|z - m\|^2 \quad (3.4)$$

Podemos, então, calcular  $J$  por

$$J = \frac{S_T - S_W}{S_W} \quad (3.5)$$

Com isso, temos valores de  $J$  maiores para imagens com regiões de cores mais homogêneas. Quando a distribuição de classes é mais homoganeamente distribuída pela imagem,  $J$  assume um valor baixo. Posteriormente, um método de crescimento de região para segmentar a imagem com base no valor  $J$  é aplicado, determinando as regiões finais da imagem.

O algoritmo ainda permite que o utilizador especifique o tamanho da janela para computar o valor de  $J$ , o que torna o método bastante flexível para imagens de naturezas diferentes. Em imagens aéreas de escala considerável, como as utilizadas neste trabalho, pode-se usar uma janela diminuta, já que os detalhes importantes podem ser bem pequenos. Para diminuir super-segmentação, os segmentos encontrados na segunda etapa são fundidos de acordo com seus histogramas coloridos.

A abordagem utilizando mean-shift (ou mudança de média, numa tradução literal) desenvolvida por Comaniciu e Meer (2002) oferece uma ferramenta interessante para resolver o problema de segmentação de imagens. Em um primeiro momento, uma filtragem por mean-shift para suavização da imagem é aplicada e os dados desta imagem pós-filtrada são projetados em um domínio  $d$ -dimensional que considera informações espaciais e cromáticas. Estas informações são agrupadas utilizando um critério de proximidade. Nesta etapa, as amostras cuja distância no domínio espacial estão abaixo de  $h_s$ , e no domínio de cores estão abaixo de  $h_r$ , serão agrupados em um mesmo *cluster*.

Após a convergência do processo de agrupamento, os *clusters* menores que um tamanho mínimo estipulado são absorvidos pelo *cluster* adjacente com maior similaridade, seguindo os mesmos critérios do processo de agrupamento original. Por fim, os segmentos são delimitados utilizando as bordas dos *clusters* encontrados. Parâmetros como distância espacial, distância de cores e o tamanho mínimo dos *clusters* podem ser determinados pelo usuário, para adequar o algoritmo ao problema em questão. A figura 7 exemplifica o processo de segmentação utilizando o algoritmo de mean-shift para uma imagem hipotética em tons de cinza.

O algoritmo *Multi-resolution Region Merging Segmentation* (MSEG), descrito por Felzenszwalb e Huttenlocher (2004), é amplamente usado pela comunidade de sensoriamento remoto. Baseado em grafos, este algoritmo tem preocupação com a previsibilidade da segmentação e com o seu custo computacional, da ordem de  $O(n \log n)$ .

O MSEG representa o problema de segmentação como um grafo  $G = (V, E)$  não direcionado, onde cada vértice  $v \in V$  corresponde a um pixel da imagem a ser segmentada, e cada aresta  $(v_i, v_j) \in E$  conecta dois vértices  $v_i$  e  $v_j$ . As arestas possuem um peso  $w(v_i, v_j)$ , um valor não-negativo que representa a diferença entre os vértices  $v_i$  e  $v_j$ . Nesta abordagem, uma segmentação  $S$  é a partição de  $V$  em regiões de tal forma que cada região  $C \in S$  corresponde a um componente conectado no grafo  $G' = (V, E')$ , onde  $E' \subseteq E$ . O conceito principal gira em torno dos pesos entre os vértices de uma mesma região, que devem ser baixos, enquanto os pesos altos devem ser utilizados para encontrar bordas entre as regiões.

O aumento na heterogeneidade no momento da junção de um par de segmentos é computado como uma soma ponderada de medidas de coloração e morfologia. O procedimento de junção é realizado iterativamente, e junta os pares de segmento que

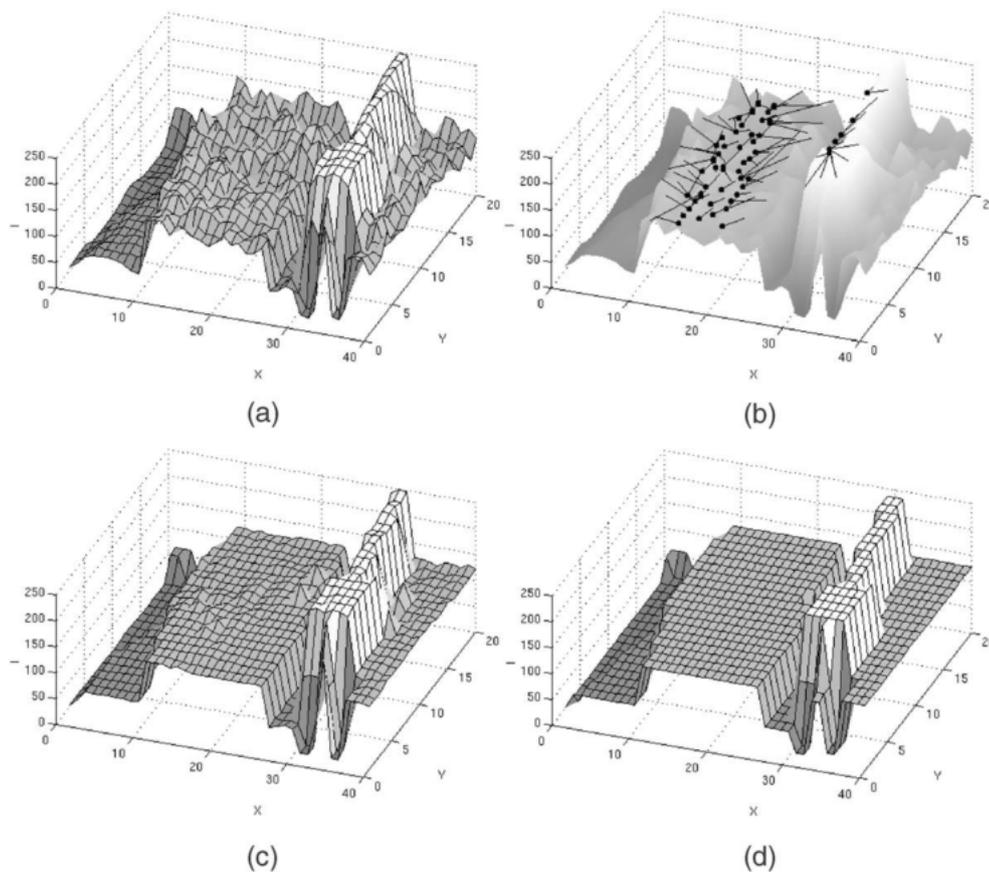


Figura 7 – Exemplo de segmentação mean-shift em uma imagem em tons de cinza. (a) Imagem de entrada. (b) Filtragem mean-shift para suavização. (c) Resultado do agrupamento sob parâmetros  $h_s$  e  $h_r$ . (d) Resultado da segmentação. Fonte: (COMANICIU; MEER, 2002).

resultam no menor aumento de heterogeneidade possível, até que a soma exceda um limiar, que pode ser parametrizado pelo utilizador. Um parâmetro é utilizado para definir a escala observada na segmentação, o que influi no tamanho e no número dos segmentos resultantes.

A principal vantagem deste método reside no desempenho computacional, com custo próximo ao linear em função do número de pixels da imagem. Os autores do algoritmo sugerem que ele seja perfeitamente adequado para o processamento de imagens de grandes dimensões, ou mesmo segmentação de vídeo em tempo real.

O algoritmo de segmentação baseada em crescimento/junção de regiões *Statistical Region Merging* (SRM), publicado por Nock e Nielsen (2004), utiliza um procedimento simples de junção acompanhado por uma operação de ordenação para segmentar imagens com eficiência. A ordenação da junção de regiões é realizada através de um algoritmo guloso, que percorre sempre o menor custo de caminhamento entre pixels de 4-vizinhança (pixels vizinhos ao norte, sul, leste e oeste, ignorando as diagonais). O custo do caminhamento é

obtido através de uma função objetiva, definida por

$$f_a(p, p') = |p_a - p'_a| \quad (3.6)$$

onde  $f_a(p, p')$  é a distância máxima entre os canais de cores dos pixels  $p$  e  $p'$ . Esta simples função, segundo o artigo de Nock e Nielsen (2004), não representa um problema, pois evidências experimentais mostram que o ganho ao usar uma função mais objetiva, como os kernels de convolução, não melhoram significativamente os resultados.

Duas regiões da imagem são unidas se os valores médios dos pixels das duas regiões estão mais próximos que um limiar previamente definido. A coesão da segmentação pode ser controlada por um parâmetro  $Q$  definido pelo usuário. Segundo os autores, este algoritmo sofre de um problema descrito como sobrejunção (do inglês *overmerging*), ou seja, ele tende a juntar duas regiões que deveriam ser separadas. O algoritmo, em compensação, parece não sofrer de subjunção, quando uma região é desnecessariamente dividida em diversos segmentos.

O trabalho de Yuan e Wang (2013) apresenta um algoritmo chamado *Factorisation-based segmentation* (FSEG). O FSEG primeiramente computa o histograma espectral para cada pixel da imagem, que é um amalgama de diversas respostas aos filtros em uma janela local. A saída deste procedimento é uma matriz  $M \times N$ , onde  $M$  é o tamanho do vetor de características extraído de cada pixel e  $N$  é o número de pixels na imagem.

Cada vetor de características em cada pixel é tratado como uma combinação linear de características, e que possuem a capacidade de identificar bordas entre os segmentos da imagem. Dados que  $Y$ , representação de uma imagem  $M \times N$  pixels e  $L$  o número de características representativas, podemos expressar a imagem  $Y$  por:

$$Y = Z\beta + \varepsilon \quad (3.7)$$

onde  $Z$  é uma matriz  $M \times L$  cujas colunas são características representativas,  $\beta$  é uma matriz  $L \times N$  cujas colunas são vetores de peso e  $\varepsilon$  é um fator de erro. A matriz  $\beta$  pode ser encontrada através da estimativa de quadrados mínimos:

$$\beta = (Z^T Z)^{-1} Z^T Y \quad (3.8)$$

A segmentação é obtida ao examinar o vetor normalizado de  $\beta$  ( $\hat{\beta}$ ), quando cada pixel é dito pertencente à região com maior respectivo peso. Para uma segmentação completamente automática,  $Z$  e  $\beta$  são desconhecidas, e o objetivo do algoritmo é fatorizar a matriz  $Y$  para estimar essas duas matrizes desconhecidas.

O algoritmo FSEG utiliza decomposição de valores singulares e fatoração de matrizes não-negativas para aumentar a eficiência computacional da segmentação, o que é atrativo quando o tempo de processamento é um fator importante.

O algoritmo gPb-owt-ucm (*Oriented Watershed Transform Ultrametric Contour Maps with globalPb*) introduzido por Arbelaez et al. (2011) realiza segmentação em várias etapas. Primeiramente, uma versão multi-escala de um detector  $Pb$  é utilizada para identificar as prováveis bordas dos segmentos da imagem. O detector  $Pb$ , originalmente introduzido por Martin, Fowlkes e Malik (2004), define a função  $Pb(x, y, \theta)$ , responsável por prever a probabilidade a posteriori de uma borda de orientação  $\theta$  para um pixel  $(x, y)$ , através da medição de diferença local de brilho, coloração e textura.

A adaptação da função  $Pb$  para um contexto multi-escalar ( $mPb$ ), realizada por Arbelaez et al. (2011), consiste na execução de  $Pb$  em três escalas diferentes,  $\sigma/2$ ,  $\sigma$  e  $2\sigma$ , para cada canal de brilho, cor e textura. Os autores utilizam  $\sigma = 5$  pixels para o canal de brilho e  $\sigma = 10$  pixels para os canais de cor e textura, considerando os resultados satisfatórios.

Uma sucessão de filtros gaussianos aplicados à imagem resultante do  $mPb$  e o posterior cálculo dos autovetores das informações de orientação dos segmentos é responsável por normalizar e globalizar ( $gPb$ ) a segmentação inicial da imagem. Após este ponto, o algoritmo já é capaz de determinar contornos abertos na imagem em questão.

Um processo de fechamento de contornos é realizado. Para tal, uma transformada de *watershed* orientada (do inglês *oriented watershed transform*, ou OWT) é utilizada. Esta etapa é responsável pela produção de uma segmentação hierárquica da imagem em nível primário, o que possivelmente gera uma supersegmentação, com vários segmentos apontados pelo algoritmo pertencendo ao mesmo segmento real.

Por último, um mapa ultramétrico de contornos (do inglês *Ultrametric Contour Map*, ou UCM) é utilizado para prover a segmentação final da imagem. Construído por um algoritmo guloso baseado em grafos, a técnica ordena as arestas por similaridade e mescla as regiões mais similares. O resultado desta etapa é uma árvore onde a raiz é a imagem completa e cada nó é uma região. Isto é conveniente, pois cada nível  $k$  da árvore representa uma escala  $k$  de segmentação, o que facilita a recuperação da segmentação em qualquer um dos valores de  $k$  presentes. Esta é uma das principais vantagens do algoritmo gPB-owt-ucm, a possibilidade de ajustar a escala de segmentação, portanto o usuário pode escolher uma escala que mais se adequa ao tipo de imagem do problema, evitando segmentação excessiva em imagens com muitos detalhes, por exemplo.

A tabela 2 exhibe os resultados encontrados por Yuan, Gleason e Cheriadat (2013), juntamente com uma descrição das características das diversas técnicas de segmentação de imagens levantadas neste trabalho.

Técnica	Características	Prec. bordas	Prec. regiões
Manual	Não aplicável	69%	84%
gPb-owt-ucm	Brilho, cor, textura	65%	69%
FSEG	Textura	61%	66%
SRM	Cor, intensidade	60%	60%
JSEG	Cor, borda	56%	66%
MSEG	Cor, morfologia	57%	50%
Mean-shift	Cor, posição	58%	48%

Tabela 2 – Comparação entre as técnicas de segmentação de imagens, ordenados por desempenho decrescente, conforme resultados em Yuan, Gleason e Cheriadat (2013).

Tanto na revisão feita por Arbelaez et al. (2011) com a base de imagens naturais BSD500, quanto na comparação feita por Yuan, Gleason e Cheriadat (2013) em uma base de imagens aéreas, o algoritmo gPb-owt-ucm de Arbelaez et al. (2011) possui um desempenho superior aos demais algoritmos avaliados. Nenhum dos trabalhos faz qualquer avaliação sobre o custo computacional dos algoritmos, nem sobre o tempo de execução durante os experimentos.

Conforme descrito em detalhes no capítulo 4, a abordagem escolhida neste trabalho para encontrar elementos antrópicos em imagens aéreas da floresta amazônica passa pela classificação das regiões de imagens produzidas por um algoritmo de segmentação. Por este motivo, um levantamento bibliográfico também foi feito sobre classificação de imagens aéreas.

## 3.2 Classificação de imagens aéreas

Os trabalhos relacionados sobre classificação de imagens aéreas foram selecionados de acordo com suas semelhanças com o trabalho proposto neste documento. Todos tratam de imagens aéreas ortogonais, ou seja, com inclinação de aproximadamente 90° em relação ao solo, em cenas naturais e com bases de dados com forte presença de vegetação. Nem todas as bases de dados são obtidas através de veículos aéreos, algumas advêm de satélites em órbita terrestre. Os trabalhos relacionados expressam diferentes formas de classificação das imagens: classificação de pixels individuais, blocos, segmentos ou superpixels.

O trabalho de Dubuisson-Jolly e Gupta (2000) apresenta uma técnica de segmentação focada em imagens aéreas coloridas que realiza segmentações separadamente por cor e textura, para no final unir as duas e chegar a uma segmentação final utilizando um algoritmo de classificação por máxima verossimilhança (*Maximum Likelihood*). Os autores chegaram à conclusão que informações de cor são mais eficientes para a localização de bordas, enquanto a textura provê uma classificação menos ruidosa das regiões da imagem. A computação isolada das características foi importante para entender melhor que tipos

de atributos contribuem melhor em que etapas do trabalho, mas a metodologia utilizada para mesclar as segmentações é pouco explicada no artigo. O objetivo do trabalho de Dubuisson-Jolly e Gupta (2000) é a atualização de mapas antigos a partir de imagens recentes, mas pode facilmente ser utilizado na classificação de cobertura de terreno, ou de regiões previamente segmentadas.

De acordo com Sadgal, Fazziki e Ouahman (2005), o processamento de imagens digitais que representam cenas naturais requer elaboração substancial em todos os níveis: pré-processamento, segmentação, reconhecimento e interpretação. O trabalho apresentado sugere uma abordagem onde todas essas etapas acontecem em um único nível, e propõe um modelo de visão que tenta generalizar o reconhecimento de objetos utilizando categorização e cooperação. A solução proposta combina processos estocásticos, dentre os quais Inferência Bayesiana, Campos Aleatórios de Markov, com métodos não-estocásticos como Redes Neurais Artificiais. Esta diversidade de métodos é utilizada na segmentação e na extração de características de cores, texturas e formas, que depois são usadas na classificação dos objetos. Uma vantagem importante deste método é a possibilidade de paralelizar o processo de classificação, uma vez que as diversas técnicas são fundidas no final do processo, ao invés de serem aplicadas em cascata. Embora os resultados pareçam ser satisfatórios em imagens naturais, pouco é dito sobre como o processo de fusão de classificadores é feito, e nenhuma implementação ou base de imagens está disponível publicamente.

A pesquisa apresentada por Munoz-Mari et al. (2007) propõe uma abordagem que consiste em combinar diversos classificadores unários para classificar imagens multiespectrais de satélite pelo tipo de cobertura de terreno entre as classes: urbano, não-urbano e desconhecido. O trabalho realiza uma série de experimentos que combinam diversos classificadores unários, cada um treinado para reconhecer uma das classes do problema e chega à um veredito de classificação utilizando médias e probabilidades *a posteriori* das classes. Os algoritmos utilizados nos experimentos são a Mixture of Gaussians (MoG) e o *Support Vector Domain Description* (SVDD), uma variação de SVM para classificação unária. O trabalho alcança acurácia superior a 97% em alguns casos, o que indica que a abordagem de utilizar um conjunto de classificadores unários pode ser válida para a problemática aqui apresentada.

A dissertação de mestrado de Fernandes (2008) descreve uma solução de detecção de áreas de desmatamento em imagens de radar e satélite da região amazônica. Para tanto, uma segmentação das imagens utilizando a técnica de *meanshift* é realizada, cabendo ao algoritmo de aprendizado SVM a classificação destes segmentos em "área desmatada" ou "área não-desmatada". Com a precisão média geral de 87% de acurácia para imagens de satélite e 74% para imagens de radar, a autora considera que os resultados foram satisfatórios, visto que apenas 2,3% dos segmentos analisados apresentam erros graves para a classe de interesse. A mesma métrica gira em torno de 5,8% para as imagens de

radar. Este é o único trabalho relacionado que se utilizou de imagens de satélite, mas foi incorporado à bibliografia por se tratar de imagens da região amazônica e com vários objetivos em comum.

A pesquisa realizada por Ahmadi (2013) tem como objetivo fazer segmentação e classificação de imagens aéreas, pixel a pixel. Para tal, diversos classificadores e atributos das imagens são testados, chegando-se a conclusão de que o uso do algoritmo de KNN em características de cor e textura, mais precisamente o filtro de Gabor (FOGEL; SAGI, 1989) dos canais de matiz, saturação e intensidade (HSV) de cada pixel, obtiveram os melhores resultados dentre os algoritmos testados. Este trabalho argumenta que métodos estabelecidos na literatura costumam classificar as imagens com base em segmentos, o que supostamente costuma levar mais tempo que uma abordagem que classifique diretamente os pixels, mas os resultados dos experimentos não são particularmente precisos, com acurácia máxima de 82,23% na base de imagens testada.

O trabalho de Ghiasi e Amirfattahi (2013) realiza segmentação e classificação de tipos de terreno em imagens aéreas através de dois passos: primeiramente a imagem é dividida em superpixels, utilizando a técnica de fluxos geométricos de Levinshtein et al. (2009); posteriormente, cada superpixel tem suas características de textura e cor extraídas e é classificado através do algoritmo KNN. As características apontadas como mais úteis pelos autores são o Local Binary Pattern Histogram Fourier (LBP-HF) (AHONEN et al., 2009) para informações de textura e histograma dos canais RGB para informações sobre cores. Os autores do artigo alegam realizar o processo em tempo real, com precisão superior a 95% em todas as classes utilizadas. Apesar do cenário e região apresentados pelo trabalho de Ghiasi e Amirfattahi (2013) serem diferentes dos nossos, o tipo de imagem, as condições de aquisição e a diferenciação de elementos antrópicos (edificações, neste caso) estabelecem uma forte relação entre os dois trabalhos.

A tabela 3 apresenta um sumário dos trabalhos levantados nesta seção, ordenados conforme citação.

<b>Trabalho</b>	<b>Algoritmos</b>	<b>Amostra</b>	<b>Problema investigado</b>	<b>Acurácia</b>
Dub-Jolly	Máxima verossimilhança	Pixel	Atualização de mapas	91,8%
Sadgal	Redes neurais	Blocos	Classificação de terreno	-
Munoz-Mari	Combinação de unários	Pixels	Detecção de áreas urbanas	97,2%
Fernandes	SVM	Segmentos	Detecção de áreas desmatadas	87,0%
Ahmadi	KNN	Pixel	Classificação de terreno	82,2%
Ghiasi	KNN	Superpixel	Busca por objetos de interesse	95,0%

Tabela 3 – Comparação entre os trabalhos sobre classificação de imagens aéreas.

Os trabalhos relacionados à classificação de imagens aéreas exploram pouco a problemática de detecção de anomalias. O trabalho aqui apresentado tem como objetivo final a classificação e detecção de regiões anômalas à paisagem natural, portanto definidas como antrópicas. Tais regiões em um ambiente natural vasto como o da floresta amazônica,

tendem a ter ocorrência muito baixa, portanto considerar que os elementos antrópicos nas imagens sejam anomalias é uma modelagem com validade estatística.

Com a preocupação de que modelos de aprendizado supervisionados convencionais tenham dificuldade em reconhecer classes tão pouco representadas em uma base de treinamento, um levantamento foi feito por trabalhos na área de detecção de anomalias utilizando aprendizagem de máquina.

### 3.3 Detecção de anomalias

Recentes trabalhos na literatura passaram a utilizar algoritmos de classificação unária para encontrar anomalias em imagens dos mais diversos tipos. Foi realizado um levantamento de trabalhos na área de detecção de anomalias que utilizam algoritmos de aprendizado de máquina e imagens. Uma especial atenção foi dada à aspectos que pudessem ser utilizados neste trabalho, tais como os algoritmos, as características extraídas da imagem, combinação de resultados e a forma de avaliação do desempenho.

O trabalho de Hegenbart, Uhl e Vécsei (2012) utiliza o algoritmo *One-class Support Vector Machines* (OC-SVM) em diagnósticos médicos. A pesquisa utiliza o classificador unário baseado em SVM juntamente com o padrão local binário (*Local Binary Pattern* ou LBP), um popular descritor de texturas invariante à rotação da imagem, para detectar mudanças na textura de regiões do aparelho digestório, especialmente na análise de presença de doença celíaca. Embora a temática seja distinta do trabalho aqui apresentado, os resultados de Hegenbart, Uhl e Vécsei (2012) vem somar aos achados de Ghiasi e Amirfattahi (2013) no uso de um robusto descritor de texturas como o LBP para problemas de classificação de imagens.

Poucos trabalhos na área de sensoriamento remoto realizam experimentos com detecção de anomalias. O trabalho de Pla, Carmona e Sotoca (2013) consiste em um breve resumo do estado da arte de classificação unária para reconhecimento de imagens, seguido de uma aplicação real de classificação de pixels. Os autores apresentam os resultados da detecção de vegetação (classe majoritária) *versus* solo nu (classe minoritária) em uma base de imagens hiperespectrais de satélite com 207 bandas, formado da fusão de diversas fontes. Com 100% de verdadeiros positivos e 9% de falsos positivos para a classe majoritária, o trabalho não dá detalhes sobre a distribuição das classes na base de dados nem apresenta resultados para a classe minoritária, portanto, não é possível saber se os resultados são satisfatórios. A utilidade e relação do trabalho de Pla, Carmona e Sotoca (2013) com este aqui apresentado residem no uso do algoritmo *One-class Support Vector Machines* (OC-SVM), considerado estado da arte em diversas aplicações de aprendizado unário e na temática do problema apresentado, que se utiliza de imagens aéreas de regiões com intensa cobertura vegetal.

O artigo de Wang et al. (2013) busca comparar métodos de detecção de anomalias utilizando aprendizagem de máquina, especificamente classificadores unários, para detectar elementos anômalos em imagens hiperespectrais de satélite. As duas bases de dados utilizadas pelo trabalho são um conjunto de imagens de regiões florestais e agrárias do estado de Indiana, nos Estados Unidos, e um conjunto de imagens urbanas de San Diego, no mesmo país. A novidade do trabalho é a mesclagem de características espaciais e espectrais, com a finalidade de melhorar os resultados de trabalhos anteriores. Os resultados apresentados pelos autores apontam para taxas de acurácia de até 94% para o algoritmo SCSVDD (Spatial-Contextual Support Vector Data Description), uma variação de SVM para problemas de classificação unária, mas com heurísticas bastante específicas para o problema apresentado pelos autores.

Os trabalhos citados nesta seção foram dispostos na tabela 4, onde podem ser comparados quanto à suas aplicações, algoritmos utilizados e acurácia, quando disponível.

<b>Trabalho</b>	<b>Algoritmos</b>	<b>Amostra</b>	<b>Problema investigado</b>	<b>Acurácia</b>
Hegenbart (2012)	OC-SVM	Imagem inteira	Diagnóstico médico	82,9%
Pla (2013)	OC-SVM	Pixel	Detecção de vegetação	-
Wang (2013)	SCSVDD	Pixel	Detecção de objetos incomuns	94%

Tabela 4 – Comparação entre os trabalhos sobre detecção de anomalias.

Este trabalho pretende adaptar ou enriquecer os métodos utilizados na literatura, aplicando-os especificamente à detecção de elementos antrópicos em imagens aéreas da floresta amazônica, que possui seus desafios característicos, visto que o tipo de terreno e vegetação apresentam padrões diferentes dos vários trabalhos realizados em áreas urbanas ou florestas temperadas. Os trabalhos encontrados relacionados à floresta amazônica comumente utilizam outros tipos de sensores como RADAR e LIDAR, além de câmera de espectro visível, como pode ser visto nos trabalhos de Linhares et al. (2014) e Santos, Maldonado e Graça (2005) e Fernandes (2008). Esses trabalhos também tendem a se preocupar com dados temporais, como avanço do desmatamento e levantamento de grandes áreas.

O diferencial deste trabalho está na aplicação ao tema de vigilância ambiental através de VANTs e a preocupação com disponibilização dos conjuntos de dados rotulados para trabalhos futuros na mesma problemática. Mais detalhes sobre o método proposto e as estratégias estudadas serão discutidos no capítulo 4.

## 4 Metodologia

Neste capítulo é descrita a sequência de etapas que serão realizadas neste trabalho para que os objetivos de pesquisa sejam alcançados.

Em imagens aéreas, como as utilizadas neste trabalho, é comum que os elementos que indicam presença humana sejam relativamente grandes (pista de pouso, estradas, clareiras, etc.), podendo ser definidas como uma região durante a segmentação da imagem. Portanto, classificar pixels tende a apresentar custo computacional elevado, visto que mesmo uma pequena imagem provê dezenas de milhares deles, que devem ter suas características extraídas e providas ao modelo de aprendizado, para que possam ser classificados. Portanto, utilizar técnicas de segmentação de imagem para agrupar os pixels espacialmente e caracteristicamente relacionados em uma única amostra não só possibilita uma execução mais rápida da solução computacional, como também torna o resultado final menos ruidoso.

Por este motivo, a arquitetura para a solução proposta neste trabalho prevê uma etapa de segmentação das imagens, seguida de uma etapa de classificação, responsável pela determinação do tipo de cada região encontrada na segmentação. Desta forma, elementos antrópicos podem ser separados como mais uma das regiões das imagens e classificados como tal. O diagrama da arquitetura geral da solução pode ser visto na figura 8.

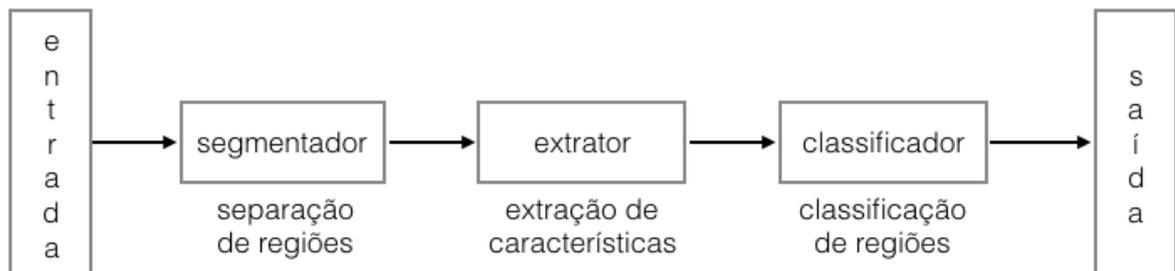


Figura 8 – Arquitetura geral da solução desenvolvida para detecção de elementos antrópicos em imagens aéreas da floresta amazônica.

A ideia geral é que imagens aéreas de regiões florestais da Amazônia legal sirvam de entrada para o problema. Estas mesmas imagens serão particionadas em regiões por um segmentador. Em seguida, um extrator será utilizado para criar o vetor de características de cada segmento provido pela etapa anterior. Por fim, um classificador composto por um ou mais algoritmos de classificação será utilizado para rotular as regiões, com a finalidade de encontrar as regiões de elementos antrópicos. A saída do sistema pode, então, ser composta pelas regiões segmentadas e suas classificações finais.

Embora esta seja uma arquitetura simples, bastante difundida na literatura, nuances para o problema apresentado neste trabalho devem ser levadas em conta. Para chegarmos

à conclusão de quais segmentadores, extratores e classificadores devem ser utilizados, são necessárias várias etapas de experimentação, validação e análise de resultados.

## 4.1 Entrada

Uma base de dados com imagens aéreas de floresta tropical precisa ser formada. As imagens não precisam ter as mesmas dimensões, mas devem ter condições visuais claras do cenário e pertencerem à região de floresta amazônica, excluindo imagens de cidades e povoados da região.

Como serão processadas por uma etapa de segmentação, as imagens não sofrerão nenhum tipo de filtragem, ficando estas a cargo dos segmentadores a serem avaliados. A base de imagens gerada nesta etapa é um dos objetivos específicos deste trabalho.

## 4.2 Segmentador

Para encontrarmos o segmentador ideal, os métodos de segmentação de imagens considerados estado-da-arte serão aplicados à uma parte da base de imagens aéreas da floresta amazônica. Conforme levantado no capítulo 3, os seguintes métodos serão avaliados: Mean-shift, JSEG, MSEG, SRM, FSEG e gPb-owt-ucm. Estes métodos foram escolhidos por terem os melhores desempenhos no *benchmark* de segmentação de uma base de imagens aéreas de Yuan, Gleason e Cheriyyadat (2013).

Também será realizado um experimento que funde as etapas de segmentação e classificação em uma só, utilizando algoritmos de aprendizado de máquina supervisionados para segmentar as imagens diretamente para suas classes (floresta, água, etc) pixel a pixel. Isto é importante para determinar se há ganhos reais na separação entre segmentação e classificação das imagens, ou se realizar todo o procedimento em um único passo produz melhores resultados de acurácia e custo computacional.

A base de dados também precisa ser segmentada manualmente por seres humanos, pois este será o *ground-truth* usado para aferir o desempenho dos algoritmos de segmentação avaliados. Ainda é preciso aferir a consistência da segmentação manual realizada por seres humanos. Métricas de erros de consistência como o *Local Consistency Error* (LCE) e *Global Consistency Error* (GCE), descritos na seção 2.1.4, devem ser aplicados para a medição de desempenho de todos os algoritmos, bem como para validação da consistência interna da segmentação manual.

Esta etapa deve determinar o método de segmentação com melhores resultados, a ser utilizado na solução descrita pelo trabalho. A precisão e o tempo de execução devem ser utilizados para a avaliação do desempenho dos algoritmos testados. Em seguida, é preciso extrair as características dos segmentos gerados.

### 4.3 Extrator

Os algoritmos de classificação lidam com variáveis numéricas inteiras, de ponto flutuante e em alguns casos, dados textuais. No entanto, para classificar imagens, ou no caso deste trabalho, segmentos de imagens, é preciso que um conjunto de características seja extraído dos pixels destas imagens ou regiões e disponibilizados para o modelo de classificação.

Nesta etapa serão testadas diversas características disponíveis na literatura de processamento digital de imagens, especialmente informações sobre cor, intensidade, textura e morfologia destas amostras. A escolha das características foi baseada, inicialmente, nas características utilizadas em trabalhos relacionados de classificação de imagens aéreas: canais RGB para descrição de cores, histogramas de intensidade, *Local Binary Patterns* para descrição de texturas e transformada de Hough para detecção e descrição de formas geométricas, em especial linhas retas.

Os atributos também deverão passar por um processo de seleção baseada em correlação, o CFS. O objetivo é diminuir a dimensionalidade do vetor de características e avaliar se há ganho no desempenho e generalização dos algoritmos de aprendizado, utilizando este subconjunto do vetor de características original.

Como requisito da composição da base de dados de segmentos, bem como para avaliação e otimização do vetor de características, todas as amostras geradas pela etapa de segmentação devem ser rotuladas manualmente e contar com a avaliação de especialistas na inspeção deste tipo de imagem. Com a base devidamente criada e rotulada, experimentos para determinar os classificadores mais adequados podem ser feitos.

### 4.4 Classificador

Nesta etapa, quatro estratégias de classificação presentes na literatura serão abordadas: classificadores multi-classe, binários, unários e conjuntos de classificadores unários. Essas estratégias não demandam alterações na arquitetura geral da solução, mas mudam a forma como a base é rotulada e organizada. As métricas de avaliação do aprendizado precisam ser ponderadas para cada estratégia.

A motivação para explorar estas quatro abordagens é a diversidade de estratégias encontradas na literatura relacionada. É importante destacar que a classe de interesse, elementos antrópicos, permanece a mesma em todas as abordagens, e sua inspeção mais cuidadosa é vital na avaliação de todos os métodos utilizados. Como esta classe é absolutamente minoritária (<1% das instâncias) na base de imagens utilizada no trabalho, abordagens que lidam com anomalias de forma mais robusta serão investigadas.

### 4.4.1 Classificador multi-classe

Nesta estratégia, a base de segmentos gerada será rotulada entre cinco classes possíveis: floresta, vegetação rasteira, água, terra e elemento antrópico. A escolha das classes foi feita através de uma observação das classes de trabalhos relacionados.

Também observando os algoritmos utilizados em trabalhos relacionados que modelam o problema como uma questão multi-classe, diversos métodos de classificação multi-classe serão testados: *Random Forest*, KNN, SVM, árvore de decisão e *Naive Bayes*. Esta gama de métodos também serve para cobrir os testes com abordagens de aprendizado simbólico, bayesiano, aprendizagem preguiçosa (do inglês *lazy learning*) e ensemble de classificadores multi-classe.

Todos os métodos devem ser testados com a base completa de segmentos, mas também com a base pré-processada por um seletor de características, que será responsável pela redução do vetor de características do problema. Este detalhe do experimento servirá para determinar se a seleção de atributos nesta abordagem reduz a complexidade dos modelos gerados e o ruído na base de dados, possibilitando melhor taxa de aprendizagem.

Cada algoritmo deve ser responsável pela classificação de toda a base. Ao fim, métricas de aprendizado como acurácia, precisão e revocação serão utilizadas para comparar os métodos entre si.

### 4.4.2 Classificador binário

Neste cenário, a base de segmentos gerada será rotulada entre duas classes: elementos naturais e elementos antrópicos. A classe de elementos naturais agrupa o que originalmente seriam as classes de floresta, vegetação rasteira, água e terra. A motivação desta modelagem é realizar experimentos com algoritmos que permitam o aprendizado bi-classe, abordagem pouco usada nos trabalhos encontrados na literatura, bem como avaliar se o agrupamento das classes secundárias (as classes que não são de elementos antrópicos), pode ser descrito de forma eficaz pelos algoritmos testados.

Os classificadores utilizados serão os mesmos do experimento com classificadores multi-classe, visto que não há grande diferença metodológica nas duas abordagens. Todos os métodos também serão testados com seletores de características, e o impacto desta seleção também será avaliado.

Ao fim do experimento, as métricas propostas para o problema de aprendizado serão colhidas para cada método testado, e utilizadas para comparar os métodos entre si.

### 4.4.3 Classificador unário

Assim como no experimento de classificadores binários, a base de dados de segmentos será dividida entre elementos naturais e elementos antrópicos, utilizando os mesmos critérios. À primeira vista, esta estratégia de aprendizado parece muito similar aos classificadores binários, mas a diferença está inicialmente nos algoritmos utilizados.

Os métodos de classificação unária tendem a lidar melhor com bases de dados cujas classes estão fortemente desbalanceadas, provável caso da base de imagens utilizada neste trabalho, especialmente depois do rearranjo entre elementos antrópicos e demais classes de elementos naturais. Por esta razão, a abordagem que utiliza classificadores unários precisa ser avaliada.

Como o aprendizado unário utiliza o conceito de classe majoritária e classe anômala (*outlier*), é preciso definir quais classes do problema exercerão estes papéis. Por uma simples questão de frequência estatística, fica definido que a classe de elementos naturais é a classe majoritária, enquanto a classe de elementos antrópicos é considerada a classe anômala.

Apoiando-se nos casos apresentados pela literatura, em especial nos trabalhos relacionados que se propõem a realizar detecção de anomalias, os algoritmos utilizados neste experimento serão os *One-Class Support Vector Machines* (OC-SVM) e o REPTree. Aqui também deve ser avaliado o impacto de um seletor de atributos aplicado ao vetor de características da base de dados, a fim de entender se há mudança significativa na taxa de aprendizado da classe anômala.

### 4.4.4 Conjunto de classificadores unários

Esta abordagem consiste na criação de modelos de aprendizado unários para cada classe do problema, exceto para a classe anômala, e posteriormente na combinação dos resultados destes classificadores unários para um resultado multi-classe. Sendo assim, um ou mais classificadores unários serão treinados para cada uma das classes de floresta, vegetação rasteira, água e terra. A motivação por trás do uso desta abordagem é o crescente número de trabalhos na literatura de classificação de imagens que utiliza este grupo de técnicas, em especial trabalhos que discorrem sobre descrição de objetos em grandes coleções de imagens.

A preparação da base de dados é um importante passo nesta estratégia de aprendizado, visto que diversas versões da base de segmentos terão de ser geradas, cada uma representando a classe não-anômala em questão como a classe majoritária e todas as demais como anomalias.

O classificadores criados serão agrupados em um *ensemble* de classificadores e cada um dos modelos será utilizado na composição do veredito final, para a classificação da

amostra. Por uma limitação de escopo do trabalho, uma abordagem simples de votação com pesos deve ser utilizada. Os pesos utilizados, bem como as implicações desta abordagem devem ser discutidas nos resultados do experimento.

Assim como todas as outras estratégias de classificação presentes neste trabalho, as métricas mais relevantes serão a acurácia, precisão e revocação da classe de interesse, de elementos antrópicos, que aqui é a classe anômala. As abordagens de classificação com melhor desempenho nestes quesitos ou mais promissoras devem ser apontadas ao fim destes experimentos.

## 4.5 Saída

Os resultados dos experimentos anteriores devem sustentar a conclusão sobre quais métodos de segmentação e classificação são mais indicados para a solução. A saída da solução deve indicar quais segmentos da coleção de imagens original possuem elementos antrópicos. A avaliação do desempenho na detecção destes elementos é a métrica definitiva da adequação da solução para o problema proposto.

O próximo capítulo descreve os experimentos realizados e discute os resultados encontrados.

## 5 Experimentos

Neste capítulo são apresentados o protocolo experimental e resultados de todas as etapas e experimentos deste trabalho. Os resultados das seções sobre a base de dados (5.1) e segmentação de imagens (5.2) são apresentados de forma cronológica, enquanto os resultados da seção de classificação (5.3) e suas subseções são apresentados de forma não-cronológica, privilegiando a ordem das abordagens conforme descritas na metodologia (capítulo 4).

### 5.1 Base de dados

A base de dados (imagens) utilizada advém do projeto GEOMA (INPE, 2013), financiado pelo Instituto Nacional de Pesquisas Espaciais (INPE). Trata-se de imagens coloridas, codificadas em JPEG e com 640 pixels de largura por 480 pixels de altura. A base é composta por fotografias ortogonais ao relevo (como pode ser visto na figura 9), de altitudes variadas e tiradas a partir de aeronaves tripuladas, durante o trajeto entre diversas cidades da região amazônica.

No momento do início dos experimentos deste trabalho, estas imagens tiradas de aviões tripulados eram as únicas da região da Amazônia legal disponíveis publicamente. Podemos considerá-las válidas por terem sido tiradas em altitude de voo compatível com as missões de VANTs de vigilância, entre 900 e 1.100 metros do solo. Como este trabalho tem como objetivo utilizar apenas câmeras de espectro visível, são dispensáveis comparações de sensores com VANTs que eventualmente possuam sonar, câmeras infravermelho ou outros tipos de sensores.

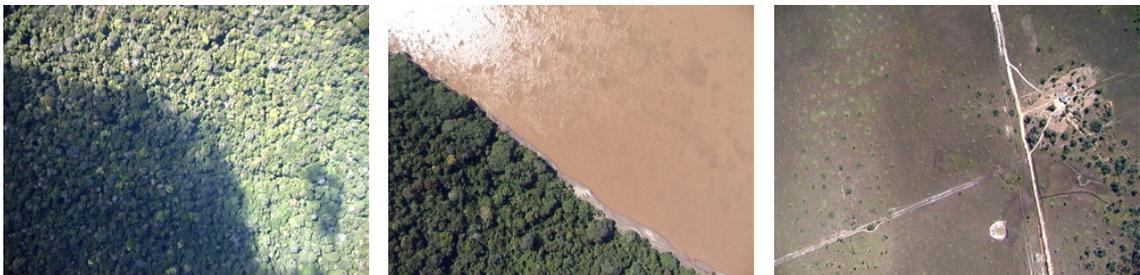


Figura 9 – Amostras da base de dados.

A base possui um total de 3.044 imagens, com dimensão total de 1,02 Gigabytes de dados. Todas as imagens foram utilizadas no presente trabalho sem tratamento ou manipulação prévios.

Para criar uma referência (*ground-truth*) para a segmentação das imagens da base

de dados, uma ferramenta computacional executável em navegadores web foi construída (figura 10). A saída deste aplicativo é uma coleção, para cada imagem, de informações sobre bordas das regiões da imagem. Estas informações serviram de referência para avaliar o desempenho dos algoritmos de segmentação testados neste trabalho.

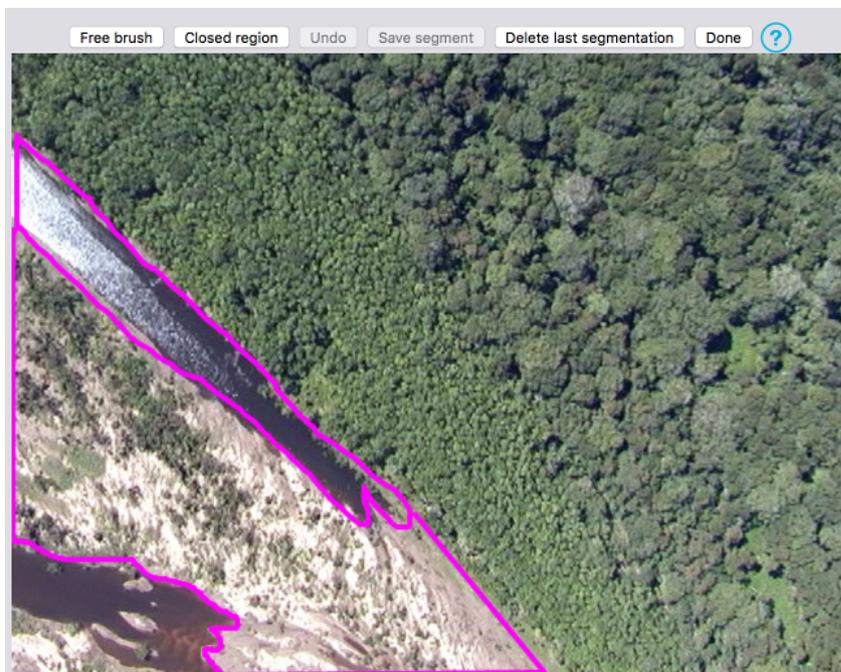


Figura 10 – Ferramenta para segmentação manual das imagens.

Um dos objetivos específicos deste trabalho é a criação e disponibilização de uma base de imagens aéreas da floresta amazônica, com seus segmentos definidos, devidamente rotulados e elementos antrópicos assinalados. Esta base de imagens está publicamente disponível<sup>1</sup>, e pode ser utilizada em diversos trabalhos futuros com temas relacionados.

## 5.2 Segmentação

Um experimento de comparação entre diversos algoritmos de segmentação de imagens foi planejado. Cada imagem da base de dados foi segmentada por seres humanos, segmentações estas que constituiram a base de referência para o experimento. Cada imagem tem sua segmentação de referência consolidada a partir da segmentação manual de pelo menos 5 indivíduos.

### 5.2.1 Protocolo experimental

Para criação da segmentação manual de referência (*ground-truth*), 31 voluntários, todos alunos de pós-graduação em informática ou áreas relacionadas, foram convidados a

<sup>1</sup> <http://github.com/luizcavalcanti/geoma-database>

realizar a segmentação manual das imagens da base de dados, através de uma ferramenta<sup>2</sup> criada com esta finalidade.

A ferramenta de software web criada consiste de uma interface gráfica onde o usuário pode desenhar sobre uma imagem a ser segmentada. A ideia é que nesta imagem sejam circunscritas as bordas das regiões definidas pelo usuário, de acordo com instruções fornecidas pela ferramenta e que foram lidas obrigatoriamente por cada voluntário antes do início do experimento.

Em linhas gerais, as instruções orientaram os participantes do experimento a segmentar as imagens de acordo com a cobertura ou tipo de terreno, formação geológica ou vegetação, utilizando os critérios e granularidade que lhes pareçam mais adequados:

*Your mission is to manually segment the given images as accurately as possible, according to your own judgment. The criteria here is terrain coverage. So, we would like to separate different vegetations, geological formations and human-made objects. (CAVALCANTI, 2015)*

Em tradução livre:

Sua missão é segmentar manualmente as imagens disponibilizadas o mais precisamente possível, de acordo com o seu julgamento. O critério aplicado é a cobertura de terreno. Portanto, nós gostaríamos de separar diferentes vegetações, formações geológicas e objetos feitos por seres humanos. (CAVALCANTI, 2015)

O conteúdo integral das instruções pode ser encontrado no apêndice B.

## 5.2.2 Resultados

Antes de analisar o desempenho de cada algoritmo de segmentação investigado neste trabalho em relação à segmentação manual, foi preciso aferir a validade da própria segmentação manual, visto que ela foi realizada por diferentes participantes com diferentes interpretações das instruções fornecidas. Segundo Martin et al. (2001), uma forma de validar a consistência das segmentações manuais é medir a consistência entre segmentações de uma mesma imagem feita por diferentes pessoas.

Utilizando as métricas de erro de consistência local (LCE) e global (GCE) para medir a similaridade das segmentações manuais, os resultados confirmam a consistência da segmentação manual feita pelos participantes do experimento. Conforme pode ser visto nas figuras 11 e 12, os coeficientes de erro global e local, respectivamente, ficam abaixo dos 20% e 10% para todos os casos.

<sup>2</sup> <http://amazonsegmentation.ddns.net/>

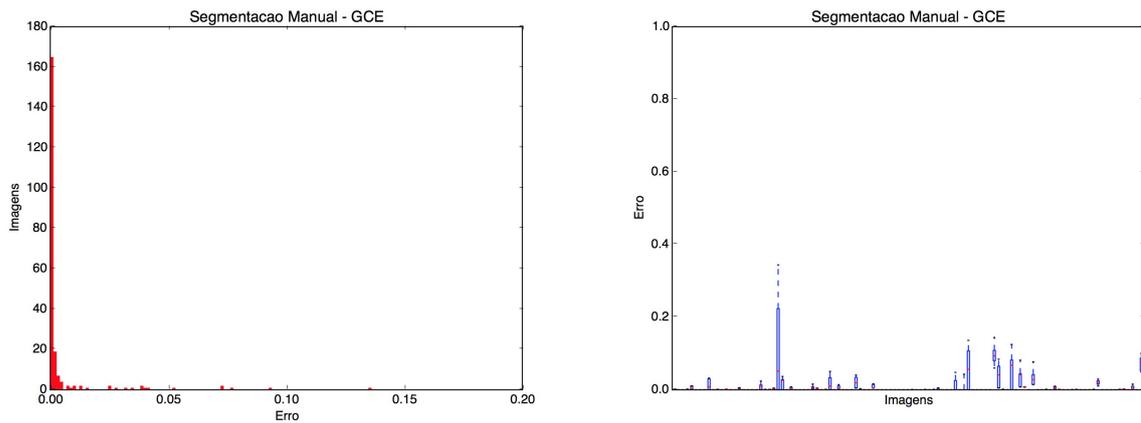


Figura 11 – Coeficiente de erro global das diferentes segmentações de mesma imagem.

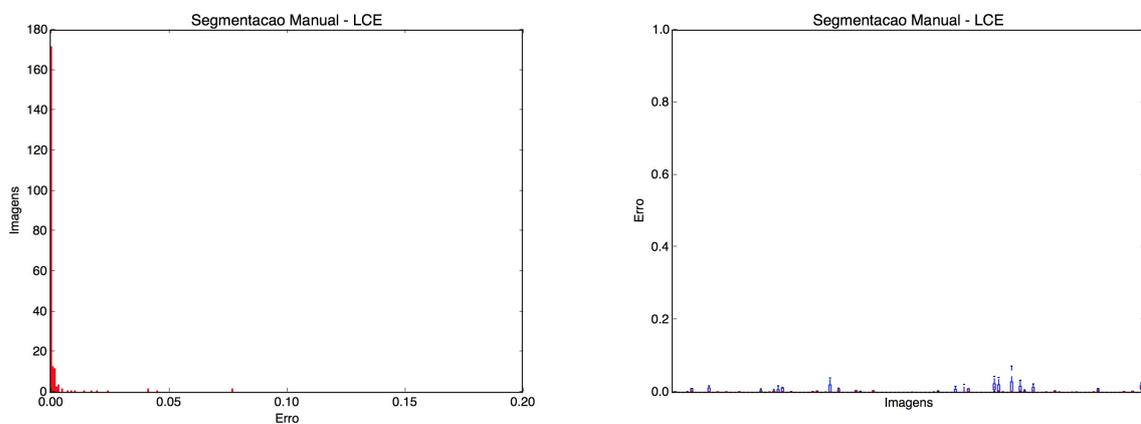


Figura 12 – Coeficiente de erro local das diferentes segmentações de mesma imagem.

Para determinar qual dos algoritmos de segmentação levantados na pesquisa bibliográfica teria melhor desempenho na base de dados utilizada neste trabalho, todos foram implementados ou adaptados. Os algoritmos foram testados em todas as imagens da base de dados do trabalho que possuíam segmentação manual por pelo menos 5 voluntários, um total de 203 imagens, utilizando as mesmas medidas de erro global (GCE) e erro local (LCE) apresentados no trabalho de Martin et al. (2001) e na validação da base de segmentação manual construída para este trabalho.

Os resultados do experimento são apresentados na tabela 5, com o melhor resultado para cada critério destacado em coloração cinza.

O método SRM conseguiu uma média de erros global e local substancialmente menor que os demais algoritmos e foi considerado o método com melhor desempenho do experimento na base de imagens deste trabalho, embora o tempo de segmentação deste algoritmo seja uma ordem de magnitude maior que o algoritmo MSEG, que obteve o melhor tempo de execução. A imagem 13 mostra a saída de alguns dos métodos testados, para fins de comparação visual.

Algoritmo	GCE médio	Desvio	LCE médio	Desvio	Tempo (s)
Manual	0.01822	$\pm 0.04612$	0.00537	$\pm 0.04795$	-
FSEG	0.03063	$\pm 0.05644$	0.00273	$\pm 0.06352$	13,91
gPb-owt-ucm	0.00655	$\pm 0.01368$	0.00297	$\pm 0.01409$	237,32
JSEG	0.02990	$\pm 0.05527$	0.00486	$\pm 0.06105$	14,82
Mean-shift	0.02237	$\pm 0.05818$	0.00271	$\pm 0.06214$	6,39
MSEG	0.01005	$\pm 0.01429$	0.00072	$\pm 0.01429$	0,33
SRM	0.00622	$\pm 0.00822$	0.00066	$\pm 0.00999$	4,66

Tabela 5 – Comparação de métodos de segmentação em parte da base de imagens deste trabalho, em ordem alfabética. Os erros médios da segmentação manual foram calculados entre diferentes segmentações da mesma imagem.

Adicionalmente, experimentos foram realizados com técnicas de classificação. O objetivo era saber se poderíamos utilizar apenas uma etapa para realizar segmentação e a classificação dos segmentos. Nesse experimento, um conjunto de imagens da mesma base utilizada neste trabalho foi rotulado de acordo com a cobertura de terreno, pixel a pixel, sendo submetido posteriormente à treinamento e teste de diversos métodos de aprendizado de máquina supervisionado.

Todos os algoritmos obtiveram precisão inferior ao que foi conseguido na etapa de segmentação isoladamente. Além disso, a quantidade de ruído na classificação - pixels de uma classe  $c_1$  espalhados por grandes regiões de uma classe  $c_2$  - foi muito alta. Outra desvantagem encontrada nesta abordagem é o alto tempo de processamento para cada imagem. Os resultados foram publicados no 10<sup>th</sup> *International Conference on Computer Vision Theory and Applications* e o artigo de Cavalcanti, Carvalho e Santos (2015) completo pode ser visto no apêndice A.

Para fins de informação, os resultados deste experimento com técnicas de segmentação são apresentados na tabela 6, com o melhor resultado para cada critério destacado em coloração cinza. O tempo de execução da classificação para cada imagem no artigo publicado ignora o tempo de extração de características da imagem. Para uma comparação correta deste aspecto dos resultados com os métodos de segmentação apresentados anteriormente, o tempo gasto em extração de características foi acrescido na tabela desta seção. Como a classificação pixel a pixel da imagem não necessariamente resulta em regiões conexas de pixels pertencentes à mesma classe, não é adequado fazer a comparação dos resultados utilizando as métricas de LCE e GCE.

Com um tempo de processamento baixo e excelentes resultados nas métricas de erro local e global, o algoritmo SRM foi escolhido como método para determinar os segmentos de imagem a servir de entrada para a próxima etapa da solução, responsável pela classificações destes mesmos segmentos.

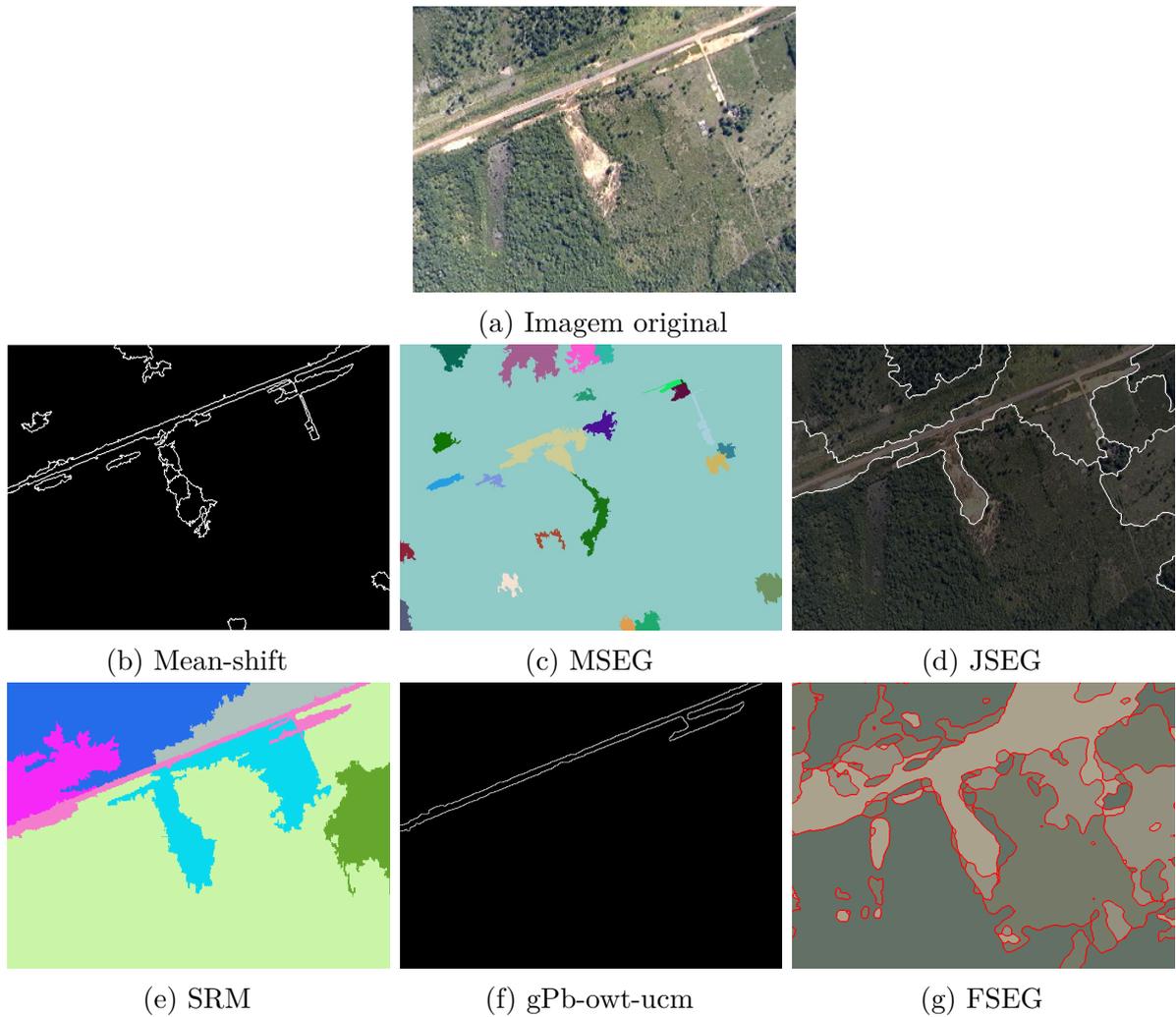


Figura 13 – Comparação visual de métodos de segmentação.

Algoritmo	Acurácia	Tempo/imagem
Random forest	96,0%	12,72 s
KNN	92,6%	22,89 s
Naive Bayes	92,8%	8,36 s
Decision tree	82,2%	14,49 s

Tabela 6 – Comparação de métodos de classificação para segmentação das imagens em uma única etapa, ordenados por acurácia.

### 5.3 Classificação

A segunda etapa da solução consiste em classificar os segmentos de imagens produzidos na etapa de segmentação. Um conjunto de experimentos foi realizado com base nas quatro abordagens para classificação supervisionada, descritas no capítulo 4:

- Classificadores multi-classe
- Classificadores binários

- Classificadores unários
- Conjunto de classificadores unários

### 5.3.1 Protocolo experimental

Primeiramente, todas as imagens foram segmentadas utilizando o algoritmo SRM, por conta de seu melhor desempenho no experimento de segmentação anterior. Cada um dos 10.057 segmentos produzidos nesta segmentação foi devidamente classificado manualmente, e suas características foram extraídas. Cada segmento foi classificado como uma das possíveis classes para o problema: floresta, vegetação rasteira, água, terra ou elemento antrópico.

Cerca de 70 segmentos não puderam ser classificados, pois continham mais de um tipo de terreno dentro de seus limites, tornando a rotulação dúbia. É importante destacar que nenhum destes segmentos continha elementos antrópicos, portanto, foram descartados da base utilizada no treinamento e validação dos modelos de aprendizado.

Todos os elementos antrópicos encontrados foram validados pela Dra. Solange Costa, especialista do Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM), órgão governamental federal responsável pelo patrulhamento e sensoriamento remoto da Amazônia legal brasileira.

Em seguida, uma análise da base de segmentos rotulada foi feita, observando a estrutura e a distribuição dos dados ao longo de toda a base, levando em consideração como os tipos de dados utilizados e sua representação podem influenciar os resultados dos algoritmos de aprendizado utilizados no experimento.

Após a rotulação de todas as amostras utilizadas no experimento, a distribuição das classes do problema em toda a base foi medida e pode ser vista na tabela 7.

Classe	Amostras	Percentual
Floresta	8.684	86,3 %
Vegetação rasteira	939	9,3 %
Água	287	2,8 %
Terra	136	1,3 %
Elementos antrópicos	31	0,3 %

Tabela 7 – Distribuição de classes na base de segmentos.

Com um conjunto de dados altamente desbalanceado, algumas precauções na avaliação de desempenho dos algoritmos precisaram ser tomadas. A avaliação do aprendizado precisou considerar os cálculos de precisão e revocação das classes menos representadas no conjunto de dados, especialmente a classe de elementos antrópicos.

O vetor de características escolhido inicialmente para representar cada amostra compreende informações de cor, intensidade, morfologia e textura de cada segmento a ser classificado. Este conjunto de características foi escolhido pela alta representatividade de valores em determinadas classes, baixa dimensionalidade e simples depuração, bem como a presença em trabalhos relacionados de classificação de imagens aéreas.

A Cor média para os canais vermelho, verde e azul de todo o segmento trazem informação do tom de cor, enquanto a intensidade média e o histograma em tons de cinza ajudam a representar informações de intensidade. Para guardar informações sobre a textura e variância de intensidade do segmento, um histograma de *Local Binary Pattern*, descrito por Ahonen et al. (2009), foi extraído.

Finalmente, para tentar representar características de morfologia do segmento, a transformada de Hough foi utilizada. Neste último conjunto de características, a contagem de linhas retas e o comprimento da maior linha foram extraídos e passaram a compor o vetor de características das amostras.

Com a discretização e normalização das variáveis que compõem o vetor de características, obtivemos um total de 48 atributos numéricos. Um total de 10.057 amostras foram classificadas e o método de validação cruzada k-fold foi utilizado, com a finalidade de aferir não só a precisão, mas também a baixa variância e generalização dos modelos criados. A lista de atributos, seus tipos e dimensões podem ser vistos na tabela 8.

<b>Atributo</b>	<b>Tipo</b>	<b>Dimensão</b>
Vermelho médio	Real	1x1
Verde médio	Real	1x1
Azul médio	Real	1x1
Intensidade média	Real	1x1
Intensidade - histograma	Inteiro	16x1
LBP - histograma	Inteiro	26x1
Hough - número de retas	Inteiro	1x1
Hough - maior reta	Inteiro	1x1

Tabela 8 – Atributos gerados a partir da base de segmentos.

Para reduzir a dimensionalidade do vetor de características, a seleção de subconjunto de atributos baseada em correlação (*Correlation-based Feature Subset Selection*, ou CFS), introduzida por Hall (1998), foi utilizada. O vetor de características resultante varia de acordo com a abordagem de aprendizado utilizada, visto que o agrupamento de classes altera a distribuição da base de dados, e por consequência, modifica a correlação entre atributos e a classe da amostra.

Esta redução da dimensionalidade do vetor de atributos pode ser importante para aumentar a generalização e simplificar os modelos gerados no experimento, bem como diminuir problemas de dimensionalidade enfrentados por alguns dos algoritmos testados.

Para efeitos de comparação, todos os algoritmos foram testados com os atributos originais e com o subconjunto de atributos selecionados pela técnica CFS.

Finalmente, todos os algoritmos processaram os conjuntos de treinamento para criar um modelo de aprendizado e em seguida, utilizaram a base de testes para medir o quanto o modelo pode ser generalizado para bases de imagens diferentes.

Para todos os algoritmos testados, o método de validação cruzada foi utilizado para avaliar a variância na taxa de aprendizado e inferir a generalização dos modelos criados. Mais especificamente, o método k-fold foi utilizado, com um valor de  $k = 10$ .

Cada algoritmo testado foi refinado e parametrizado para que consiga produzir uma maior acurácia para todas as classes e para otimizar a revocação da classe de elementos antrópicos. As seções a seguir detalham o protocolo e os resultados da experimentação para cada abordagem de classificação supervisionada testada no trabalho.

### 5.3.2 Classificadores multi-classe

Utilizando as técnicas de K vizinhos mais próximos (KNN), máquinas de vetores de suporte (SVM), árvores de decisão, *Naive Bayes* e *Random Forest*, um experimento foi conduzido com o intuito de classificar os segmentos de imagem gerados na etapa anterior do trabalho e definir que método mais se adequa para a solução do problema proposto neste trabalho.

Cada algoritmo teve seus parâmetros ajustados de forma a maximizar os índices de precisão e revocação gerais, refletidos pela média da medida F1 de todas as classes do problema, observando o resultado obtido na validação cruzada. Para o algoritmo KNN, um valor de  $k = 3$  e a utilização de distância euclidiana obtiveram os melhores resultados. Na árvore de decisão, um fator de confiança para poda de 0,3 foi utilizado. Para o *ensemble* de classificadores *Random Forest*, um número máximo de 200 árvores, com agrupamentos de 10 características obtiveram os melhores resultados. Para o SVM, um kernel de base radial com  $\gamma = 0.1$  obteve a melhor medida F1 geral. *Naive Bayes*, por sua vez, não precisa ser parametrizado.

Todas as amostras disponíveis foram classificadas. Todos os classificadores foram treinados e avaliados com o conjunto integral de atributos, e também com o conjunto reduzido, utilizando a técnica de seleção de atributos CFS. O conjunto de atributos selecionados para as classes deste experimentos é apresentado na tabela 9. A acurácia, precisão e revocação são apresentadas na tabela 10, com o melhor resultado para cada critério destacado em coloração cinza.

Tanto por questões de balanceamento das classes como por questões contextuais do problema, falsos positivos para a classe de elementos antrópicos são toleráveis, enquanto falsos negativos devem ser minimizados, mesmo que em detrimento da precisão e revocação

Atributo	Tipo	Dimensão
Intensidade - histograma (5/15)	Inteiro	5x1
LBP - histograma (1/26)	Inteiro	1x1
Hough - maior reta	Inteiro	1x1

Tabela 9 – Atributos selecionados pela técnica de CFS para a abordagem de classificação multi-classe.

Método	Acurácia	Precisão	Revocação	F1
Random Forest	0.930	0.917	0.930	0.921
KNN	0.919	0.907	0.920	0.909
SVM	0.915	0.899	0.915	0.900
Random Forest (CFS)	0.909	0.893	0.910	0.899
Árvore de decisão	0.904	0.898	0.904	0.901
KNN (CFS)	0.898	0.880	0.898	0.887
Árvore de decisão (CFS)	0.893	0.880	0.894	0.886
SVM (CFS)	0.865	0.798	0.866	0.808
Naive Bayes (CFS)	0.816	0.845	0.816	0.824
Naive Bayes	0.563	0.857	0.563	0.101

Tabela 10 – Comparação de métodos de classificação para regiões segmentadas das imagens, ordenados por acurácia.

de outras classes do problema. A razão contextual para esta afirmação é a seguinte: é importante que o maior número de elementos antrópicos seja detectado, mesmo que isso signifique um aumento do número de alarmes falsos para os agentes de segurança responsáveis pela patrulhamento das regiões de floresta. Esta é uma linha de raciocínio intuitiva, não houve consulta a agentes de segurança sobre a questão.

A avaliação dos resultados de revocação, bem como a análise da curva ROC da classe de elementos antrópicos para cada algoritmo testado podem ser vistas na tabela 11, com o melhor resultado para cada critério destacado em coloração cinza.

Método	Precisão	Revocação	F1	Área ROC
KNN	0.889	0.727	0.800	0.907
Random Forest	0.917	0.500	0.647	0.999
SVM (CFS)	0.813	0.591	0.684	0.795
Árvore de decisão	0.700	0.636	0.667	0.839
Random Forest (CFS)	0.846	0.500	0.629	0.926
KNN (CFS)	0.722	0.591	0.650	0.806
Árvore de decisão (CFS)	0.588	0.455	0.513	0.724
Naive Bayes	0.054	0.955	0.101	0.984
Naive Bayes (CFS)	0.030	0.909	0.058	0.978
SVM	0.030	0.591	0.057	0.806

Tabela 11 – Comparação de métodos de classificação multi-classe em relação à classe de elementos antrópicos, ordenados pela medida F1.

Embora tenha tido os melhores resultados em todos os quesitos em uma análise geral da base de dados, o algoritmo *Random Forest* obteve baixa revocação para a classe de elementos antrópicos, indicando que houve bom aprendizado apenas para as classes mais abundantes na base. Apesar do bom desempenho na análise geral, o SVM obteve resultados abaixo do esperado na classe de elementos antrópicos, só apresentando melhora quando a seleção de atributos com o método CFS foi realizada. O método *Naive Bayes* obteve desempenho ruim em todas as análises, indicando que a abordagem bayesiana não é suficiente para distinguir as classes deste problema com os atributos escolhidos.

Tanto no aprendizado geral quanto especificamente no aprendizado da classe de elementos antrópicos, o algoritmo KNN obteve bom desempenho, apresentando melhor balanço entre precisão e revocação para a classe de maior interesse no problema. É seguro afirmar, também, que a seleção de atributos utilizando o método CFS não teve impacto positivo na generalização e precisão do aprendizado deste experimento, uma vez que praticamente todos os métodos obtiveram melhor desempenho com o conjunto completo de atributos.

Embora os resultados da classificação multi-classe tenham apresentado elevadas taxas de precisão e revocação, é importante investigar outras estratégias de aprendizado. Com o foco na classe de elementos antrópicos, é preciso averiguar se a modelagem do problema para duas classes, onde uma apenas é a classe de interesse, pode melhorar significativamente os resultados. É especialmente importante entender se as demais classes de menos interesse podem ser agrupadas em uma única classe, mantendo a generalização do aprendizado.

### 5.3.3 Classificadores binários

O experimento com classificação binária utilizou-se dos mesmos algoritmos testados no experimento de classificação multi-classe. Assim como nos experimentos anteriores, os métodos de aprendizado utilizados nesta abordagem foram treinados e avaliados com o conjunto completo de atributos, bem como com o vetor de atributos com sua dimensionalidade reduzida pela técnica de seleção CFS. O conjunto reduzido de atributos para a abordagem binária pode ser visto na tabela 12.

Atributo	Tipo	Dimensão
Intensidade - histograma (1/15)	Inteiro	1x1
Hough - maior reta	Inteiro	1x1

Tabela 12 – Atributos selecionados pela técnica de CFS para a abordagem de classificação binária.

Para a preparação da base de dados, as amostras previamente rotuladas entre as cinco classes originais do problema foram mapeadas para apenas duas: elementos naturais

e elementos antrópicos. A distribuição das classes nesta versão da base de dados de segmentos é apresentada na tabela 13.

Classe	Amostras	Percentual
Elementos naturais	10.026	99,7 %
Elementos antrópicos	31	0,3 %

Tabela 13 – Distribuição de classes na base de segmentos para classificação binária.

Os algoritmos tiveram seus parâmetros definidos com o objetivo de maximizar os valores de precisão e revocação de todas as classes, medidas em última instância pelo valor da medida F1. Para o algoritmo KNN, um valor de  $k = 3$  e a utilização da distância euclidiana obtiveram os melhores resultados. Para a árvore de decisão, um fator de confiança para poda de 0,4 foi utilizado. Para o algoritmo de *Random Forest*, um número máximo de 200 árvores e um agrupamento de 9 características obtiveram os melhores índices de medida F1. Para o algoritmo SVM um kernel de base radial com  $\gamma = 0.13$  obteve melhor desempenho de precisão e revocação.

Com uma distribuição de classes ainda mais irregular que a base original, resultados gerais de precisão na base inteira se tornam irrelevantes, uma vez que um simples método de aprendizado que sempre escolha a classe majoritária teria taxa de acurácia superior a 99%. A tabela 14 mostra como todos os métodos obtiveram bons resultados gerais, com o melhor resultado para cada critério destacado em coloração cinza.

A causa para estes resultados reside em dois fatores: a desigualdade na distribuição das classes do conjunto de dados e a baixa tolerância dos métodos convencionais de aprendizagem supervisionada à bases de dados fortemente desbalanceadas. Os métodos utilizaram os mesmos parâmetros do experimento anterior, de classificadores multi-classe.

Método	Acurácia	Precisão	Revocação	F1
KNN (CFS)	0.999	0.999	0.999	0.999
Random Forest	0.998	0.999	0.999	0.999
KNN	0.998	0.999	0.999	0.999
Random Forest (CFS)	0.998	0.998	0.998	0.998
Árvore de decisão	0.998	0.999	0.999	0.999
Árvore de decisão (CFS)	0.998	0.999	0.999	0.999
SVM (CFS)	0.998	0.999	0.999	0.999
SVM	0.997	0.995	0.997	0.996
Naive Bayes (CFS)	0.972	0.997	0.973	0.984
Naive Bayes	0.939	0.997	0.939	0.966

Tabela 14 – Comparação de métodos de classificação binária para regiões segmentadas das imagens, ordenados por acurácia.

Em um caso como este, foi ainda mais importante considerar os resultados de precisão e revocação da classe de interesse (elementos antrópicos). A tabela 15 exhibe os

resultados da medida F1, calculada a partir da média harmônica entre precisão e revocação da classe de elementos antrópicos, medida escolhida para selecionar os melhores resultados para este experimento. O melhor resultado para cada critério se encontra destacado em coloração cinza

<b>Método</b>	<b>Precisão</b>	<b>Revocação</b>	<b>F1</b>	<b>Área ROC</b>
KNN (CFS)	1.000	0.615	0.762	0.867
Random Forest	0.941	0.615	0.744	0.979
Árvore de decisão (CFS)	0.941	0.615	0.744	0.694
SVM (CFS)	0.889	0.615	0.727	0.808
Árvore de decisão	0.842	0.615	0.711	0.802
KNN	0.733	0.654	0.708	0.892
Random Forest (CFS)	0.727	0.615	0.667	0.903
Naive Bayes (CFS)	0.075	0.846	0.137	0.933
Naive Bayes	0.039	0.962	0.076	0.974
SVM	0.030	0.591	0.057	0.806

Tabela 15 – Comparação de métodos de classificação binária em relação à classe de elementos antrópicos, ordenados pela medida F1.

Neste experimento, a seleção de atributos através do método CFS obteve bons resultados, reduzindo a dimensionalidade e melhorando a revocação e precisão em quase todos os casos, com exceção do algoritmo *Random Forest*. Os melhores métodos deste experimento não obtiveram desempenho significativamente inferior ao experimento de classificadores multi-classe.

Considerando os resultados para a classe de elementos antrópicos, o método KNN com seleção de atributos por CFS obteve os melhores resultados gerais graças a uma precisão de 100%, embora a revocação tenha sido relativamente baixa. O algoritmo Naive Bayes obteve a melhor revocação entre as classes, mas obteve muitos falsos positivos, fazendo com que a precisão fosse abaixo de 10%, tanto com a seleção de atributos quanto com o vetor de características inteiro.

Com exceção do algoritmo Naive Bayes e do SVM sem seleção de atributos, todos os métodos obtiveram medida F1 acima de 60% para a classe de elementos antrópicos. Isto pode ser considerado um bom resultado, visto que o esperado era que a capacidade de generalização da classe majoritária fosse prejudicada pela diversidade das amostras, que verdadeiramente não pertencem à mesma classe.

Ainda assim, o uso de métodos de aprendizado que não foram concebidos para lidar com uma base de dados com distribuição de classes tão desigual não é o cenário ideal. Por este motivo, experimentos utilizando algoritmos que levem em consideração a detecção de classes extremamente minoritárias devem ser realizados.

### 5.3.4 Classificadores unários

De forma muito semelhante ao experimento de classificadores binários, a base de dados para o experimento de classificadores unários foi criada a partir da base original com cinco classes rotuladas, mapeadas para apenas duas: elementos naturais (*target*) e elementos antrópicos (*outlier*). A distribuição de amostras da base de dados entre as classes também é a mesma do experimento de classificadores binários, e pode ser vista na tabela 16.

Classe	Amostras	Percentual
Elementos naturais	10.026	99,7 %
Elementos antrópicos	31	0,3 %

Tabela 16 – Distribuição de classes na base de segmentos para classificação unária.

Assim como nas duas abordagens de classificação supervisionada anteriores, este experimento utilizou a base completa de segmentos. Os algoritmos de classificação unária foram treinados e avaliados utilizando o conjunto integral de atributos, e também um subconjunto de atributos selecionados pela técnica CFS. O conjunto de atributos selecionados para este experimento pode ser visto na tabela 17.

Atributo	Tipo	Dimensão
Intensidade - histograma (1/15)	Inteiro	1x1
Hough - maior reta	Inteiro	1x1

Tabela 17 – Atributos selecionados pela técnica de CFS para a abordagem de classificação unária.

Os dois algoritmos testados neste experimento foram a variação uniclasse do SVM (OC-SVM) e uma versão unária da árvore de decisão chamada REPTree. Ambos os algoritmos foram executados tendo como entrada o vetor de características das amostras completo e posteriormente o vetor reduzido, selecionado através do método CFS de seleção de características.

O algoritmo OC-SVM obteve melhores resultados utilizando um kernel de base radial com  $\gamma = 0.18$ , enquanto o algoritmo REPTree obteve as melhores taxas de medida F1 utilizando um mínimo de 2 instâncias por folha e redução de erro de poda com 3 passadas.

A tabela 18 mostra uma evidente diferença na taxa de aprendizado entre os métodos OC-SVM e REPTree, onde o último obteve resultados gerais visivelmente inferiores. O bom desempenho do OC-SVM se mantém quando analisadas precisão e revocação para a classe de elementos antrópicos, na tabela 19. Ambas as tabelas apresentam o melhor resultado para cada critério destacado em coloração cinza.

Método	Acurácia	Precisão	Revocação	F1
OC-SVM (CFS)	0.998	0.999	0.999	0.999
OC-SVM	0.997	0.995	0.997	0.996
REPTree	0.464	0.996	0.465	0.632
REPTree (CFS)	0.383	0.997	0.384	0.552

Tabela 18 – Comparação de métodos de classificação unária para regiões segmentadas das imagens, ordenados por acurácia.

Método	Precisão	Revocação	F1	Área ROC
OC-SVM (CFS)	0.842	0.615	0.711	0.808
REPTree (CFS)	0.004	0.962	0.008	0.727
REPTree	0.003	0.692	0.007	0.520
OC-SVM	0.000	0.000	0.000	0.000

Tabela 19 – Comparação de métodos de classificação unária em relação à classe de elementos antrópicos, ordenados pela medida F1.

Assim como no experimento de classificação binária, a seleção de atributos com o método CFS foi responsável por visível melhora no resultado do método OC-SVM, o algoritmo de melhor resultado neste experimento. Apesar dos bons resultados, os resultados deste experimento foram inferiores aos experimentos anteriores. A hipótese é de que as classes de menor interesse (floresta, vegetação rasteira, água e terra) são diversas demais para que um único modelo de aprendizagem possa distinguí-las das amostras anômalas (elementos antrópicos). Com a intenção de separar o modelo de aprendizagem para estas classes e ainda assim obter robustez na detecção de amostras anômalas, experimentos foram conduzidos com conjuntos de classificadores unários.

### 5.3.5 Conjunto de classificadores unários

Para este experimento, a base de dados rotulada originalmente foi dividida e as amostras foram separadas de acordo com sua classe. Dois algoritmos classificadores unários foram treinados exclusivamente com exemplos de cada uma das quatro classes majoritárias, excluindo as amostras das demais classes, totalizando oito modelos de classificação unária a serem combinados entre si e avaliados neste experimento.

Como forma de utilizar um conjunto de classificadores unários para resolver um problema de classificação multi-classe, um método de votação com pesos foi utilizado. O peso de cada classificador foi estimado através de um método força-bruta. Os pesos variam de 0 a 1, com intervalos de 0,1, com todos os pesos somando um total de 1. À medida que os pesos foram alterados, as taxas de aprendizado foram comparadas. Ao final deste processo, um vetor de pesos foi determinado de forma a otimizar as taxas de aprendizado e a medida F1 da classe de elementos antrópicos.

Assim como nas abordagens de classificação supervisionada anteriores, este experi-

mento utilizou a base completa de segmentos. Os algoritmos de classificação unária foram treinados e avaliados utilizando o conjunto integral de atributos, e também um subconjunto de atributos selecionados pela técnica CFS. O conjunto de atributos selecionados para este experimento para cada classe individualmente treinada pode ser visto na tabela 20.

Classe majoritária	Atributo	Tipo	Dimensão
Floresta	Intensidade - histograma (5/15)	Inteiro	7x1
	LBP - histograma (1/26)	Inteiro	1x1
	Hough - maior reta	Inteiro	1x1
Água	Intensidade - histograma (7/15)	Inteiro	7x1
	LBP - histograma (2/26)	Inteiro	2x1
	Hough - número de retas	Inteiro	1x1
Vegetação	Intensidade - histograma (6/15)	Inteiro	7x1
	LBP - histograma (1/26)	Inteiro	1x1
	Hough - número de retas	Inteiro	1x1
Terra	Intensidade - histograma (7/15)	Inteiro	7x1
	LBP - histograma (1/26)	Inteiro	1x1
	Hough - número de retas	Inteiro	1x1
	Hough - maior reta	Inteiro	1x1

Tabela 20 – Atributos selecionados pela técnica de CFS para a abordagem de classificação unária.

O conjunto de classificadores foi escolhido, em um primeiro momento, de acordo com a precisão de cada modelo treinado para uma classe específica. A tabela 21 exhibe os valores de acurácia, precisão e revocação para cada classe do problema em cada um dos algoritmos de classificação unária testados neste experimento.

Os valores de precisão, revocação e medida F1 para a classe de elementos antrópicos, *outlier* em todos os classificadores deste experimento, estão apresentados na tabela 22, com o melhor resultado para cada critério destacado em coloração cinza.

A combinação de todos os métodos de aprendizado testados neste experimento, somados à variante da redução do conjunto de características, totalizam 256 combinações possíveis. Todos os experimentos foram executados e o resultado de todas estas combinações foram colhidos. Por uma questão de clareza e brevidade, apenas os 6 conjuntos de classificadores com melhor desempenho de medida F1 foram apresentados neste texto.

Como pode ser visto, ao se comparar as tabelas 21 e 22, o melhor conjunto de classificadores não é determinado pelos melhores classificadores individuais. Isso pode ser constatado ao perceber que vários dos métodos com melhor desempenho individual não aparecem entre os melhores conjuntos de classificadores. Este fenômeno é considerado uma característica de métodos *ensemble* e é amplamente documentado na literatura de aprendizagem de máquina. A ideia central deste fenômeno é de que o conjunto de classificadores termina por fortalecer classificadores mais simples.

Classe/Algoritmo	Acurácia	Precisão	Revocação	F1
Floresta/OC-SVM	0.942	0.941	0.943	0.939
Floresta/OC-SVM (CFS)	0.942	0.941	0.943	0.939
Floresta/REPTree	0.856	0.847	0.856	0.792
Floresta/REPTree (CFS)	0.805	0.821	0.805	0.812
Água/OC-SVM	0.975	0.971	0.976	0.969
Água/OC-SVM (CFS)	0.508	0.353	0.763	0.482
Água/REPTree	0.538	0.954	0.538	0.675
Água/REPTree (CFS)	0.445	0.952	0.445	0.590
Vegetação/OC-SVM	0.929	0.920	0.930	0.919
Vegetação/OC-SVM (CFS)	0.873	0.433	0.871	0.578
Vegetação/REPTree	0.538	0.837	0.539	0.633
Vegetação/REPTree (CFS)	0.338	0.887	0.333	0.407
Terra/OC-SVM	0.986	0.983	0.987	0.981
Terra/OC-SVM (CFS)	0.632	0.954	0.453	0.614
Terra/REPTree	0.305	0.980	0.305	0.452
Terra/REPTree (CFS)	0.363	0.979	0.363	0.519

Tabela 21 – Acurácia, precisão, revocação e medida F1 para as classes e algoritmos do experimento de conjunto de classificadores unários, sem ordenação específica.

Floresta	Água	Vegetação	Terra	Precis.	Revoc.	F1
OC-SVM	OC-SVM (CFS)	REPTree	OCSVM	0.947	0.692	0.800
REPTree	OC-SVM	REPTree	REPTree	0.850	0.730	0.785
REPTree	OC-SVM	OC-SVM	REPTree (CFS)	0.809	0.653	0.723
OC-SVM	REPTree (CFS)	OC-SVM	OC-SVM	0.239	0.653	0.350
OC-SVM	OC-SVM	REPTree	OC-SVM	0.156	0.769	0.259
OC-SVM	REPTree	REPTree (CFS)	OC-SVM	0.153	0.653	0.248

Tabela 22 – Comparação dos 6 melhores conjuntos de classificadores unários em relação à classe de elementos antrópicos, ordenados pela medida F1.

O conjunto de pesos utilizados para os experimentos foi de 0,5 para o classificador da classe de floresta, 0,3 para a classe água, 0,1 para vegetação rasteira e 0,1 para a classe de terra. Os pesos foram multiplicados ao valor de confiança de classificação de cada algoritmo e o maior valor determinou a classe da amostra. Para que uma amostra pudesse ser considerada uma anomalia, o resultado de todas as classificações precisava ser 0.

Embora o conjunto de classificadores com melhor desempenho tenha apresentado o melhor resultado para classificação de amostras de água ao utilizar um subconjunto dos atributos, selecionados através da técnica CFS, pode-se dizer que esta seleção foi pouco eficaz, visto que os 6 melhores resultados possuem pouca presença de algoritmos que utilizam um subconjunto das características.

Uma vez que todos os experimentos foram concluídos e os seus dados colhidos, a próxima seção se ocupa em discutir os resultados.

## 5.4 Discussão

Dentre as técnicas de segmentação analisadas, o método SRM, que realiza o agrupamento de pixels em uma região de acordo com a similaridade da intensidade e cor dos pixels de regiões vizinhas, obteve erros de consistência global e local inferiores aos demais. É importante notar também que o algoritmo MSEG, que se utiliza de informações de cor e textura, obteve desempenho aproximado ao do SRM, mas com o tempo de execução cerca de quatorze vezes menor. Isto torna o algoritmo MSEG um forte candidato a substituir a técnica SRM em uma aplicação prática, em que o tempo de processamento da segmentação seja um fator de impacto e um pouco da precisão possa ser descartado.

Para a classificação dos segmentos e, por consequência, a detecção de elementos anômalos nas imagens, quatro abordagens de modelagem do problema de aprendizagem de máquina foram avaliadas.

A primeira delas utiliza classificadores multi-classe em uma base com rótulos de cinco classes (floresta, vegetação rasteira, água, terra e elementos antrópicos). Neste experimento o algoritmo *K-Nearest Neighbor* (KNN) com  $k = 3$  obteve os melhores resultados de todo o trabalho. Apesar dos bons resultados, técnicas de aprendizado preguiçoso (*lazy learning*) como o KNN não criam um modelo de aprendizado sucinto, necessitando de boa parte das amostras de treinamento para realizar a classificação de novas amostras, o que pode ser problemático em um ambiente real, com limitações de tempo de processamento e memória disponível.

Adicionalmente, um problema da abordagem com classificadores multi-classe é a pouca generalização do aprendizado sobre elementos antrópicos, visto que se trata de uma classe pouco previsível. Como os elementos antrópicos fazem parte do modelo de treinamento, novas amostras de objetos antrópicos cujas características se distanciem das amostras de elementos antrópicos da base de treinamento tendem a reduzir a precisão do modelo.

A segunda abordagem analisada, que utiliza classificadores binários, obteve desempenho acima do esperado. A expectativa era de que os modelos gerados por esta abordagem tivessem um baixo desempenho na classificação da classe de elementos antrópicos, uma vez que esta correspondia a apenas 0.3% da base de dados. Esperava-se também que, por conta da re-classificação das amostras de segmentos que reagrupou as classes de floresta, vegetação rasteira, água e terra em uma única classe, o modelo gerado fosse pouco descritivo, hipótese que não foi comprovada.

Neste experimento, o algoritmo KNN também obteve o melhor desempenho para a classe de elementos antrópicos, desta vez se beneficiando da seleção de características feita através da técnica CFS, que utiliza a correlação entre os atributos do problema para reduzir a dimensão do vetor de características. Os resultados indicam que a seleção

de atributos contribuiu significativamente para a simplificação do problema, tornando o classificador mais generalista e mais apto a descrever a classe de elementos naturais, resultado inverso ao experimento com classificadores multi-classe.

Na terceira abordagem, pudemos avaliar o desempenho de alguns algoritmos de classificação unária, também conhecidos como detectores de anomalias (*outliers*). Assim como na abordagem de classificação binária, aqui foi preciso agrupar as amostras das classes majoritárias em uma única classe de elementos naturais, agravando o desbalanceamento das classes na base de treinamento.

Os resultados foram bastante similares à abordagem binária, com visível melhora dos resultados quando se usa a seleção de atributos através do método CFS. No entanto, ao contrário da abordagem binária, classificadores unários tendem a responder melhor quando amostras não-usuais da classe de *outliers* aparecem.

A última abordagem analisada neste trabalho se utiliza de um conjunto de classificadores unários. Uma vez que cada classificador é responsável por aprender e criar um modelo para reconhecer uma classe do problema, esta estratégia permitiu que modelos mais especializados fossem criados e que diferentes maneiras de chegar a um consenso sobre a classificação das amostras fossem testadas.

Com bons resultados desde o início dos experimentos, o conjunto de classificadores unários se mostra uma abordagem com boa conciliação entre a capacidade de classificar corretamente as amostras de interesse e a capacidade de generalização e detecção de elementos antrópicos ausentes da base de treinamento.

Como métrica para avaliação das abordagens de aprendizado ficou definido que a medida  $F1$ , que leva em consideração um compromisso entre precisão e revocação, seria utilizada para julgar o desempenho dos algoritmos. Por conta da importância da classe, bem como forma de compensar o desequilíbrio na distribuição das classes do problema na base de dados de treinamento, a medida  $F1$  da classe de elementos antrópicos foi o critério principal na avaliação dos métodos.

Utilizando este critério para ranquear as abordagens de aprendizado de acordo com seus respectivos melhores resultados, podemos afirmar que o melhor desempenho foi obtido pelos métodos KNN como classificador de um problema multi-classe e por uma combinação de métodos unários, conforme exibido na tabela 23.

Considerando que existem inúmeros exemplos de elementos antrópicos possíveis em imagens da categoria utilizada neste trabalho e que a base de treinamento possui quantidade e variedade limitadas destes mesmos elementos, é razoável supor que uma abordagem que produza um modelo capaz de reconhecer elementos naturais e não pressupõe características específicas para elementos antrópicos, possa ter melhor desempenho em aplicações reais.

<b>Abordagem</b>	<b>Precisão</b>	<b>Revocação</b>	<b>F1</b>
Classificadores multi-classe (KNN)	0.889	0.727	0.800
Conjunto de classificadores unários	0.947	0.692	0.800
Classificadores binários (KNN)	1.000	0.615	0.762
Classificadores unários (OC-SVM)	0.842	0.615	0.711

Tabela 23 – Comparação do desempenho das abordagens de aprendizagem de máquina para a classe de elementos antrópicos, ordenados pela medida F1.

Considerando esse cenário, é válido afirmar que o conjunto de classificadores unários é a abordagem mais promissora dentre as avaliadas para resolver o problema da detecção de elementos antrópicos em imagens aéreas da floresta amazônica. Mesmo que o algoritmo KNN em uma abordagem multi-classe tenha se saído igualmente bem nos experimentos deste trabalho, é esperado que este modelo de aprendizado obtenha pior desempenho em conjuntos de imagens com novos tipos de elementos antrópicos.

Com uma precisão superior a 85% e baixa taxa de falsos positivos para a classe de elementos antrópicos, é possível considerar que a abordagem que utiliza conjunto de classificadores unários atinge o objetivo geral deste trabalho: realizar a identificação automática de elementos antrópicos com elevadas taxas de acerto.

Todos os conjuntos de dados, implementações dos algoritmos, ferramentas de apoio e avaliação dos resultados deste trabalho estão publicamente disponíveis<sup>3</sup> e podem ser verificados, reutilizados e reproduzidos por trabalhos futuros na mesma temática ou em áreas relacionadas.

Após esta discussão de resultados, o próximo capítulo se encarrega de tecer conclusões sobre o trabalho como um todo, revisar seus objetivos e sugerir pesquisas futuras sobre o tema.

<sup>3</sup> <https://github.com/luizcavalcanti/ForestClassifier>

## 6 Conclusão e Trabalhos Futuros

Este trabalho teve como objetivo propor uma abordagem eficiente, utilizando processamento digital de imagens e aprendizagem de máquina, para o problema de detecção automática de elementos antrópicos em imagens aéreas da floresta amazônica. A intenção era demonstrar que a identificação automática destes elementos pode trazer benefícios, em especial agilidade, ao processo que hoje é completamente dependente de agentes humanos, e produzir elevadas taxas de acerto.

Em um primeiro momento, uma base de dados foi criada, rotulada e disponibilizada publicamente para que experimentos pudessem ser executados, e também que pudesse servir de referência para futuros estudos nesta área de pesquisa e em áreas relacionadas.

Posteriormente, investigamos e definimos métodos de segmentação de imagens adequados à este tipo de imagem e problema. O algoritmo SRM obteve o melhor desempenho e foi indicado para esta etapa da solução. É importante ressaltar que outros algoritmos de segmentação como o MSEG também obtiveram resultados promissores, e não devem ser descartados em problemáticas similares.

Em seguida, características foram extraídas das imagens, para que um vetor de atributos pudesse ser usado na etapa de classificação dos segmentos. Elementos de baixo nível como cor e textura foram mesclados à características de alto nível como textura, para representar as amostras de segmentos de imagens aéreas. A técnica de seleção de atributos CFS foi utilizada para escolher um subconjunto de atributos a ser utilizado em experimentos específicos, muitas vezes com resultados melhores ao conjunto completo de atributos.

Por último, quatro abordagens de aprendizagem de máquina supervisionada foram avaliadas: classificação multi-classe, binária, unária e conjuntos de classificadores unários. Esta última abordagem obteve o melhor desempenho, visto que teve boa taxa de aprendizado geral e elevada taxa de detecção da classe de elementos antrópicos, assim como tem um bom potencial de generalização da capacidade de detectar novos elementos antrópicos, não presentes na base utilizada neste trabalho.

Todos os conjuntos de dados, implementações dos algoritmos, ferramentas de apoio e avaliação dos resultados deste trabalho estão publicamente disponíveis<sup>1</sup> e podem ser verificados, reutilizados e reproduzidos por trabalhos futuros na mesma temática ou em áreas relacionadas.

---

<sup>1</sup> <https://github.com/luizcavalcanti/ForestClassifier>

## 6.1 Trabalhos futuros

Para o desenvolvimento mais aprofundado de uma solução para a problemática proposta, é preciso continuar explorando o uso de conjuntos de classificadores unários, em especial métodos de *ensemble* de classificadores, com a intenção de aperfeiçoar o método de determinação da classificação final deste conjunto de modelos de aprendizado.

Além disso, é importante que outros algoritmos de classificação unária sejam experimentados, visto que apenas dois, OC-SVM e REPTree, foram utilizados nos experimentos. Testar outros algoritmos pode trazer ganhos na taxa de aprendizado e generalização do modelo gerado.

Também se faz necessária a ampliação do conjunto de imagens disponível para o problema. Embora a base de dados utilizada neste trabalho conte com uma boa quantidade de imagens aéreas da região de floresta amazônica, é preciso incrementar o número de amostras com presença de elementos antrópicos, reduzindo o desequilíbrio na distribuição das classes. Se possível, novos tipos de elementos antrópicos precisam ser inseridos.

# Referências

- AHMADI, P. A simple and efficient method for segmentation and classification of aerial images. In: *Image and Signal Processing (CISP), 2013 6th International Congress on*. [S.l.: s.n.], 2013. v. 01, p. 566–570. Citado na página 34.
- AHONEN, T. et al. Rotation invariant image description with local binary pattern histogram fourier features. In: *Image Analysis: 16th Scandinavian Conference, SCIA 2009, Oslo, Norway, June 15-18, 2009. Proceedings*. [S.l.: s.n.], 2009. p. 61–70. Citado 2 vezes nas páginas 34 e 50.
- ALPAYDIN, E. *Introduction to Machine Learning*. 2nd. ed. [S.l.]: The MIT Press, 2010. ISBN 026201243X, 9780262012430. Citado na página 13.
- ARBELAEZ, P. et al. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Washington, DC, USA, v. 33, n. 5, p. 898–916, maio 2011. ISSN 0162-8828. Citado 3 vezes nas páginas 26, 31 e 32.
- AZEVEDO, A. R. D. et al. Caracterização de uso e cobertura da terra na Amazônia utilizando imagens duais multitemporais do COSMO-SkyMed. *Acta Amazonica*, v. 44, n. 1, p. 87–98, 2014. ISSN 00445967. Citado na página 25.
- BREIMAN, L. Random forests. *Machine Learning*, Kluwer Academic Publishers, v. 45, n. 1, p. 5–32, 2001. ISSN 0885-6125. Citado na página 20.
- CAVALCANTI, L. C. A. M. *Amazon rainforest image segmentation experiment*. 2015. Disponível em: <<http://amazonsegmentation.ddns.net/>>. Citado na página 45.
- CAVALCANTI, L. C. A. M.; CARVALHO, J. R. H.; SANTOS, E. M. d. A comparison on supervised machine learning classification techniques for semantic segmentation of aerial images of rain forest regions. In: *Computer Vision Theory and Applications, 2015. VISAPP 2015. Proceedings. Eighth IEEE International Conference on*. [S.l.: s.n.], 2015. v. 2, p. 416–423 vol.2. Citado na página 47.
- 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. Citado 2 vezes nas páginas 28 e 29.
- COVER, T.; HART, P. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, v. 13, n. 1, p. 21–27, 1967. ISSN 0018-9448. Citado na página 17.
- DENG, Y.; MANJUNATH, B. Unsupervised segmentation of color-texture regions in images and video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 23, n. 8, p. 800–810, Aug 2001. ISSN 0162-8828. Citado 2 vezes nas páginas 26 e 27.
- DUBUISSON-JOLLY, M.-P.; GUPTA, A. Color and texture fusion: application to aerial image segmentation and GIS updating. *Image and Vision Computing*, v. 18, n. 10, p. 823–832, 2000. ISSN 0262-8856. Citado 2 vezes nas páginas 32 e 33.

- DUDA, R. O.; HART, P. E. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, ACM, v. 15, n. 1, p. 11–15, 1972. Citado na página 10.
- ESPÍRITO-SANTO, F. D. B.; SHIMABUKURO, Y. E. Validação do mapeamento de uma área de floresta tropical com o uso imagens de videografia aérea e dados de levantamento de campo. *Revista Árvore*, v. 29, n. 2, p. 227–239, 2005. ISSN 0100-6762. Citado na página 25.
- FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *International Journal of Computer Vision*, Springer, v. 59, n. 2, p. 167–181, 2004. Citado na página 28.
- FERNANDES, C. da E. *Classificação de imagens de sensoriamento remoto com área desmatada*. Dissertação (Mestrado) — Universidade Federal do Amazonas, Manaus, 2008. Citado 2 vezes nas páginas 33 e 36.
- FERREIRA, L. V. Classificação da cobertura da terra na Amazônia com base em imagens de satélite e caracterização das classes com relação à superfície do terreno. *Geografia (Londrina)*, v. 21, n. 3, p. 115–131, 2012. Citado na página 25.
- FILHO, R. A.; NASCIMENTO, P. S. R.; BATISTA, G. T. Avaliação de técnicas de segmentação e classificação automática de imagens landsat-tm no mapeamento do uso do solo na Amazônia. *Acta Amazonica*, v. 28, n. 1, p. 41–41, mar 1998. ISSN 1809-4392. Citado na página 25.
- FOGEL, I.; SAGI, D. Gabor filters as texture discriminator. *Biological Cybernetics*, Springer-Verlag, v. 61, n. 2, p. 103–113, 1989. ISSN 0340-1200. Citado na página 34.
- GHIASI, M.; AMIRFATTAHI, R. Fast semantic segmentation of aerial images based on color and texture. In: *Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on*. [S.l.: s.n.], 2013. p. 324–327. ISSN 2166-6776. Citado 2 vezes nas páginas 34 e 35.
- GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. [S.l.]: Prentice hall Upper Saddle River, NJ:, 2002. Citado 5 vezes nas páginas 6, 7, 8, 11 e 16.
- HALL, M. A. *Correlation-based Feature Subset Selection for Machine Learning*. Tese (Doutorado) — University of Waikato, Hamilton, New Zealand, 1998. Citado 2 vezes nas páginas 16 e 50.
- HEGENBART, S.; UHL, A.; VÉCSEI, A. On the implicit handling of varying distances and gastrointestinal regions in endoscopic video sequences with indication for celiac disease. In: *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*. [S.l.: s.n.], 2012. p. 1–6. ISSN 1063-7125. Citado na página 35.
- HOUGH, P. V. *Method and means for recognizing complex patterns*. [S.l.], 1962. Citado na página 10.
- IBGE. *Instituto Brasileiro de Geografia e Estatística*. 2015. Disponível em: <<http://www.ibge.gov.br>>. Citado na página 3.
- INPE. *Projeto GEOMA*. 2013. Disponível em: <<http://www.geoma.lncc.br/>>. Citado na página 43.

JAIN, A. K. *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989. ISBN 0-13-336165-9. Citado na página 15.

JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. In: *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann, 1995. p. 338–345. Citado na página 19.

LATORRE, M. L. et al. Integração de dados de sensoriamento remoto multi resoluções para a representação da cobertura da terra utilizando campos contínuos de vegetação e classificação por árvores de decisão. *Revista Brasileira de Geofísica*, v. 25, n. 1, p. 1689–1699, mar 2007. ISSN 0102-261X. Citado na página 25.

LEVINSHTEIN, A. et al. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 31, n. 12, p. 2290–2297, Dec 2009. ISSN 0162-8828. Citado na página 34.

LINHARES, M. M. et al. Vant's aplicados ao monitoramento da evolução vegetativa de culturas agrícolas. *Congresso Brasileiro de Agricultura de Precisão*, 2014. Citado na página 36.

LU, D. et al. Land use/cover classification in the Brazilian Amazon using satellite images. *Pesquisa Agropecuária Brasileira*, v. 47, n. 9, p. 1185–1208, 2012. ISSN 0100-204x. Citado na página 25.

MARTIN, D. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.: s.n.], 2001. v. 2, p. 416–423 vol.2. Citado 3 vezes nas páginas 13, 45 e 46.

MARTIN, D. R.; FOWLKES, C. C.; MALIK, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 5, p. 530–549, May 2004. ISSN 0162-8828. Citado na página 31.

MITCHELL, T. M. Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, v. 45, 1997. Citado na página 19.

MUNOZ-MARI, J. et al. Combination of one-class remote sensing image classifiers. In: *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. [S.l.: s.n.], 2007. p. 1509–1512. Citado na página 33.

NIXON, M.; AGUADO, A. *Feature extraction & image processing*. [S.l.]: Academic Press, 2008. Citado 3 vezes nas páginas 8, 10 e 16.

NOCK, R.; NIELSEN, F. Statistical region merging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 26, n. 11, p. 1452–1458, 2004. Citado 2 vezes nas páginas 29 e 30.

OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision amp; Image Processing., Proceedings of the 12th IAPR International Conference on*. [S.l.: s.n.], 1994. v. 1, p. 582–585 vol.1. Citado na página 9.

- PLA, F.; CARMONA, P. L.; SOTOCA, J. M. One-class classification techniques in image recognition problems. In: *Information Optics (WIO), 2013 12th Workshop on*. [S.l.: s.n.], 2013. p. 1–3. Citado na página 35.
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. [S.l.]: Prentice Hall, 2010. (Prentice Hall series in artificial intelligence). ISBN 9780136042594. Citado 2 vezes nas páginas 14 e 19.
- SADGAL, M.; FAZZIKI, A. E.; OUAHMAN, A. A. Aerial image processing and object recognition. *The Visual Computer*, Springer, v. 21, n. 1-2, p. 118–123, 2005. Citado na página 33.
- SANTOS, J. R. dos; MALDONADO, F. D.; GRAÇA, P. M. L. A. Integração de imagens landsat/etm+ e cbers-2/ccd para detecção de mudanças em área da amazônia sob domínio da floresta de transição. *Revista Brasileira de Cartografia*, v. 1, n. 57, 2005. Citado na página 36.
- SCHÖLKOPF, B. et al. Support vector method for novelty detection. *NIPS*, v. 12, p. 582–588, 1999. Citado na página 21.
- SILVA, E. T. d. J. B. Veículos aéreos não tripulados: panorama atual e perspectivas para o monitoramento de atividades ilícitas na amazônia. In: *Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 16, Anais*. [S.l.: s.n.], 2013. p. 9324–9331. ISBN 978-85-17-00066-9. Citado na página 3.
- VAPNIK, V.; CORTES, C. Support-vector networks. *Machine Learning*, Kluwer Academic Publishers, v. 20, n. 3, p. 273–297, 1995. ISSN 0885-6125. Citado 2 vezes nas páginas 18 e 21.
- VASCONCELOS, C. H.; NOVO, E. M. L. D. M. Mapeamento do uso e cobertura da terra a partir da segmentação e classificação de imagens – fração solo, sombra e vegetação derivadas do modelo linear de mistura aplicado a dados do sensor TM / Landsat5, na região do reservatório de Tucuruí - PA. *Instituto Nacional de Pesquisas da Amazônia - ACTA Amazônica*, v. 34, n. 3, p. 487–493, 2001. ISSN 0044-5967. Citado na página 25.
- WANG, X. et al. A one-class classification by spatial-contextual for remotely sensed image. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*. [S.l.: s.n.], 2013. p. 437–440. ISSN 2153-6996. Citado na página 36.
- YADAV, A.; YADAV, P. *Digital Image Processing*. [S.l.]: Laxmi Publications Pvt Limited, 2009. ISBN 9788131805244. Citado na página 6.
- YUAN, J.; GLEASON, S. S.; CHERIYADAT, A. M. Systematic benchmarking of aerial image segmentation. *IEEE*, 2013. Citado 5 vezes nas páginas ix, 26, 31, 32 e 38.
- YUAN, J.; WANG, D. Factorization-based texture segmentation. *Dept. Comput. Sci. Eng., The Ohio State Univ., Columbus, OH, USA, Tech. Rep. OSU-CISRC-1/13-TR01*, 2013. Citado na página 30.

# Apêndices

# A comparison on supervised machine learning classification techniques for semantic segmentation of aerial images of rain forest regions

Luiz Carlos A. M. Cavalcanti<sup>1,2</sup>, Jose Reginaldo Hughes Carvalho<sup>2</sup> and Eulanda Miranda dos Santos<sup>2</sup>

<sup>1</sup>*Nokia Institute of Technology, Manaus, Brazil*

<sup>2</sup>*Universidade Federal do Amazonas, Manaus, Brazil*

*luiz.cavalcanti@indt.org.br, reginaldo@icomp.ufam.edu.br, emsantos@icomp.ufam.edu.br*

**Keywords:** Semantic Segmentation, Image Processing, Machine Learning, Classification

**Abstract:** Segmentation is one of the most important operations in Computer Vision. Partition of the image in several domain-independent components is important in several practical machine learning solutions involving visual data. In the specific problem of finding anomalies in aerial images of forest regions, this can be specially important, as a multilevel classification solution can demand that each type of terrain and other components of the image are inspected by different classification algorithms or parameters. This work compares several common classification algorithms and assess their reliability on segmenting aerial images of rain forest regions as a first step into a multi-level classification solution. Finally, we draw conclusions based on the experiments using real images from a publicly available dataset, comparing the results of those classification algorithms for segmenting this kind of images.

## 1 INTRODUCTION

One of the most important operations in Computer Vision is segmentation. The aim of image segmentation is the domain-independent partition of the image into a set of regions which are visually distinct and uniform with respect to some property, such as grey level, texture or colour (Freixenet et al., 2002). Semantic segmentation aims at pixel-wise classification of images into high-level components and is needed in challenging tasks such as remote sensing, driving assistance systems or precise object localisation in general. The input image is divided into regions, which correspond to the high-level objects of the scene (Heitz and Koller, 2008).

The global community have spent a reasonable attention on the deforestation records of the Amazon rainforest. However, the Amazon region is threatened by agents that take advantage of its dimension (larger than most countries on earth). So far, most of the academic efforts are concentrated on statistics and modelling of the ecosystem behaviour. This work, on the other hand, is part of an effort to move from statistics into action. Our goal to equip the preservation agents and authorities with methodologies and tools that enables them to take immediate action, so that the deforestation in course can be stopped, not only

measured. A critical success factor of any tool is its ability to automatic acquire and classify sensory data, specially imagery information taken from aerial aircrafts (manned or unmanned). It is important to mention that the project's final goal is not to provide a fully automatic classification of aerial forest images, but, rather, a supporting decision making tool to dramatically reduce the number of images to be analysed by any human being from thousands to a few dozen of candidate images. This work is about how the segmentation process of Amazon's rain forest aerial images could be done, before any relevant classification of anomalies on those images.

In aerial images (orthographic photos), different types of terrain have different characteristics that are not always easily separable by a single segmentation criteria. Color or multi-spectral information is very useful to the segmentation of aerial images, but most existing methods which use multi-spectral features alone tend to produce very noisy segmentation maps (Dubuisson-Jolly and Gupta, 2000). If texture features are used alone, for example, the localisation of the region boundaries is not very accurate.

This is specially true when low-resolution images are the only ones available. Many are the issues that could degenerate an aerial image quality and

cause some impact on a classification algorithm performance: water bodies reflecting the sunlight can cause saturation (fig. 1) and regions shadowed by clouds can darken a region's colour (fig. 2).



Figure 1: Sun's reflex on water causing saturation (white patch in lower right)



Figure 2: Clouds shadowing a patch of forest (the darker lower part), which typically causes segmentation problems

As seen in fig 3(a), a single criteria segmentation can present good results to certain images, but very bad to others. In the given example, Otsu's threshold segmentation method (Otsu, 1979) with two thresholds fails to provide the desired number of connected components (fig. 3(b)), finding too many high level components (fig. 3(c)) and compromising the classification steps afterwards. A similar problem happens when we apply Watershed segmentation which, as pointed out by (Li and Wan, 2010), often produces over-segmented regions due to image noise and detail information (fig. 3(d)).

As part of an effort to find anomalies in aerial images of rain forest regions such as human-made structures and environmental menaces, we surveyed segmentation techniques that could separate the image

in its high-level components (i.e.: vegetation, water bodies, human-made objects, etc) and provided an acceptable semantical segmentation of the images. This necessity comes from the idea that everything that is not common to the region's scenery (something that is neither vegetation nor water body) is of potential interest.

We compare several machine learning techniques that use combined criteria (texture, colour, morphology, etc.) to provide a more accurate segmentation for our needs. The same set of publicly available images is used on all techniques investigated in this work, so they could be compared to each other and with an ideal segmentation that is expected by a specialist.

This work is organised as follows. Section 2 discusses related work both in single-criteria and multi-criteria techniques. Section 3 details how the dataset is organised, which features were extracted, why they matter to the problem, which classification algorithms were used and what was the overall workflow of the experiment. Section 4 discusses the results of the experiment. Finally, section 5 draw conclusions on the experiment and discusses possible future works on the subject.

## 2 RELATED WORK

(Pal and Pal, 1993) do a thorough survey on single-criteria techniques for images segmentation, most of them implemented by this work with poor results due to the complexity of textures, colours and morphology on the dataset used on this our work.

In his work, Dubuisson-Jolly (Dubuisson-Jolly and Gupta, 2000) segments aerial images using multi-criteria segmentation. They create two distinct segmentations, by colour and by texture, and finally fuses both using a Maximum Likelihood algorithm. The outcome is a single segmentation for the image based on the fusion of both criteria.

(Freixenet et al., 2002) survey several methods of segmentation that integrate boundary and region information, exploring embedded integration and post-processing integration of segmentation criteria. All experiments were done in both real and synthetic images, finding very similar results. (Freixenet et al., 2002) concluded that post-processing methods have better results than embedded ones.

(Bosch et al., 2007) use Random Forest algorithm to segment images and compared it to multi-class Support Vector Machines, with comparable performance on the problem of categorising objects in scenes. The main advantage of Random forest over multi-class SVM, the authors argue, is the simplicity

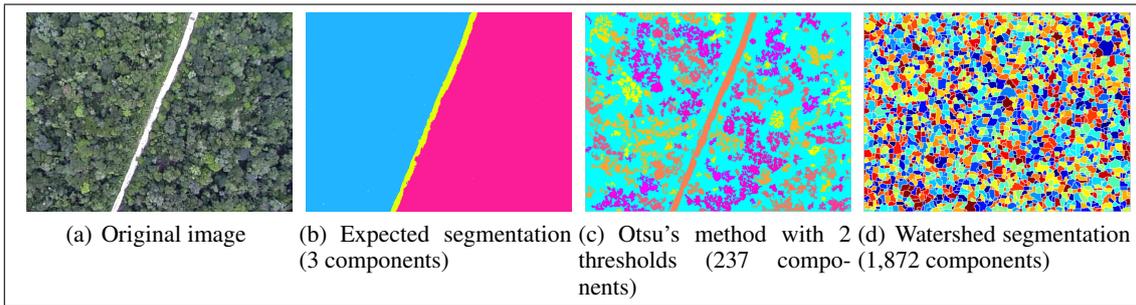


Figure 3: Single-criteria segmentation techniques in this work's dataset

of training and test costs.

(Ghiasi and Amirfattahi, 2013) rely on colour and texture information to classify forest region images using two separate KNN classifiers. New images are broke down into super-pixels and then features are extracted. Using RGB channels and Local Binary Pattern Histogram Fourier features (LBP-HF) (T. et al., 2009), they achieved a success rate over 95% on the test dataset.

Our work combines several features extracted by the aforementioned works and experiment on several commonly used classification algorithms to find a better technique for segmenting aerial images of Amazon rain forest regions.

### 3 EXPERIMENTAL FRAMEWORK

Like (Heitz and Koller, 2008), we aim to classify pixels in an image to generate a segmentation map for it. Every pixel must be in one of the three classes of the problem:

- Vegetation: Grass, trees, swamped plants, etc
- Water bodies: Rivers, lakes, lagoons, swamps, waterfalls, margins, etc
- Other: Everything that does not belong to the aforementioned classes, potentially something human-made or a strange natural object

The image database comes from the GEOMA project (INPE, 2013) dataset. This dataset is composed of aerial images of the Amazonian rain forest region, taken with VGA cameras on board of manned airplanes during a number of flights between Amazon's major urban centres. The complete database has approximately 40,000 images (13 GB), from which we used a single flight outcome, comprising of 3,031 images (1.02 GB). All images use 24-bit colour space with dimensions of 640 per 480 pixels

(0.3 MegaPixels). This dataset is publicly available at <https://github.com/luizcavalcanti/geoma-database>.

Gaussian blur is used on all images prior to any feature extraction. This was found useful because rain forest vegetation is very irregular and causes edge detector filters to produce incorrect border information (i.e. too many borders). Using a Gaussian Blur filter before edge detection aims to reduce the level of noise in the image (Deng and Cahill, 1993) and smooth borders, which improves the result of edge-detection algorithms (Shapiro and Stockman, 2001). In this work a Gaussian filter with a  $5 \times 5$  convolution matrix and  $\sigma = 2$  were used.

A comparison is shown in Figure 5, where the results of a Sobel operator are presented for the same original image, with (Figure 5(c)) and without (Figure 5(b)). In this example a reduction in improper border detection can be seen.

From the whole dataset, 17 images were chosen as good representations of the overall terrain found in the remaining images. Some patches on those images were labeled by a specialist as one of the three classes of the problem and every pixel on those patches had their features extracted to compose the training dataset. A total of 4,000 pixels were used as training data, distributed among the classes as shown in table 1.

Class	Samples
Vegetation	1,200
Water body	1,600
Other	1,200
Total	4,000

Table 1: Training dataset of pixels and its distribution among the classes

For every pixel, which represents a sample in this work, we obtained a vector of features. Those features represent relevant information on colour, brightness, borders, neighbourhood and texture of the pixel or its region.

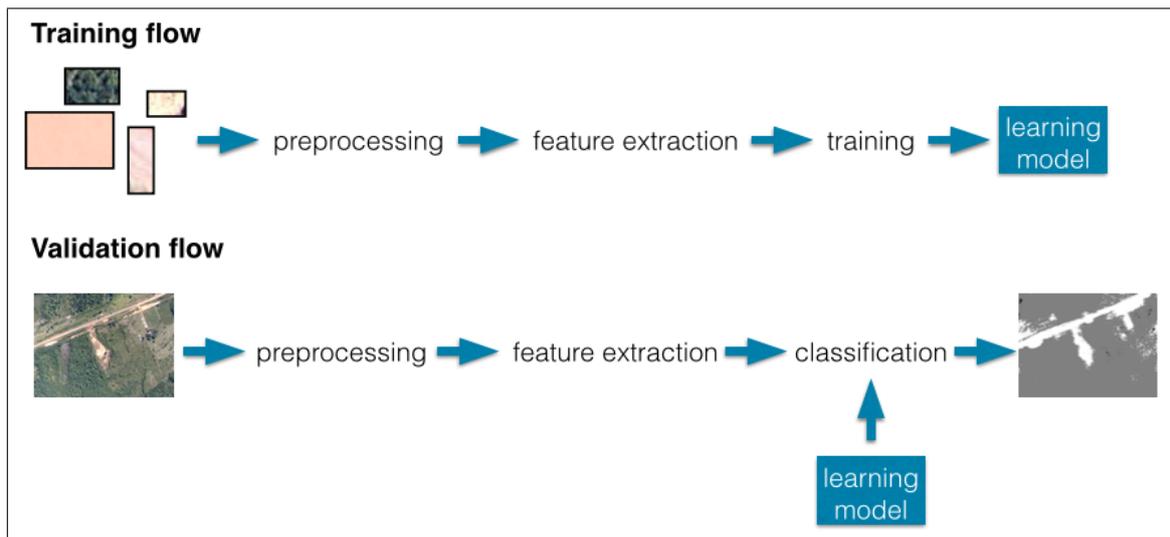


Figure 4: Overview of our approach. The training flow generates the learning models and the validation flow generates the segmentation maps for validation images

To obtain colour and brightness information we convert the RGB pixel into a HSB representation and its channels (Hue, Saturation and Brightness) compose the feature vector. For edge detection and border information, we extracted the difference of Gaussians and Sobel operator information on the neighbourhood (Nixon and Aguado, 2008).

Hessian features are also extracted to provide information on morphology and texture (Mikolajczyk and Schmid, 2002). This includes a scalar representation of the Hessian matrix, trace, determinant, first and second eigenvalues, orientation, gamma-normalized square eigenvalue difference and square of gamma-normalized eigenvalue difference.

Provided that those are relatively complex features, their representation in unidimensional variables for every pixel and its neighbours makes the feature vector grow quickly. In this work, using the aforementioned features, the features vector contains 81 discrete numeric features. This is done for the labeled pixels involved in the training phase but also for every non-labeled pixels involved in the validation/test phase, to compose the actual training and validation datasets respectively.

This data was used for training all the algorithms in this work, with the exact same attributes. A learning model was created for every algorithm using the Waikato Environment for Knowledge Analysis tool (Hall et al., 2009), also known as WEKA. The algorithms in this work are all supervised. Decision trees, KNN, Naive Bayes and Random forests were used.

For each learning algorithm, the parametrisation used in this work was decided during the training

phase, using cross-validation in the training dataset as the accuracy performance criteria. For KNN, a range of  $k = [1, 7]$  was tested and the best results were with  $k = 5$  and Euclidean distance, so that's the one used throughout the experiment. In Random forests, the maximum number of trees was 200 and the number of features per tree was 40.

From the dataset, 500 images were chosen as validation dataset. Each of their pixels had features extracted and were classified by the learning model generated on the training phase. Those same classified pixels were used to render a segmentation map for every image (Figure 6).

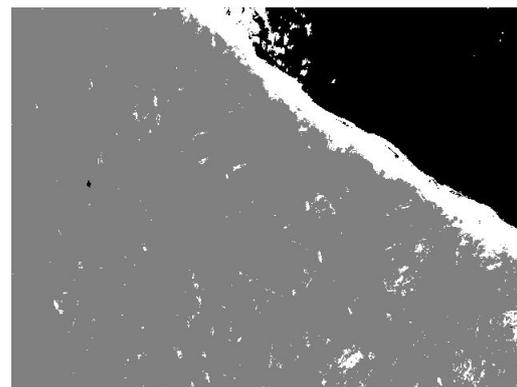
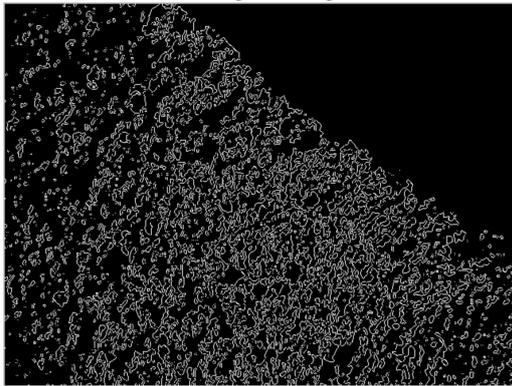


Figure 6: Example of segmentation map generated by pixel classification

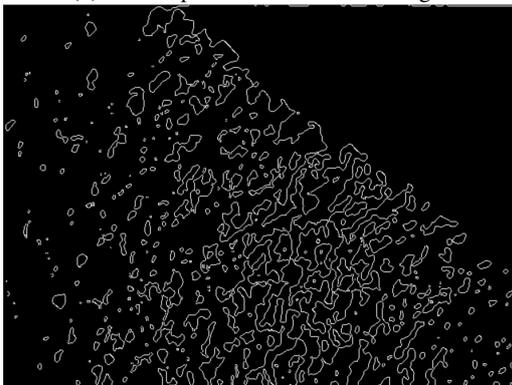
Evaluating semantic segmentation is often a purely visual and subjective process, hence a specialist was responsible for analysing each one of the seg-



(a) Original image



(b) Sobel operator on unfiltered image



(c) Sobel operator on blurred image

Figure 5: A comparison between using a unfiltered and a gaussian-blurred one in edge detection (Sobel operator)

mented images for all techniques applied. For every resulting segmentation map, the specialist was responsible to classify the whole image as an "acceptable" or "unacceptable" semantic segmentation. The percentage of "acceptable" segmentation composes the accuracy score of the methods in this work, and comparisons are made by this criteria (Section 4).

The specialist was instructed that "acceptable" segmentations are comprised by well-defined and continuous segmented regions compatible with the expected semantic segmentation of the original image. Some noise, or pixels with wrong classification, is acceptable as long as it don't form another continuous region or an expressive area inside the correct region. Were deemed "non-acceptable" the segmentation maps that violate any of those criteria.

The overview of our approach can be seen in Figure 4.

## 4 RESULTS

All algorithms were trained with the same training dataset and tested against the same validation dataset.

After a specialist evaluated every output for all algorithms, we came with an accuracy percentage, represented by the segmentations assigned by the specialist as "acceptable".

The figure 7 shows two examples of the experiment's output: a segmentation map, for every learning algorithm used on both images. Black areas represent "water bodies" class, grey areas represent "vegetation" class, white areas represent "other". During the experiment, Random forest algorithm (Figure 7(f)) had the most problems with "water bodies" false-positives, but had a superior semantic segmentation in images without water elements. Decision tree (Figures 7(c) and 7(g)) did not cause many false-positive problems, but usually overflowed classes' edges in more complex images, sometimes ruining the semantic segmentation. KNN (Figures 7(d) and 7(h)) and Naive Bayes attained a good overall segmentation but had a lot of classification noise in regions that should be contiguous.

The accuracy and average execution time (per sample) for all tested algorithms are presented in table 2. The feature extraction processing time for a whole image of the dataset (307,200 pixels) took longer than expected: around 7023 milliseconds in a 2.4 GHz Intel Core i5 processor. This feature extraction time is not considered in the algorithms average execution time. Figures 8 and 9 show graphical comparisons among methods' error rate (linear scale) and average execution time (logarithmic scale) respectively.

All algorithms, scripts and datasets used in this work are publicly available under *GNU General Public License v2 (GPLv2)* at <http://github.com/luizcavalcanti/ForestClassifier> and it's free to reproduce, modify and distribute accordingly.

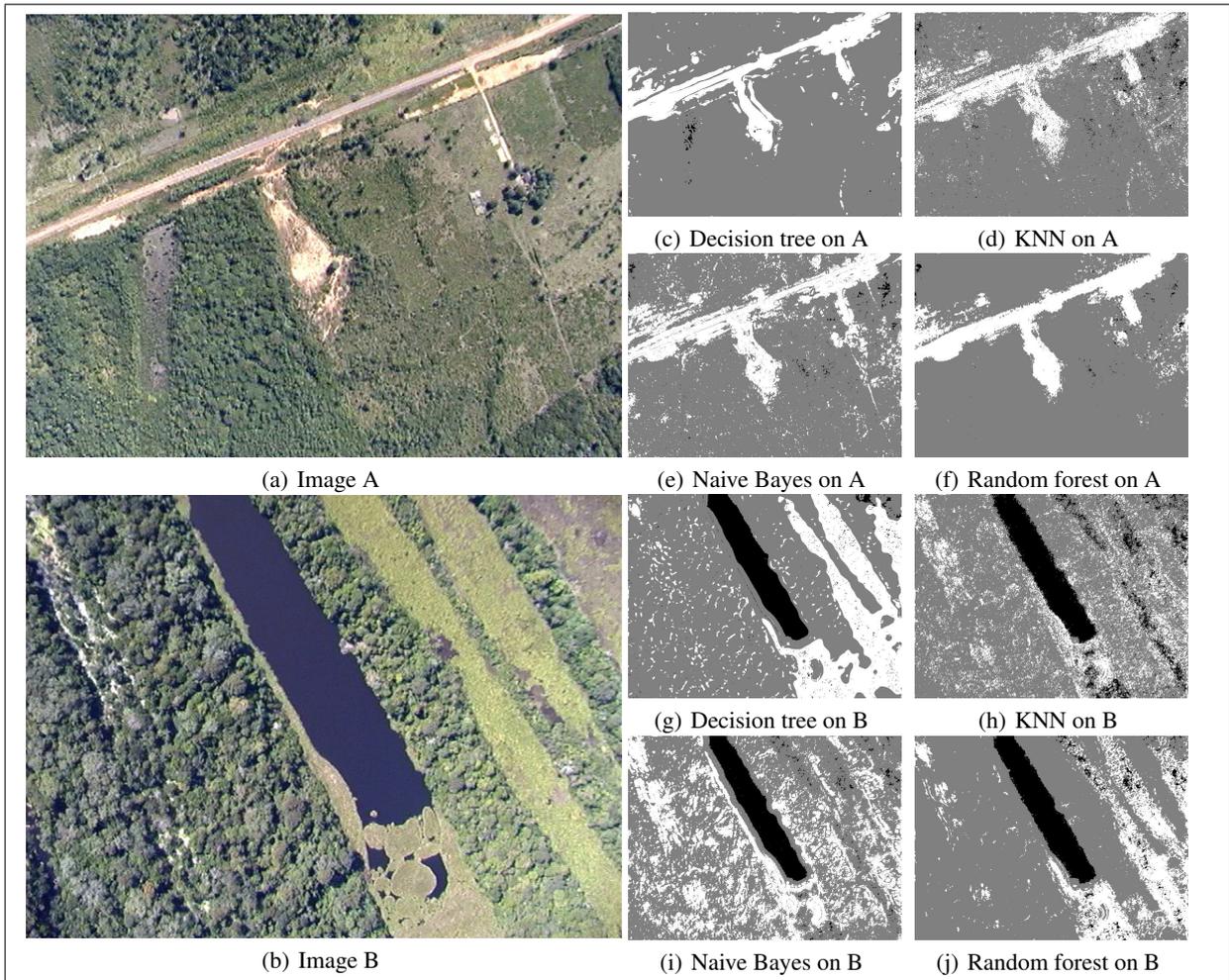


Figure 7: Visual comparison on segmentation results

Algorithm	Accuracy	Execution time
Decision tree	82.2%	136 ms
KNN	92.6%	420,098 ms
Naive Bayes	92.8%	320 ms
Random forest	96%	3,542 ms

Table 2: Accuracy and average execution time for all algorithms tested

## 5 CONCLUSION AND FUTURE WORKS

With over 90% of accuracy, some of the tested machine learning algorithms are good approaches to classification of aerial forest images with low resolution. Naive Bayes and KNN algorithms had a fairly good performance, but are still suffering from noise in its learning models. KNN also has a computational performance problem that needs further investigation,

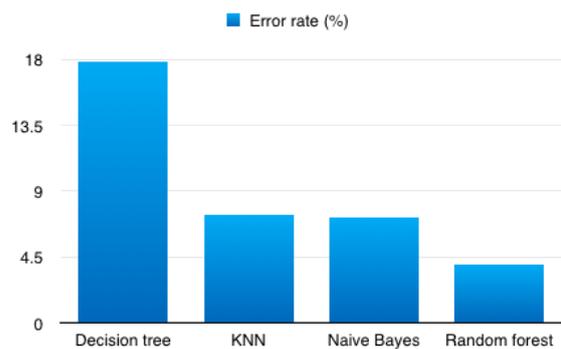


Figure 8: Comparison on error rate (linear scale)

since it takes 3 orders of magnitude more time than Naive Bayes to classify samples. This could be related to the number of training samples that every new sample needs to be compared to, but also to the di-

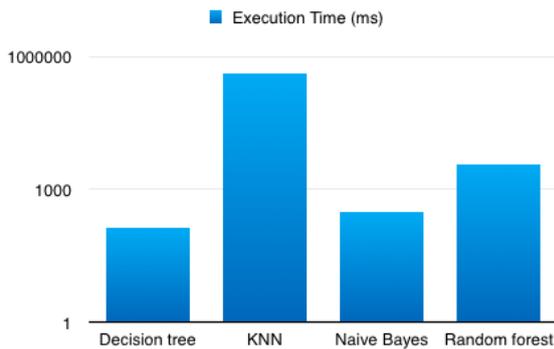


Figure 9: Comparison on execution time in milliseconds (logarithmic scale)

dimensionality of the feature vector. A technique to reduce the number of comparisons should be tested.

Random forest clearly performs best in this dataset, but we must find a way to address the false-positives on water bodies, which is visibly larger than in other methods. This is definitely a theme for future works, as the overall performance of Random forest was promising.

Future works should also include an improvement on samples representation, making the feature extraction faster and reduce the feature vector dimensionality. Algorithms like KNN and SVM should benefit from those improvements, specially the later, making possible to reduce the support vector complexity and make model generalisation easier. Another area of possible improvement is in the image preprocessing, prior to the feature extraction, specially reducing noise. It is also important to test the same dataset with unsupervised machine learning techniques and compare to those in this work.

In general, the results were satisfactory in providing good directives on how to implement a efficient and robust segmentation tool for the rainforest operation scenarios. Amazon forest has been suffering from years with systematic degradation and this work is a small part of an effort to provide information and supporting actions to mitigate the deforestation activities in the region.

## 6 Acknowledgements

This work was partially sponsored by *Fundacao de Amparo a Pesquisa do Estado do Amazonas (FAPEAM)* under the ARTES project (No. 114/2014). The Nokia Institute of Technology (INDT) also funded partially this work.

## REFERENCES

- Bosch, A., Zisserman, A., and Munoz, X. (2007). Image classification using random forests and ferns.
- Deng, G. and Cahill, L. (1993). An adaptive gaussian filter for noise reduction and edge detection. In *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record.*, pages 1615–1619 vol.3.
- Dubuisson-Jolly, M.-P. and Gupta, A. (2000). Color and texture fusion: application to aerial image segmentation and gis updating. *Image and Vision Computing*, 18(10):823 – 832.
- Freixenet, J., Munoz, X., Raba, D., Marti, J., and Cufi, X. (2002). Yet another survey on image segmentation: Region and boundary information integration. In Heyden, A., Sparr, G., Nielsen, M., and Johansen, P., editors, *Computer Vision ECCV 2002*, volume 2352 of *Lecture Notes in Computer Science*, pages 408–422. Springer Berlin Heidelberg.
- Ghiasi, M. and Amirfattahi, R. (2013). Fast semantic segmentation of aerial images based on color and texture. In *Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on*, pages 324–327.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.
- Heitz, G. and Koller, D. (2008). Learning spatial context: Using stuff to find things. In *Computer Vision—ECCV 2008*, pages 30–43. Springer.
- INPE (2013). Projeto GEOMA <http://www.geoma.lncc.br/>.
- Li, G. and Wan, Y. (2010). Improved watershed segmentation with optimal scale based on ordered dither halftone and mutual information. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, volume 9, pages 296–300.
- Mikolajczyk, K. and Schmid, C. (2002). An affine invariant interest point detector. In Heyden, A., Sparr, G., Nielsen, M., and Johansen, P., editors, *Computer Vision ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 128–142. Springer Berlin Heidelberg.
- Nixon, M. and Aguado, A. S. (2008). *Feature Extraction & Image Processing, Second Edition*. Academic Press, 2nd edition.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on*, 9(1):62–66.
- Pal, N. R. and Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277 – 1294.
- Shapiro, L. G. and Stockman, G. (2001). *Computer vision: Theory and applications*.
- T., A., J., M., and M., H. C. . P. (2009). Rotation invariant image description with local binary pattern histogram fourier features. In *In: Image Analysis, SCIA 2009 Proceedings, Lecture Notes in Computer Science 5575, 61-70*. ISBN 978-3-642-02229-6.

# APÊNDICE B – Instruções da ferramenta de segmentação manual

## Objectives

Segment aerial images on the Amazon rainforest according to terrain coverage/type

## What are segments?

Segments are closed regions on images that separate some elements in the image from the rest. Take figure 1, for example:



Fig. 1 - An image where parts of it were segmented

This figure was divided in several segments. One could say that the person or algorithm segmented this image to separate different kinds of vegetation, other would argue that it was divided simply by dominant color. The point is that there is no correct or proper segmentation unless you know what are you trying to separate or analyse in the image.

Segmentations can also vary by granularity. Let's observe figure 2:

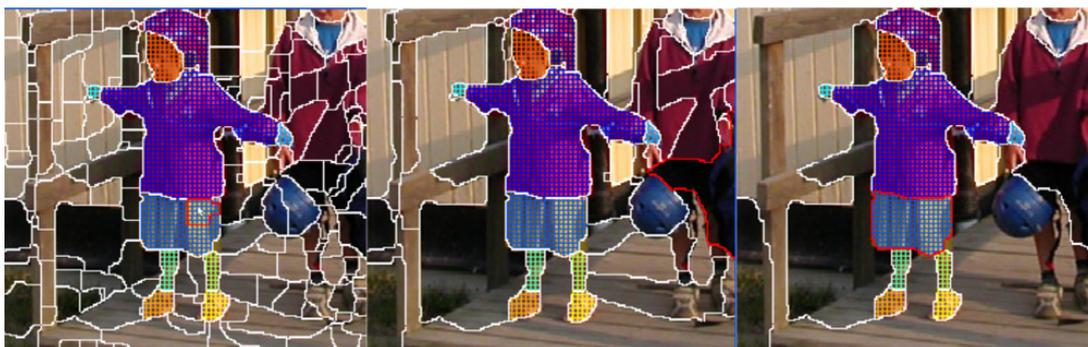


Fig. 2 - The same image segmented with different granularities

All those images segment the same image with seemingly the same valid criteria, but with different granularity. This is also perfectly normal in human or manual segmentation, since we tend to describe or view objects with different levels of detail.

## What do you need to do?

Your mission is to manually segment the given images as accurately as possible, according to your own judgment.

The criteria here is terrain coverage. So, we would like to separate different vegetations, geological formations and human-made objects. Let's have an example with figure 3:



Fig. 3 - An original image and a plausible segmentation for it, according to this experiment's criteria

The pink/magenta line on the rightmost picture separates different types of terrain coverage and objects. An overly obsessed person would maybe segment every tree in the forest patch on that imagem, but that's not necessary.

If we look closer on the lower right corner, we can see several human-made structures. In this example, the person that segmented the image chose to put all buildings in a single

segment. Other person would choose to define a segment for every individual building. None are wrong, they just decide to segment the image on a different granularity.

If an image is homogeneous, or in other words, if it has a single type of terrain/coverage, no segmentation is needed. An example of our dataset of that case is show in figure 4:



Fig. 4 - An homogeneous image. Makes little sense to segment it at all

To help you segment the images in the experiment, a fairly simple tool is available. The following section describe how it works.

## The segmentation tool

As figure 5 shows, the segmentation tool interface is divided in two parts. The toolbar on the top, with all the commands necessary for the segmentation of the image, and the canvas on the bottom, where the user can actually segment the image.

When a new image is given, you must chose which drawing tool you will use. For segments that crosses the image from one edge to the other or subdivides previous segments, the "free brush" is the ideal tool. It allows you to draw freely on the canvas, defining the segment border as you go, describing an open-ended path (figure 6).

The "closed region" tool, as the name implies, forces the drawn segment to be a closed on. It's ideal for internal segments of the image, those segments that don't collide with previous ones and don't touch the borders of the image (figure 7 shows an example).

For both drawing tools, you have two ways of interaction. You can draw continuously, as you would do in any computer-aided drawing tool, or you can just click on key

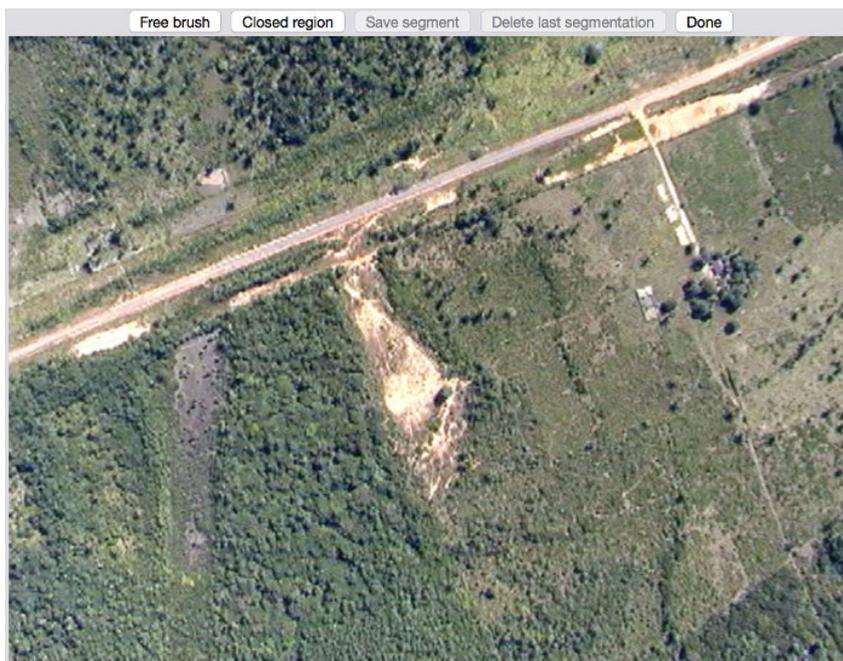


Fig. 5 - Segmentation tool interface

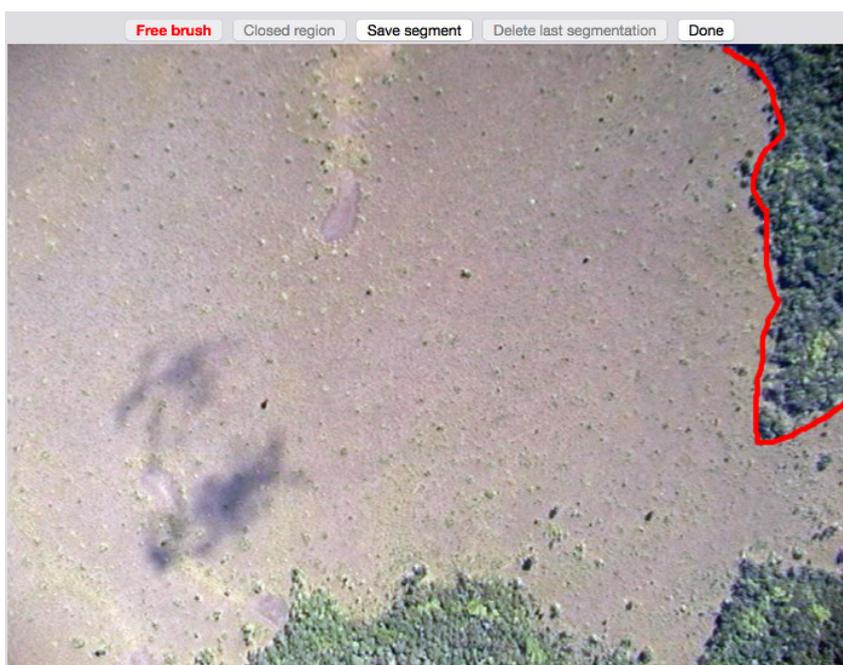


Fig. 6 - Using the "free brush" tool to draw an open-ended path

point to create a path. Experiment on those two ways until you feel comfortable.

The previous strokes on the current region you are drawing can be undone by selecting the "undo" option.

After you create a segment, the next step is to select the "save segmentation" option. This registers the current segment and changes its colour. If there is more regions on the current image to segment, you can select a new segment type (closed region or free brush)

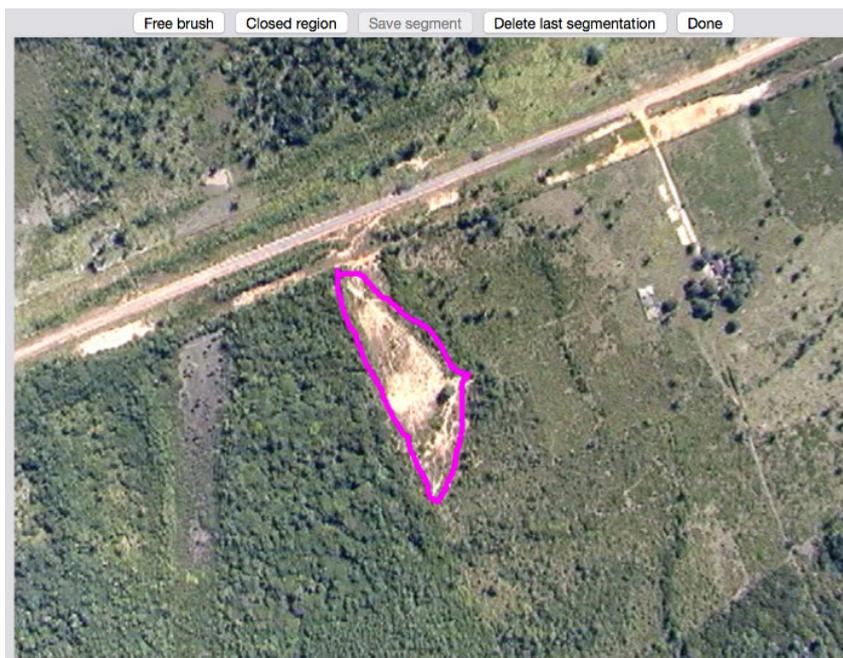


Fig. 7 - Using the "closed region" tool to determine a internal segment on the image

and start drawing it right away.

If you feel that a previous segment was not adequate, you can click the "delete last segmentation" option. As the name implies, it will remove the last defined segment on the image. You can continue to use this option until there is no more defined segments on the current image.

Once you feel that the image is properly segmented, you should select the "done" option. A brief confirmation box will appear. If you are sure that the work in this image is done, you can confirm it. In this moment, all data regarding your segmentation will be sent to the experiment server and a new image will be shown.

We understand that manually segmenting images can be a tiresome process. If you feel tired, please stop. It's better to return when you feel refreshed and willing to help. You just need to use the same user identifier on the login page to pick the segmentation experiment where you left it.