



**UNIVERSIDADE FEDERAL DO AMAZONAS  
INSTITUTO DE COMPUTAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**



**MICHEL MARIALVA YVANO**

**DETECÇÃO DE EMBARCAÇÕES POR IMAGENS NOS RIOS DA  
AMAZÔNIA**

Manaus – AM  
2016

**MICHEL MARIALVA YVANO**

**DETECCÃO DE EMBARCAÇÕES POR IMAGENS NOS RIOS DA  
AMAZÔNIA**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito para a obtenção do grau de Mestre em Informática.

**Orientadora:** PhD. Eulanda Miranda dos Santos

Manaus – AM  
2016

## Ficha Catalográfica

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

Y96d Yvano, Michel Marialva  
Detecção de embarcações por imagens nos rios da Amazônia /  
Michel Marialva Yvano. 2016  
74 f.: il. color; 31 cm.

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

1. Monitoramento fluvial. 2. Imagens. 3. Processamento digital de  
imagens. 4. Aprendizagem de máquina . I. Santos, Eulanda  
Miranda dos II. Universidade Federal do Amazonas III. Título



PODER EXECUTIVO  
MINISTÉRIO DA EDUCAÇÃO  
INSTITUTO DE COMPUTAÇÃO



UFAM

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

## FOLHA DE APROVAÇÃO

**"Detecção de Embarcações por Imagens nos Rios da Amazônia"**

**MICHEL MARIALVA YVANO**

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Profa. Eulanda Miranda dos Santos - PRESIDENTE

Prof. José Reginaldo Hughes Carvalho - MEMBRO INTERNO

Prof. José Luiz de Souza Pio - MEMBRO EXTERNO

Prof. José Pinheiro de Queiroz Neto - MEMBRO EXTERNO

Manaus, 29 de Julho de 2016

*Dedico este trabalho aos meus pais (Edaíze e Mário), minha esposa (Ananda Moura) e principalmente ao meu filho (Miguel Ryyu), que tiveram que abdicar da minha proximidade para que eu pudesse concluir com êxito esta etapa da formação acadêmica.*

## **AGRADECIMENTOS**

Meus agradecimentos são inúmeros, mas tentarei ser o mais breve possível.

Agradeço a Deus por sempre estar comigo fazendo com que eu conhecesse inúmeras pessoas boas e a possibilidade de conseguir um arsenal de conhecimento.

Agradeço à minha orientadora, professora Eulanda dos Santos, por toda paciência, compreensão e conhecimento a mim dedicados desde o início do processo que foi na produção do pré-projeto para seleção. Agradeço também ao meu antigo orientador, professor José Pinheiro Neto, por dicas valorosas na primeira etapa do processo de pesquisa.

Agradeço aos meus familiares, principalmente aos meus pais, Edaíze e Mário, por me ajudarem mostrando que a educação é a base de tudo, por seus pensamentos positivos e por serem meus exemplos de dedicação e perseverança. Agradeço também à minha esposa, Ananda Moura, por ser nos momentos difíceis a “voz da razão”, além de me dedicar um grande apoio emocional.

Agradeço a todos os meus amigos feitos nesta jornada que me apoiaram, me fizeram dar risadas e principalmente me ajudaram a trilhar este caminho da melhor forma possível, dentre eles destaco: Thais Oliveira, Bernardo Gatto, Caio Gregoratto, Ádria Menezes, Rayol Neto, Carlos Ramos, Pablo Elleres, Adriana Rodrigues, Daily Daleno, Luiz Cavalcante e Anderson Gadelha. Além do meu amigo de longa data Saulo Ricardo pelo apoio na minha estadia em Manaus. E todos os demais que não conseguirei citá-los.

E por fim agradeço ao Instituto Federal do Pará que me financiou durante esta jornada. E a todos os meus colegas do colegiado de Informática do Campus Itaituba, estendendo o agradecimento ao Diretor Geral, Diretor de Ensino, Diretor Administrativo e Coordenação de Pesquisa.

*“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.*  
*(Marthin Luther King)*

## RESUMO

O monitoramento fluvial é um trabalho intensivo e necessário a fim de evitar possíveis ameaças como pirataria e agressão ambiental. Normalmente, essa tarefa é realizada manualmente por um operador humano que analisa monitores de vídeo por longos períodos de tempo, fator que torna a tarefa longa, maçante e sujeita a erros. Na Amazônia, o monitoramento fluvial é ainda mais desafiador devido à grande extensão territorial da região e ao fato desta possuir a maior bacia hidrográfica do planeta. Dentre as possibilidades de tratar este problema existe a utilização de sensoriamento remoto, em geral utilizando imagens aéreas obtidas por satélites ou veículos aéreos, tripulados ou não. Diante desse contexto, esta dissertação propõe um método computacional capaz de detectar a movimentação de embarcações em um ambiente típico dos rios da Amazônia, pois os métodos existentes na literatura não se adequam ao cenário amazônico devido a inúmeros fatores, como por exemplo, a diversidade de composição de fundo da região. Além disso, métodos propostos na literatura foram desenvolvidos para ambientes com mar aberto, onde normalmente há apenas embarcações e água em cena. Neste trabalho, são usadas técnicas de processamento digital de imagens para a detecção do objeto em movimento e técnicas de aprendizagem de máquina para determinar a probabilidade do objeto observado ser um barco ou não. Essa dupla detecção reduz os falsos alertas emitidos pelo método proposto. Os experimentos realizados mostram que o método obteve 79% de acurácia e 91% de precisão, considerando todas as imagens, e precisão de 71% em cenas com embarcações, mostrando-se eficiente ao ser comparado com outras estratégias.

**Palavras-chave:** monitoramento fluvial; imagens; processamento digital de imagens; aprendizagem de máquina



## ABSTRACT

Fluvial monitoring is an intensive and necessary task used as a support to avoid possible threats such as piracy and environmental aggression. Usually, this task is performed manually by a human operator, who analyzes video monitors for long duration, which leads the task to be long, tedious and prone to errors. In the Amazon rainforest, fluvial monitoring is even more challenging due to the large territory of the region and due to the fact that the largest river basin in the world is in this region. Remote sensing, usually based on aerial images obtained by satellite or unmanned/manned aerial vehicles is one possibility to deal with this problem. In this context, we propose in this work a computational method focused on detecting the movement of vessels in a typical scene of the Amazon rivers. This new method is necessary since current solutions are not suited to deal with the Amazon scenario due to several reasons, for instance, the diversity of the background of the Amazon region. In addition, several methods available in the literature were designed in the context of open sea scenarios, where there are usually only vessels and water in the scenes. In this work, digital image processing techniques are employed to detect moving objects in the scene, while machine learning algorithms are used to indicate the probability about the class of the observed object, i.e. whether or not the detected object is a boat. This double detection process is intended to reduce false alerts provided by the proposed method. The experiments conducted in this work show that the proposed method achieved 79% of accuracy and 91% precision rate when all images are considered, as well as a 71% precision rate when only scenes with vessels are analyzed. These results indicate the effectiveness of the proposed method when compared to other strategies.

**Keywords:** fluvial monitoring; images; digital image processing; machine learning

## LISTA DE FIGURAS

Figura 1 – Imagem obtida por meio de teste de uma câmara instalada.....	14
Figura 2 – Formação de uma imagem digital e sua representação numérica... 17	17
Figura 3 – Figuras que representam a formação de uma imagem colorida .....	18
Figura 4 – Quadros (frames) de uma cena de vídeo .....	18
Figura 5 – Passos Fundamentais em Processamento de Imagem.....	19
Figura 6 – Principais Componentes de um Sistema de Segurança.....	21
Figura 7 – Três regiões em uma imagem binária.....	24
Figura 8 – Elementos estruturantes.....	25
Figura 9 – Processo de Erosão .....	25
Figura 10 - Processo de Dilatação.....	26
Figura 11 – Processo de Abertura e Fechamento .....	27
Figura 12 - Processo de Preenchimento .....	28
Figura 13 - Fases do Reconhecimento de Padrões Clássico .....	31
Figura 14 - Arquitetura proposta .....	49
Figura 15 – Arquitetura detalhada do método de detecção proposto .....	50
Figura 16 - Etapa de detecção inicial .....	51
Figura 17 - Etapa de detecção Final (Confirmação) .....	53
Figura 18 – Identificação do corpo da embarcação .....	56
Figura 19 – Exemplos de marcações da embarcação nos quadros .....	63
Figura 20 - Marcação do <i>Ground Truth (GT)</i> .....	65

## LISTA DE TABELAS

Tabela 1 – Síntese de trabalhos e métodos em cada trabalho.....	45
Tabela 2 – Mediana das marcações dos vídeos .....	58
Tabela 3 - Quantidade de imagens total.....	59
Tabela 4 - Resultado dos algoritmos de classificação.....	61
Tabela 5 - Resultados obtidos na primeira série de experimentos .....	62
Tabela 6 - Resultados obtidos na segunda série de experimentos.....	65

## LISTA DE SIGLAS

VC	–	Visão Computacional
RGB	–	<i>Red, Green e Blue</i>
HSI	–	<i>Hue, Saturation e Intensity</i>
HSV	–	<i>Hue, Saturation e Value</i>
FP	–	Falso Positivo
FN	–	Falso Negativo
SFDA	–	<i>Sequence Frame Detetion Accuracy</i>
HOG	–	<i>Histogram of Oriented Gradients</i>
KLT	–	Kanade – Lucas - Tomasi
2D	–	Duas Dimensões
3D	–	Três Dimensões
SVM	–	<i>Support Vector Machine</i>
PCA	–	<i>Principal Component Analysis</i>
ROC	–	<i>Receiver Operating Characteristic</i>
VP	–	Verdadeiro Positivo
VN	–	Verdadeiro Negativo
PR	–	Precisão
RC	–	Revocação ou Sensibilidade
EP	–	Especificidade
AC	–	Acurácia
GT	–	<i>Ground Truth</i>
KNN	–	<i>k-Nearest Neighbors</i>

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	12
1.1. CONTEXTUALIZAÇÃO E MOTIVAÇÃO .....	12
1.2. OBJETIVOS .....	15
<b>1.2.1. Geral</b> .....	15
<b>1.2.2. Específicos</b> .....	15
1.3. ESTRUTURA DO DOCUMENTO .....	15
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	17
2.1. IMAGEM E VÍDEO .....	17
2.2. SISTEMA DE SEGURANÇA POR VÍDEOS .....	20
2.3. PROCESSAMENTO DE BAIXO NÍVEL.....	21
<b>2.3.1. Transformação em tons de cinza</b> .....	22
<b>2.3.2. Realce de Imagens</b> .....	22
2.4. PROCESSAMENTO DE NÍVEL INTERMEDIÁRIO .....	22
<b>2.4.1. Limiarização</b> .....	23
<b>2.4.2. Componentes Conectados</b> .....	23
<b>2.4.3. Processamento Morfológico de imagens</b> .....	24
2.4.3.1. Elemento Estruturante .....	24
2.4.3.2. Erosão .....	25
2.4.3.3. Dilatação .....	26
2.4.3.4. Abertura e Fechamento.....	26
2.4.3.5. Preenchimento de buracos .....	27
2.4.3.6. Fecho convexo.....	29
<b>2.4.4. Técnicas de segmentação/deteção de objetos em movimento</b> .....	29
2.4.4.1. Subtração de Fundo.....	29
2.4.4.2. Fluxo Ótico.....	29
2.4.4.3. Diferença Temporal.....	30
2.5. PROCESSAMENTO DE ALTO NÍVEL .....	30
<b>2.5.1. Reconhecimento de Padrões</b> .....	30
<b>2.5.1.1. Extração de Características</b> .....	31
<b>2.5.1.2. Classificação</b> .....	33
2.6. MÉTRICAS DE AVALIAÇÃO.....	34
2.7. DETECÇÃO DE EMBARCAÇÕES.....	36
<b>3. TRABALHOS RELACIONADOS</b> .....	38
3.1. ANÁLISE DE CONTRIBUIÇÃO DOS TRABALHOS RELACIONADOS.....	44
<b>4. MÉTODO PROPOSTO</b> .....	48
4.1. ARQUITETURA GERAL DO MÉTODO.....	48
4.2. DETECÇÃO INICIAL.....	50
4.3. DETECÇÃO FINAL .....	52
<b>5. EXPERIMENTOS E RESULTADOS</b> .....	55
5.1. PROTOCOLO EXPERIMENTAL .....	55
5.2. BASE DE DADOS E DELIMITAÇÃO DOS EXPERIMENTOS.....	57
5.3. RESULTADOS DOS EXPERIMENTOS .....	59
5.4. DISCUSSÃO .....	66
<b>6. CONSIDERAÇÕES FINAIS</b> .....	67
6.1. CONSIDERAÇÕES.....	67
6.2. LIMITAÇÕES DO MÉTODO.....	69
6.3. CONTRIBUIÇÕES.....	69
6.4. TRABALHOS FUTUROS .....	69
<b>REFERÊNCIAS</b> .....	71

## 1. INTRODUÇÃO

O desenvolvimento de ferramentas computacionais capazes de auxiliar no setor de segurança, proteção e prevenção ambiental é o foco de linhas de diversas pesquisas realizadas atualmente. O uso dos recursos computacionais para fins preventivos virou rotina em inúmeras atividades, tornando-se uma área em amplo desenvolvimento e que gera muitos benefícios para o ser humano. Na área ambiental, é comum desenvolver ferramentas de sensoriamento para auxiliar na proteção de áreas de reserva e afins. Os estudos para esses fins são variados e englobam diversas áreas da computação aplicadas em pesquisas para análise e reconhecimento de inúmeras ameaças que podem existir, tais como: desmatamento em florestas, monitoramento de áreas críticas de queimadas e até mesmo reconhecimento da fauna ou flora (SOUSA, 2010).

As técnicas de visão computacional (VC) são utilizadas em grande escala em trabalhos relacionados com a segurança ambiental. Com a crescente utilização de câmeras para aplicações de vigilância fora do domínio da segurança pública tradicional, como no controle de ambientes litorâneos (CULLEN, 2012), por exemplo, os sistemas de VC se multiplicam em atividades como aquisição, realce, compressão, armazenamento, análise, reconhecimento e visualização de imagens (CHELLAPPA, 2010).

Além das técnicas de VC, outra área da computação em amplo crescimento é a aprendizagem de máquina, especialmente por meio de algoritmos de classificação usados para o reconhecimento de padrões existentes em dados adquiridos (KOTSIANTIS, 2007).

Essas duas áreas são investigadas neste trabalho, o qual propõe um método para detecção de embarcações a partir de vídeos em ambiente típico dos rios amazônicos para auxiliar tarefas de monitoramento ambiental.

### 1.1. CONTEXTUALIZAÇÃO E MOTIVAÇÃO

A região amazônica tem importância reconhecida não apenas no Brasil, mas também em diversos outros países que constantemente demonstram preocupação com o maior patrimônio natural da terra. Nesta região está localizada a maior floresta tropical do planeta, uma rede hidroviária notável e,

assim, uma reserva considerável de água doce. A região amazônica possui cerca de 72% da rede hidroviária e por volta de 26,7 mil Km navegáveis, composta pelas Bacias Amazônicas e do Tocantins (ANA, 2005).

O transporte fluvial é o mais importante meio de transporte utilizado na região amazônica, dada a grande rede de rios navegáveis e a precariedade dos demais meios, além do que este é o principal motivo para grande parte das cidades, vilas e lugarejos estarem situados próximos aos rios (BRASIL, 2006).

No Brasil, a Amazônia sempre foi símbolo de riqueza e soberania nacional, porém, devido às suas dimensões e dificuldades de acesso, essa região tem se tornado alvo fácil para ações de desmatamento, biopirataria e outras atividades ilegais (BRASIL, 2006). Portanto, a fiscalização de áreas fluviais na Amazônia é de grande importância, pois é por meio do uso de embarcações que ocorrem a maioria das ações de contrabando, tráfico, transporte ilegal de pessoas e madeiras nobres, entre outras questões ilegais.

Nesse contexto surgiu o projeto ARTES: Aeróstato Remoto de Telecomunicação e Sensoriamento, financiado pela FAPEAM (Fundação de Amparo à Pesquisa do Estado do Amazonas). Esse projeto prevê a instalação de balões dirigíveis ancorados, ou Aeróstatos, os quais serão utilizados como retransmissores para enlaces de comunicação, assim como ferramenta importante para o apoio nas pesquisas de preservação e monitoramento ambiental nas reservas de Mamirauá e Amanã, no Estado do Amazonas. As Reservas de Desenvolvimento Sustentável Mamirauá e Amanã são duas grandes unidades de conservação que formam 3.474.000 ha de área protegida e que fazem parte do Corredor Central da Amazônia.

Uma das tarefas de monitoramento previstas no projeto ARTES é a detecção de barcos a partir de vídeos capturados por câmeras instaladas na parte mais alta do balão. Assume-se que as câmeras terão suas lentes direcionadas para baixo, como exibido na Figura 1. Essa figura mostra uma imagem obtida a partir de uma câmera instalada em um aeróstato, o qual foi ancorado apenas para teste na sede do Instituto de Desenvolvimento Sustentável Mamirauá<sup>1</sup>, que é um dos órgãos envolvidos no projeto ARTES e que está localizado na cidade de Tefé, estado do Amazonas.

---

<sup>1</sup><http://www.mamiraua.org.br/pt-br>

O objetivo deste trabalho é contribuir com esse projeto por meio da detecção automática de embarcações a partir de vídeos, considerando o seguinte cenário: ambiente fluvial tipicamente amazônico e vídeos capturados por câmeras instaladas na parte mais alta do balão, cujas lentes serão direcionadas para baixo. A detecção automática, proposta do trabalho, diminuirá a tarefa de análise dos vídeos, realizada normalmente manualmente por um operador humano por longos períodos, o que produz uma longa jornada de trabalho maçante e suscetível a erros.

Figura 1 – Imagem obtida por meio de teste de uma câmera instalada em um aeróstato ancorado na sede do Instituto de Desenvolvimento Sustentável Mamirauá, Tefé – Amazonas.



Fonte: O Autor, 2016.

Entretanto, existem dificuldades relacionadas à detecção de embarcações a partir de vídeos/imagens nesse ambiente amazônico. Dentre as dificuldades existentes destaca-se a região que possui um fundo (*Background*) que é diversificado, há variedade de coloração dos rios, além da vegetação e do clima. Há também muita diferença entre a largura dos rios, fato que muitas vezes contribui com a dispersão do movimento da água. Por fim, existe a possibilidade de oclusão parcial da embarcação devido à vasta cobertura vegetal.

As dificuldades naturais do ambiente investigado tornam os métodos existentes na literatura não adequados ao cenário amazônico. Além disso, esses métodos foram desenvolvidos para ambientes com mar aberto, onde



normalmente há apenas embarcação e água em cena, características que simplifica a tarefa de detecção, mas não é comum em ambiente amazônico.

## 1.2. OBJETIVOS

Para melhorar a compreensão dos objetivos deste trabalho, os mesmos se apresentam divididos em geral e específicos.

### 1.2.1. Geral

Desenvolver e testar um método para detecção automática de embarcação por meio da identificação de movimento em imagens e classificação no contexto de rios amazônicos.

### 1.2.2. Específicos

- a) Combinar técnicas de processamento de imagens com aprendizagem de máquina para detecção das embarcações.
- b) Construir uma base de imagens obtidas em ambiente real do cenário amazônico, juntamente com as marcações reais das embarcações.
- c) Combinar características que favoreçam o processo de reconhecimento do padrão de embarcações.
- d) Remover fundo automaticamente a fim de diminuir os objetos a serem analisados para detectar a embarcação.
- e) Diminuir a quantidade de quadros finais a serem analisados pelo operador humano.
- f) Selecionar o classificador de melhor razão entre, precisão, acurácia e medida F.

## 1.3. ESTRUTURA DO DOCUMENTO

Este trabalho está organizado em cinco capítulos, incluindo este introdutório, onde são abordadas as principais etapas relacionadas ao objetivo da pesquisa realizada.

No Capítulo 2 serão apresentados os conceitos necessários para uma fundamentação teórica referente aos métodos de segmentação e detecção de embarcações.

O Capítulo 3 explanará alguns trabalhos relacionados que representam o estado da arte do tema proposto, delimitados ao período de 2008 a 2014.

O Capítulo 4 descreve o método proposto para a detecção de embarcações no ambiente amazônico, onde são definidas técnicas específicas, abordando suas relações para o resultado final da pesquisa.

No Capítulo 5 demonstrar-se-á os resultados dos experimentos realizados por meio de comparação do método proposto com um *baseline* para mostrar o desempenho do método proposto no ambiente do problema de pesquisa. Por fim, no capítulo 6 fecha com as considerações finais e as propostas de trabalhos futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

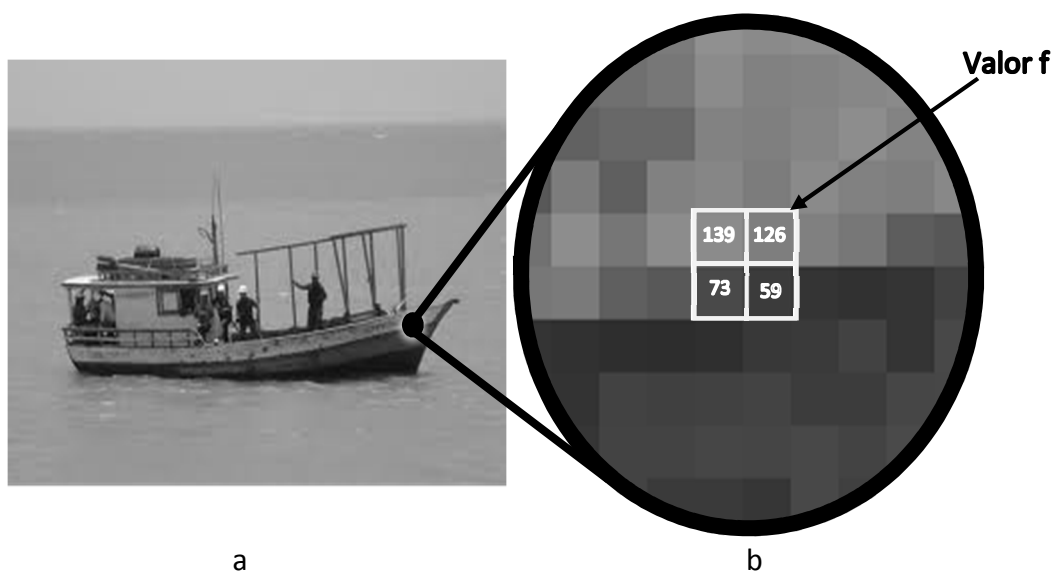
Este capítulo apresenta os principais conceitos necessários à compreensão do tema abordado nesta pesquisa. Serão apresentados conceitos referentes à detecção de objeto, detecção de embarcação e técnicas utilizadas para a detecção de embarcação em vídeos. Antes, porém, serão apresentados os conceitos de imagem e vídeo.

### 2.1. IMAGEM E VÍDEO

Os conceitos apresentados neste tópico são baseados principalmente em Gonzalez e Woods (2010).

Primeiramente, para definir-se vídeos, é importante a definição de imagem digital, a qual é uma função bidimensional  $f(x,y)$ , em que  $x$  e  $y$  representam coordenadas de um valor de  $f$  em um determinado espaço. Um ponto selecionado por estas coordenadas  $(x,y)$  é conhecido como pixel, enquanto o valor de  $f$  contido no pixel é denominado intensidade, conforme mostra na Figura 2, inspirada em Gonzalez e Woods (2010).

Figura 2 – Formação de uma imagem digital e sua representação numérica; (a) representa a imagem original; e (b) representa o *zoom in* de uma determinada região de (a) que mostra a intensidade de 4 pixels.

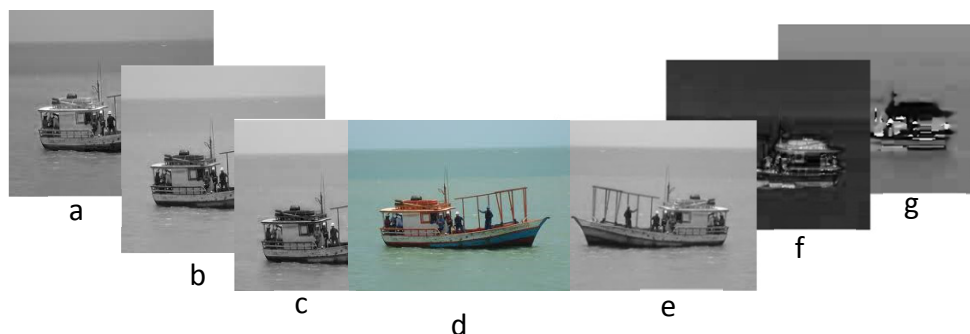


Fonte: Inspirada em Gonzalez e Woods, 2010.

Uma imagem digital colorida é a composição de uma matriz tridimensional  $M(x,y,z)$  onde cada dimensão  $z$ , pode representar um espectro de cor, que pode

ser Vermelho (*Red – R*), Verde (*Green – G*) e Azul (*Blue – B*) ou Ciano (*Cyan – C*), Magenta (*M*) e Amarelo (*Yellow – Y*) ou também a tripla de Matiz (*Hue – H*), Saturação (*Saturation – S*) e Intensidade/Valor/Luminosidade (*Intensity/Value/Luminosity – I/V/L*), como mostra a Figura 3, inspirada em Gonzalez e Woods (2010).

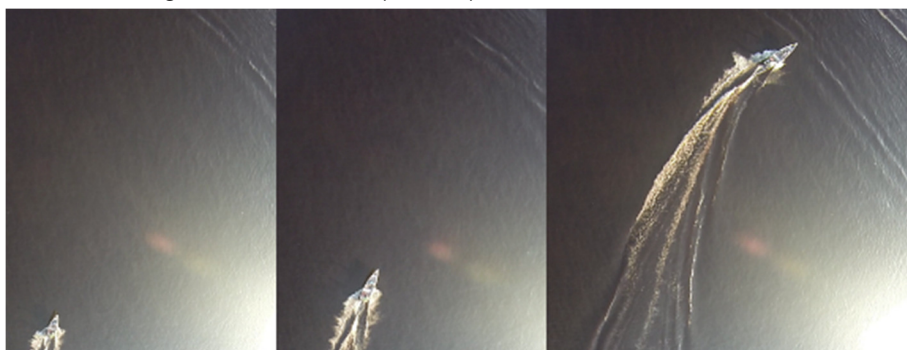
Figura 3 – Figuras que representam a formação de uma imagem colorida; (a-c), representam as camadas de cores de RGB, respectivamente; (d) é a imagem original; e (e-g) os valores de (V), saturação (S) e Matiz (H) do HSV, respectivamente.



Fonte: Inspirada em Gonzalez e Woods, 2010.

A partir da definição de imagem, vídeo pode ser entendido da seguinte maneira: um conjunto de imagens em movimento em que cada quadro (imagem) representa um determinado momento em um tempo  $t$ . Conforme pode ser visto na Figura 4, um vídeo forma um conjunto de dados, os quais podem ser analisados posteriormente em um formato dinâmico de visualização e possibilita a implantação de um sistema de segurança mais prático (MOREIRA, 2014).

Figura 4 – Quadros (frames) de uma cena de vídeo.

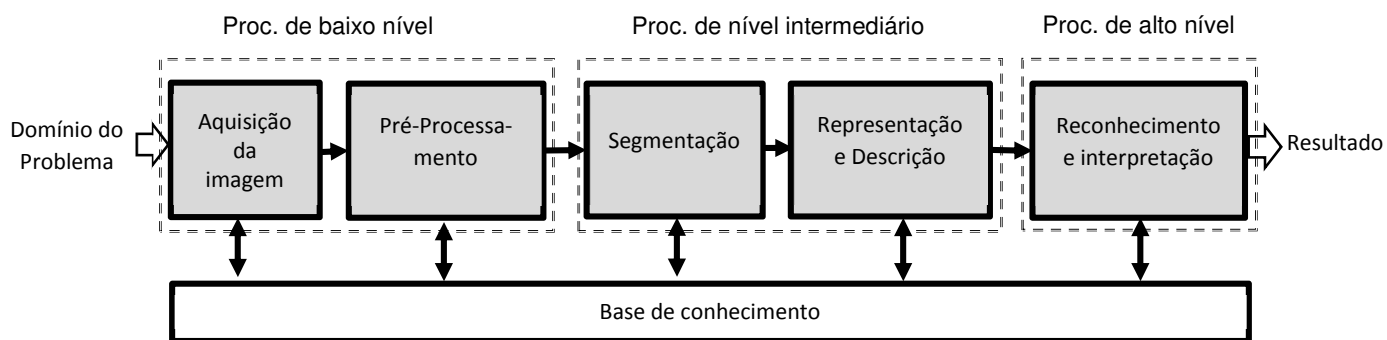


Fonte: O Autor, 2016.

As análises possíveis em uma imagem são divididas basicamente em três grupos: processamento em baixo nível, processamento em nível intermediário e

processamento de alto nível, como mostra a Figura 5, adaptada de Gonzalez e Woods (2010).

Figura 5 – Passos Fundamentais em Processamento de Imagem. A base de conhecimento é a informação do domínio do problema. O fluxo do processamento ocorre da esquerda para a direita a fim de melhorar a imagem, extrair atributos e interpretar os atributos para reconhecer o ambiente.



Fonte: Adaptado de Gonzalez e Woods, 2010.

O processamento em baixo nível se define pela melhoria das características de uma imagem por meio de pré-processamentos que envolvem operações de realce, filtragem e/ou restauração para melhoria da imagem original que, por consequência, trará resultados mais significativos nas demais etapas de processamento, assim entende-se que a entrada e a saída desta etapa são imagens.

O processamento de nível intermediário refere-se à etapa de extração de atributos da imagem por meio de identificação de formas significativas, subdividindo em partes de interesses, por exemplo a separação em plano de fundo (*background*) e primeiro plano (*foreground*).

Por fim, o processamento de alto nível analisa os atributos extraídos na etapa de nível intermediário para buscar informações associadas às funções cognitivas da visão humana, como por exemplo a identificação de um determinado objeto.

Neste trabalho, as etapas segmentação, representação e por fim reconhecimento são desenvolvidas no contexto da detecção de embarcações a

partir de vídeos para aplicações de segurança, assunto discutido na próxima seção.

## 2.2. SISTEMA DE SEGURANÇA POR VÍDEOS

Sistemas de vigilância em vídeo normalmente são utilizados para efetuar o monitoramento remoto de um determinado setor/ambiente, no intuito de facilitar a ação de vigilância tradicional feita pelo homem por meio de rondas periódicas nos locais salvaguardados. A vigilância baseada em vídeos é feita por meio da análise das sequências de quadros de imagens obtidas das câmeras instaladas nos locais estratégicos, em busca de falhas de segurança ou comportamentos inadequados que podem ocorrer em um determinado setor/ambiente. Esta forma de monitoramento baseado em vídeo, que é um ponto importante na linha de pesquisa em visão computacional, tem recebido muita atenção nas últimas décadas devido aos ambientes dinâmicos existentes nos vídeos e com vistas ao melhoramento da praticidade do monitoramento (HU et al., 2004; MAKANTASIS et al., 2013).

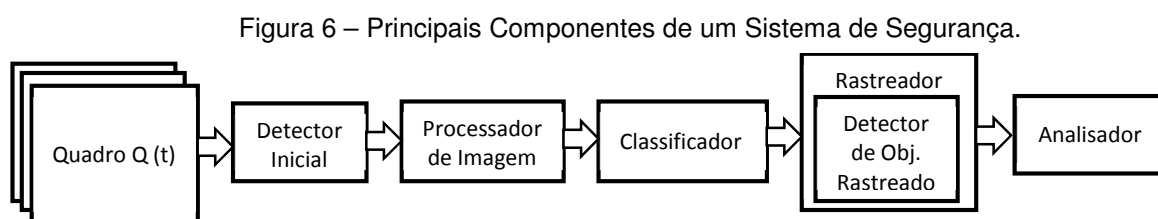
O monitoramento em vídeos se restringe basicamente à análise das sequências dos vídeos. Desta forma, Joshi e Thakore (2012) listam três maneiras de realizar esse monitoramento:

- **Monitoramento Manual:** são sistemas amplamente utilizados atualmente, porém, não muito eficazes devido às seguintes razões: a análise é executada por pessoas, e os vídeos comumente possuem horas ininterruptas de cenas repetidas de um determinado ambiente. Esses aspectos tornam esse tipo de monitoramento impraticável (CHEN; McGURR, 2014).
- **Monitoramento Semiautomático:** são sistemas que incluem alguns métodos de processamento de imagens digitais ou a utilização de sensores, porém, ainda assim necessitam de uma significativa intervenção humana. Sistemas que filmam apenas se houver a presença de algum movimento, porém, com análise dos vídeos feita por

um especialista humano (JOSHI; THAKORE, 2012), são exemplos desse tipo de monitoramento.

- **Monitoramento Automático:** são sistemas que executam todas as etapas necessárias de análise, sendo que apenas o resultado final é exibido. Isto é, o vídeo é fornecido como entrada para o sistema, o qual executa todo o processo de detecção, rastreamento e classificação de um determinado objeto/veículo/pessoa. Ao final do processo, o sistema, exibe um resultado indicando se a cena analisada é anômala ou não (CHEN; MCGURR, 2014).

Dentro de um sistema de segurança completo por vídeo, que utiliza monitoramento automático, pode-se observar cinco componentes principais que possuem tarefas que vão de baixo a alto nível, como mostra a Figura 6.



Fonte: Moreira, 2014.

Considerando o escopo do método proposto, assim como as Figuras 5 e 6, o foco deste trabalho é o processamento dos níveis intermediário e alto. Há inúmeras dificuldades na detecção de objetos em vídeos por diversos fatores: alto índice de ruído causado pela iluminação (baixa ou alta), movimentação da câmera, oclusão do objeto (parcial ou total) e diversos outros ruídos que podem afetar o bom desempenho da tarefa de detecção. Nas seções seguintes serão apresentados os passos referentes ao foco desta pesquisa.

### 2.3. PROCESSAMENTO DE BAIXO NÍVEL

Nesta etapa do processamento, definida normalmente como pré-processamento, busca-se a melhoria da imagem de entrada a fim de evidenciar suas características e facilitar o trabalho nas demais fases do processo. As

operações de filtragem, realce e aguçamento, para minimizar os ruídos na imagem de saída, podem ser parte desse processo (GONZALEZ e WOODS, 2010).

### 2.3.1. Transformação em tons de cinza.

Operação de quantização da imagem para 256 tons de cinza, realizada a fim de reduzir o espaço de cor da imagem (de  $2^{24}$  para  $2^8$ ), para que as operações com as imagens diminuam seu custo computacional. Dado que uma cor é composta por três canais de cores primárias R (*Red* - vermelho), G (*Green* - verde) e B (*Blue* - azul), a quantização de uma cor em um ponto para seu tom de cinza correspondente é realizada pela Equação 1 (GONZALEZ; WOODS, 2010).

$$G(x, y) = 0.114 R(x, y) + 0.587 G(x, y) + 0.299 B(x, y). \quad (1)$$

### 2.3.2. Realce de Imagens

O realce envolve procedimentos com a finalidade de melhorar a imagem para torná-la mais adequada para uma determinada aplicação. O processo de realce, classificado como pré-processamento, trabalha em baixo nível e tem como entrada uma imagem  $I$  e como saída uma outra imagem  $\hat{I}$ , com características melhoradas. A qualidade do realce destas características influencia diretamente à qualidade do processo seguinte de segmentação. Em resumo, se pode conceituar realce como a redistribuição das intensidades luminosas com o intuito de acentuar o contraste dos objetos de interesse (GONZALEZ e WOODS, 2010).

## 2.4. PROCESSAMENTO DE NÍVEL INTERMEDIÁRIO

A etapa de segmentação de uma imagem envolve a divisão da imagem a fim de detectar os seus elementos constituintes (regiões ou objetos). Na área de processamento digital de imagens, de forma geral, a segmentação automática de imagens de difícil separação é uma tarefa complexa. A precisão dessa etapa



define o sucesso ou o fracasso dos demais procedimentos de análise de uma imagem (GONZALEZ e WOODS, 2010).

A representação é uma tarefa que busca tornar os dados da imagem úteis para serem analisados. Dessa forma, o resultado da segmentação pode ser representado de duas formas: (1) por características externas (suas fronteiras); e (2) por suas características internas (os pixels que constituem uma região). Após a representação vem a etapa de descrição, a qual envolve a utilização dos dados das características para gerar propriedades específicas e discriminar a região (GONZALEZ e WOODS, 2010).

#### 2.4.1. Limiarização

É o processo de selecionar um limiar dentro do histograma de uma imagem em tons de cinza, a fim de realçar o primeiro plano. A saída deste processo é uma imagem binária, na qual os pixels de cor preta representam o fundo e o objeto é descrito por pixels de cor branca. Assim, considerando uma imagem  $f(x, y)$ , composta por objetos claros e um fundo escuro, o histograma de intensidades apresentará dois grupos dominantes. Desta forma, a maneira óbvia de separação entre os objetos e o fundo da imagem será dada por um Limiar  $T$  (GONZALEZ e WOODS, 2010). Onde ter-se-á uma imagem segmentada,  $g(x, y)$ , dada por:

$$g(x, y) = \begin{cases} 1 & \text{se } f(x, y) > T \\ 0 & \text{se } f(x, y) < T \end{cases} \quad (2)$$

#### 2.4.2. Componentes Conectados

Se existir um caminho entre dois pixels  $p1$  e  $p2$  de um determinado subconjunto de pixels  $S$ , afirma-se que eles são conectados. Assim, qualquer pixel  $p$  em  $S$ , com a formação de um conjunto de conexão com ele em  $S$ , é conceituado como componente conectado, ou conexo, de  $S$ . Define-se como conjunto conexo, o conjunto  $S$  de apenas um componente conexo. Se um subconjunto de pixel  $R$  de uma imagem for um conjunto conexo,  $R$  será uma Região da imagem. Considerando  $R1$  e  $R2$  duas regiões de uma imagem, afirma-se que elas podem ser adjacentes se sua união formar um conjunto conexo, caso

contrário, denominam-se de disjuntas. Se existir  $N$  regiões disjuntas, chama-se de  $R_u$  a união de todas as regiões  $N$  de uma imagem e seu complemento (conjunto de pontos fora de  $S$ ) de  $(R_u)^c$ , onde  $R_u$  será o primeiro plano (*foreground*) e  $(R_u)^c$  o fundo (*background*). A Figura 7 mostra os conceitos desta seção.

Figura 7 – Três regiões em uma imagem binária, os conjuntos de 1's representam as regiões da imagem (primeiro plano), e o conjunto de 0's representa o complemento das regiões (fundo).

0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	0	0	0	0	0
0	1	1	1	0	0	1	1	1	1
0	1	1	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	1	1
0	0	0	1	1	1	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	0	0	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Fonte: Adaptado de Gonzalez e Woods, 2010.

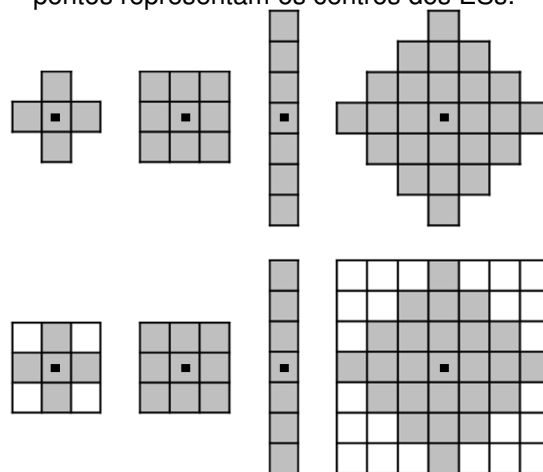
### 2.4.3. Processamento Morfológico de imagens

O processo de morfologia em imagens envolve a utilização de conceitos matemáticos como ferramenta para extrair elementos da imagem com representações úteis, tais como fronteiras, esqueletos ou fecho convexo.

#### 2.4.3.1. Elemento Estruturante

Ao efetuar operações morfológicas, normalmente se faz necessária a utilização de elementos estruturantes (ES), que são pequenos conjuntos utilizados para verificar uma imagem em busca de propriedades de interesse. Os elementos estruturantes são arranjos matriciais retangulares como mostrado na Figura 8.

Figura 8 – Elementos estruturantes; A primeira linha representa os elementos estruturantes. A segunda linha representa suas conversões em arranjos retangulares. E os pontos representam os centros dos ESs.



Fonte: Gonzalez e Woods, 2010.

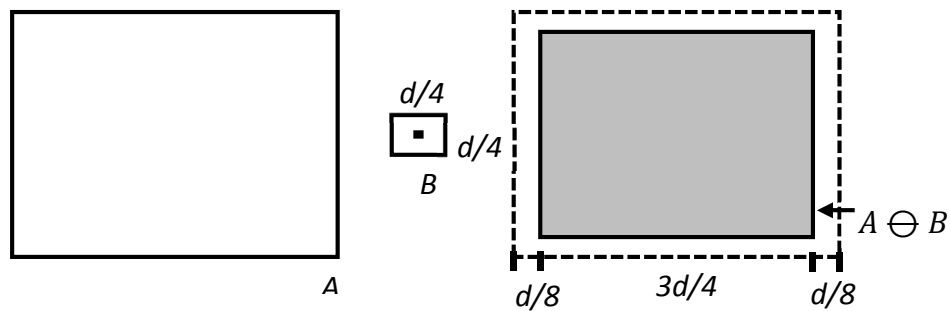
A partir do entendimento de ES, serão descritas nas próximas seções algumas operações morfológicas.

### 2.4.3.2. Erosão

O processo de erosão é indicado por  $A \ominus B$ , sendo que a erosão de A é feita por B, como mostrado na Equação 3. Pode-se dizer que a erosão de A por B é o conjunto de todos os pontos x tais que B, quando transladado por x fique contido em A, sendo B o elemento estruturante. Um exemplo de erosão é mostrado na Figura 9.

$$A \ominus B = \{ x \mid (B)_x \subseteq A \}. \quad (3)$$

Figura 9 – Processo de Erosão. Da esquerda para direita, o Conjunto A, o elemento estruturante B e a erosão de A por B mostrada sombreada.



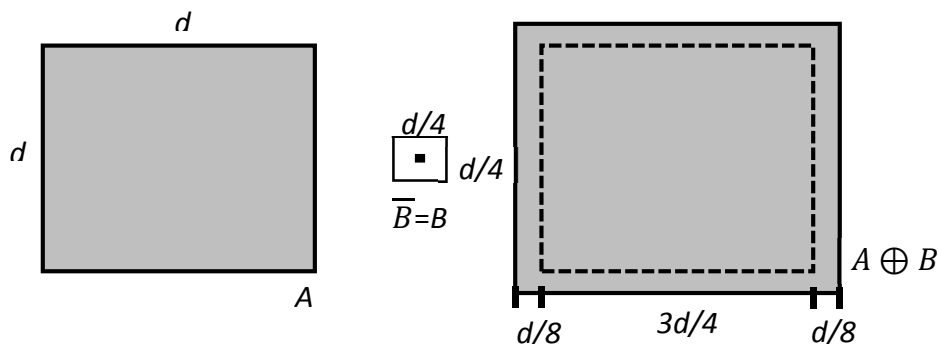
Fonte: Gonzalez e Woods, 2010.

### 2.4.3.3. Dilatação

O processo de dilatação é indicado por  $A \oplus B$ , em que a dilatação de  $A$  é feita por  $B$ , como mostrado na Equação 4. Pode-se afirmar que a dilatação de  $A$  por  $B$  é o conjunto de todos os deslocamentos  $x$  onde  $A$ , o conjunto (objetos de imagem) a ser dilatado, e  $B = \bar{B}$ , o elemento estruturante da dilatação, sobreponham-se em pelo menos um elemento. Um exemplo visual de dilatação é mostrado na Figura 10.

$$A \oplus B = \{x \mid (\bar{B})_x \cap A \neq \emptyset\}. \quad (4)$$

Figura 10 - Processo de Dilatação. Da esquerda para direita, o Conjunto  $A$ , o elemento estruturante  $B$  e a dilatação de  $A$  por  $B$  mostrada sombreada.



Fonte: Gonzalez e Woods, 2010.

### 2.4.3.4. Abertura e Fechamento

Os processos de abertura e fechamento são operações morfológicas realizadas com o intuito de suavizar contornos de um objeto. O primeiro elemento suaviza os contornos ao eliminar saliências finas, enquanto o segundo elimina pequenos buracos, funde discontinuidades e preenche lacunas no contorno, como ilustrado na Figura 11.

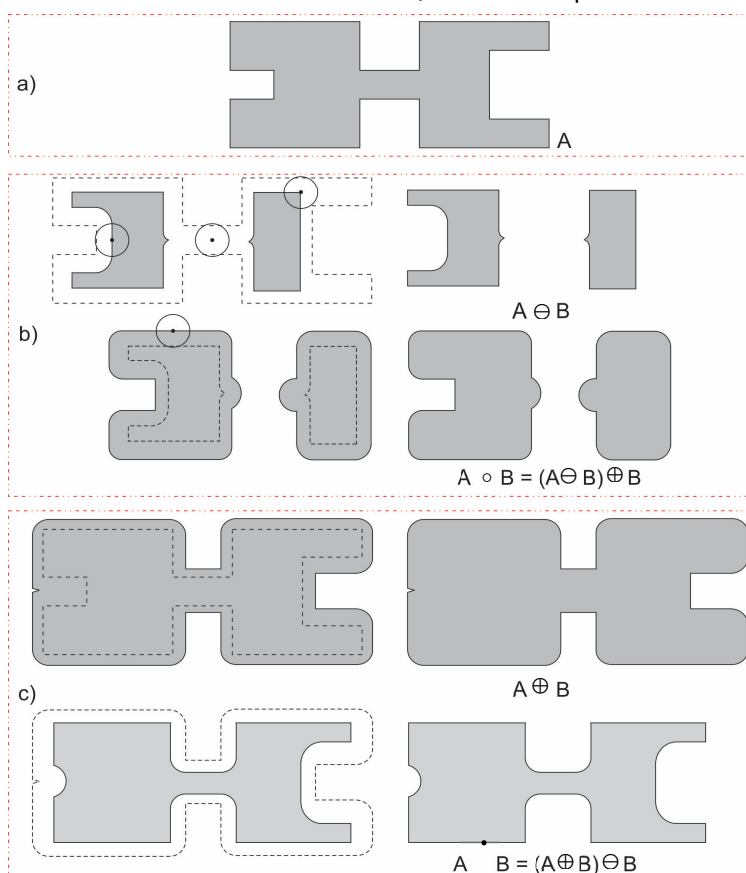
O processo de abertura é uma erosão de um objeto por um ES seguido de uma dilatação do resultado da erosão pelo ES. Este processo é indicado por  $A \circ B$  e ocorre de acordo com a Equação 5 mostrada a seguir.

$$A \circ B = (A \ominus B) \oplus B. \quad (5)$$

Ao contrário da abertura, o processo de fechamento é uma dilatação de um objeto por um ES, seguindo por uma erosão do resultado da dilatação feita pelo ES. O fechamento é indicado por  $A \bullet B$  e denotado pela Equação 6.

$$A \bullet B = (A \oplus B) \ominus B. \quad (6)$$

Figura 11 – Processo de Abertura e Fechamento. Na divisão *a*) o conjunto A. Na divisão *b*), na primeira linha, uma erosão feita pelo elemento estruturante B (pequeno círculo mostrado em várias posições) sobre o conjunto A. Ainda nesta divisão, na segunda linha o processo de dilatação com o resultado da erosão da primeira linha, formando um processo de abertura. Na divisão *c*), na primeira linha uma dilatação em A, já na segunda linha é realizado uma erosão com o resultado anterior, formando o processo de fechamento.



Fonte: Gonzalez e Woods, 2010.

#### 2.4.3.5. Preenchimento de buracos

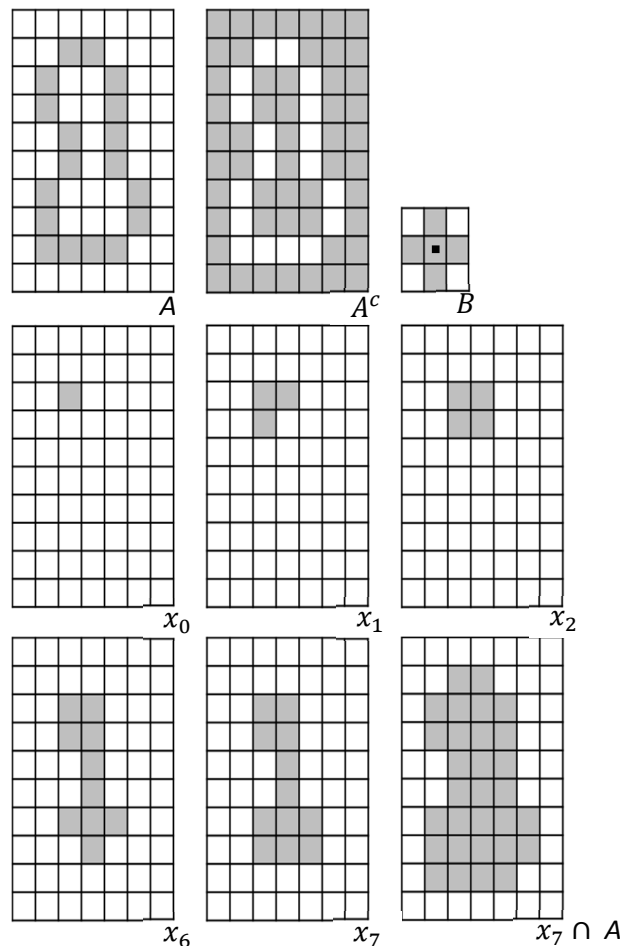
A técnica de preenchimento de região procura, a partir de um pixel dentro de uma região definida como borda, preencher a região com pixels ligados até o limite da borda. A equação chave para essa operação é mostrada na Equação 7, onde  $X_0$  é simplesmente um ponto dentro da borda,  $B$  é o elemento estruturante e  $A^c$  é o complemento de  $A$ . Esta equação deverá ser aplicada até

o momento em que  $X_k$  seja igual a  $X_{k-1}$ . Por fim, o resultado é unido com a fronteira original, como mostrado na Figura 12.

$$X_k = (X_{k-1} \oplus B) \cap A^c, \quad k = 1, 2, 3, \dots \quad (7)$$

Este algoritmo nada mais é do que um algoritmo que se utiliza de dilatação, complemento e interseção de conjuntos. O passo a passo visual é mostrado na Figura 12.

Figura 12 - Processo de Preenchimento. Na primeira linha, da esquerda para direita, a imagem original A, complemento de A e o elemento estruturante. Na segunda linha, da esquerda para direita, o ponto inicial da fronteira, as outras duas são os passos 1 e 2 da equação de preenchimento. Na terceira, da esquerda para direita, as duas primeiras, os passos 6 e 7 da equação e última, o resultado final que é a união do último passo com a imagem A.



Fonte: Gonzalez e Woods, 2010.

#### 2.4.3.6. Fecho convexo

Define-se como fecho convexo  $H$  de um conjunto arbitrário  $S$  como o menor conjunto de elementos convexos que ainda contém  $S$ . A diferença entre o conjunto  $H - S$  é denominada deficiência convexa. Estas características podem ser úteis para descrever um determinado objeto.

#### 2.4.4. Técnicas de segmentação/deteção de objetos em movimento

A segmentação em vídeos (sequências de imagens) é feita para detectar regiões que representam objetos em movimento nas cenas, como veículos (terrestre, marítimo ou aéreo) e seres vivos (animais, seres humanos, etc.). A deteção destas regiões com precisão, melhora ainda mais os processos posteriores como rastreamento e análise de comportamento. As principais abordagens de segmentação de vídeos são as seguintes: subtração de fundo, fluxo óptico e diferença temporal (JING e CHAOJIAN, 2011; HU. et al., 2004), descritas nas próximas seções.

##### 2.4.4.1. Subtração de Fundo

Técnica robusta que, primeiramente, constrói uma representação da cena, conhecida como modelo de fundo. Em seguida, tenta encontrar alterações para cada *frame* de entrada na imagem do vídeo. Portanto, o algoritmo de subtração de fundo é a diferença temporal entre o *frame* atual e o modelo subtraído. Assumindo que, para uma dada imagem,  $f$  representa a intensidade luminosa nas coordenadas  $x$  e  $y$  de um pixel no tempo  $t$ , a diferença  $\Delta I$  da imagem atual com a imagem de fundo do modelo extraído é dada como (GUJRATHI; PRIYA; MALATHI, 2014):

$$\Delta I = |f(x, y, t) - f_{fundo}(x, y)|. \quad (8)$$

##### 2.4.4.2. Fluxo Ótico

Essa técnica se utiliza de cálculos sobre a distribuição de intensidades para descobrir o deslocamento de um determinado pixel ocorrido em uma sequência de *frames* de um vídeo por meio da utilização de um vetor de movimento

relacionado a cada pixel (GUJRATHI; PRIYA; MALATHI, 2014). Assumindo que para uma dada imagem,  $I$  é a intensidade luminosa e  $x$  e  $y$  são coordenadas de pixels em tempo  $t$ , assumindo também que o intervalo de tempo  $dt$  entre duas imagens é curto e que a intensidade da imagem não se altera neste intervalo de tempo, representa-se o fluxo formalmente por:

$$I(x, y, t) = I(x + dx, y + dy, t + t). \quad (9)$$

#### 2.4.4.3. Diferença Temporal

Esse método utiliza vários *frames* consecutivos para subtrair o padrão de movimento aparente em uma cena. Essa identificação se restringe à validação dos objetos em movimento, tendo como principais dificuldades o reconhecimento da forma do objeto, assim como a identificação deste, caso venha a ficar parado na cena (GUJRATHI; PRIYA; MALATHI, 2014), (CHEN; MCGURR, 2014). Este método pode ser expressado formalmente da seguinte forma:

$$I_{dt(x,y,\tau)} = I(x, y, t) - I(x, y, t - \tau). \quad (10)$$

Onde,  $I(x, y, t)$  é o frame da imagem corrente e  $I(x, y, t - \tau)$  é o frame anterior em  $\tau$  segundos antes, sendo que o valor  $\tau$  em segundos será diretamente proporcional à velocidade do movimento do objeto e seu comprimento (CHEN; MCGURR, 2014).

## 2.5. PROCESSAMENTO DE ALTO NÍVEL

O processamento de alto nível refere-se à etapa de gerar um sentido a um conjunto de objetos detectados com os processamentos anteriores de baixo e médio nível, normalmente por meio de reconhecimento de padrões, conforme descrito a seguir.

### 2.5.1. Reconhecimento de Padrões

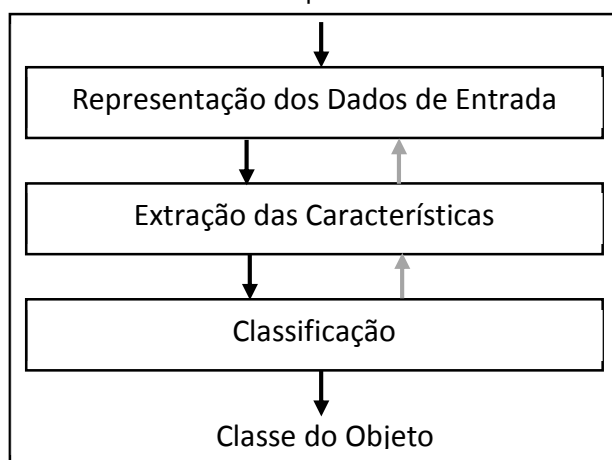
Segundo THEODORIDIS e KOUTROUMBAS (2006), reconhecimento de padrões refere-se a entender uma ordem estabelecida das propriedades características de um objeto que possibilite alocar este em uma classe correta.



O reconhecimento de padrões clássico envolve pelo menos três grandes etapas, sendo elas: a representação dos dados de entrada, a extração das características e por fim, a classificação, como mostrado na Figura 13.

A primeira fase preocupa-se com os dados de entrada e suas representatividades. Nesse caso, cada objeto de entrada pode ser descrito por um vetor  $X = [x_1, x_2, x_3, \dots, x_n]$ , onde  $x_1, x_2, x_3, \dots, x_n$  são suas características. A segunda etapa busca a redução da dimensionalidade do vetor padrão, extraindo apenas as características principais e mais discriminativas do determinado objeto. Por fim, efetua-se a classificação do objeto em destaque a fim de identificá-lo como pertencente a um determinado grupo que possui padrões muito parecidos ou iguais aos padrões do objeto a ser classificado (THEODORIDIS; KOUTROUMBAS, 2006), (DUDA; HART; STORK, 2001).

Figura 13 - Fases do Reconhecimento de Padrões Clássico: Os dados de entrada são representados em formato textual ou numérico. Em seguida ocorre a extração de características, que obtém elementos considerados importantes para serem analisados (frequência, formato, cor). Por fim, a classificação é responsável pela diferenciação dos dados em suas respectivas classes.



Fonte: Adaptado de Duda; Hart e Stork, 2001.

### 2.5.1.1. Extração de Características

A tarefa de reconhecimento de padrões investigada neste trabalho tem como objetivo a detecção de embarcações. Trata-se de uma tarefa relativamente trivial para os seres humanos, porém, ao ser desenvolvida por meio de sistemas automatizados, essa tarefa torna-se extremamente dependente de características relevantes que descrevam os dados de entrada, assim como

ocorre em outros problemas de reconhecimento de padrões automático. No contexto específico de dados de entrada na forma de imagens, estas devem ser capturadas e devidamente armazenadas a fim de serem submetidas à etapa de extração de características relevantes, para, então, tornarem-se entrada para um ou mais algoritmos de classificação (ASHT; DASS, 2012). Em seguida mostra-se as principais características que podem ser extraídas de uma imagem para sua classificação.

- **Forma:** Diferentes atributos podem ser considerados como formas em movimento, dentre eles destaca-se as representações por pontos, bordas, caixas ou regiões (bolhas) em uma imagem (PAREKH; THAKORE; JALIYA, 2014).
- **Geometria:** Os recursos geométricos são eficientes para discriminar um determinado objeto com finalidade de reduzir falsas detecções, dentre estes evidencia-se comprimento, largura e razão circular de um objeto detectado (Shi et al.,2013).
- **Movimento:** Os movimentos podem determinar informações periódicas que servem como características viáveis para classificação de objetos/veículos, pois podem mostrar determinado comportamento de um certo objeto na cena (PAREKH; THAKORE; JALIYA, 2014).
- **Cor:** A cor pode ser uma característica bastante salutar dentro de um sistema de classificação por ser relativamente constante em relação à mudança do ângulo de visão da câmera e também por ser facilmente adquirida e possuir custo computacional baixo. Contudo, não é aconselhável que se utilize apenas esse tipo de característica, pois informações de cor podem facilmente gerar confusões com outros objetos em cena (PAREKH; THAKORE; JALIYA, 2014).
- **Textura:** Trabalha com ocorrências de orientações de gradientes em regiões delimitadas de uma imagem, calculando os dados dos pontos

espaçados por meio de uma sobreposição de normalização de contraste (PAREKH; THAKORE; JALIYA, 2014).

Por fim, em muitas aplicações também são utilizadas derivações de escala de invariâncias,  $\eta_{ij}$ , de momentos de ordem  $(p + q)$ ,  $M_{pq}$ , que são médias ponderadas de pixels de intensidade de imagens, a fim de detectar por meio das derivadas de um conjunto de momentos, uma imagem em translação, mudança de escala, espelhamento ou rotação.

Após a extração de características, avança-se para o processo de classificação das instâncias para que a tarefa de reconhecimento de padrões seja finalizada.

#### **2.5.1.2. Classificação**

A classificação é o processo de atribuição de instâncias desconhecidas à uma das classes, ou grupos, do problema. O classificador pode ser um modelo gerado via aprendizagem de máquina. Nesse caso, primeiramente ocorre a aprendizagem dos padrões por algum algoritmo, e, em seguida, o modelo é gerado com base no processo de aprendizagem. Na literatura, os algoritmos de aprendizagem de máquina se dividem basicamente em três grupos: aprendizagem supervisionada, não supervisionada e por reforço.

Quando há instâncias de dados rotuladas, as quais são representadas pelo conjunto de características extraídas no passo anterior, para treinar o classificador para que o mesmo consiga aprender os padrões das instâncias, este processo é chamado de aprendizagem supervisionada. Porém, quando não se tem os rótulos das instâncias e o algoritmo deve se encarregar de aprender sobre o padrão das características dos objetos para poder agrupá-los por similaridade, esse processo é chamado de aprendizagem não supervisionada. É possível ainda, haver um terceiro tipo de classificação que trabalha com aprendizagem por reforço, onde inicialmente os objetos não são rotulados. Nesse caso, o algoritmo agrupa as instâncias e tenta aprender a informação correta durante o aprendizado por meio de *feedback* do ambiente (THEODORIDIS; KOUTROUMBAS, 2006), (DUDA; HART; STORK, 2001).

Dentre os algoritmos que trabalham com classificação supervisionada, pode-se citar o *Support Vector Machine* (SVM) que, dado um conjunto de exemplos de treino, constrói um modelo de separação que maximiza a distância entre as instâncias de treinamento de classes distintas, a fim de realizar predição de novos exemplos de forma precisa. Também existe o classificador baseado em árvore de decisão, que é um fluxograma onde cada nó representa um "teste" em uma característica e cada ramo representa o resultado do teste. Assim, o caminho da raiz à folha representa as regras de classificação (WITTEN et al, 2011).

Adicionalmente se pode mencionar o algoritmo *Randon Forest*, que trabalha com um conjunto de classificadores do tipo árvore de decisão. Há também o método *K-Nearest Neighbors* (KNN), o qual calcula a similaridade entre instâncias desconhecidas e as instâncias de treino para classificar cada instância desconhecida com base no rótulo das suas k instâncias de treino mais próximas. Por fim, o *k-means*, que é um método de agrupamento cujo objetivo é particionar os exemplos de dados, a partir dos valores de suas características, entre uma determinada quantidade de grupos. Cada exemplo pertencerá ao grupo de média de valores característicos mais próxima à sua (WITTEN et al, 2011).

Para avaliar quão preciso é o processo de classificação, utiliza-se algumas métricas existentes que serão apresentadas na seção seguinte.

## 2.6. MÉTRICAS DE AVALIAÇÃO

Em boa parte de todos os trabalhos relacionados da literatura, pelo menos quatro métricas básicas são utilizados para avaliação do desempenho em cada quadro do vídeo. Estas são: (1) Verdadeiros Positivos (VP), (2) Verdadeiros Negativos (VN), (3) Falsos Positivos (FP) e (4) Falsos Negativos (FN).

- **Verdadeiros Positivos (VP):** A rotulação de entrada dos dados coincide com a saída obtida quanto à presença de um objeto.
- **Verdadeiros Negativos (VN):** A rotulação de entrada dos dados coincide com a saída obtida quanto à ausência de um objeto.

- **Falsos Positivos (FP):** A rotulação de entrada dos dados afirma a não existência de um objeto enquanto a saída afirma a presença deste.
- **Falsos Negativos (FN):** A rotulação de entrada dos dados afirma que existe um objeto, porém a saída afirma que não existe objeto na cena.

As métricas informadas são utilizadas para calcular os pares de métricas de desempenho global do algoritmo, considerando as várias imagens existentes. Essas métricas são descritas em WITTEN (2011). Dadas diferentes imagens a avaliar, a taxa de VP é definida como a soma de todos os verdadeiros positivos que ocorreram nas imagens consideradas. Após as explicações anteriores, apresentar-se-á cinco indicadores que serão utilizados para avaliar o desempenho do método proposto neste trabalho.

- **Precisão (PR):** Taxa com que todos os exemplos classificados como positivos são realmente positivos.

$$PR = \frac{VP}{VP+FP} \cdot \quad (11)$$

- **Revocação ou Sensibilidade (RC):** Taxa que classifica como positivos todos os exemplos que são positivos.

$$RC = \frac{VP}{VP+FN} \cdot \quad (12)$$

- **Acurácia (AC):** Taxa que informa que os resultados preditos realmente são corretos.

$$AC = \frac{VP+VN}{VP+FP+FN+VN} \cdot \quad (13)$$

- **Medida F (F):** pode ser interpretada como uma média harmônica entre precisão e revocação.

$$F = 2 * \frac{PR*RC}{PR+RC} \cdot \quad (14)$$

A próxima seção apresenta com mais detalhes o problema investigado neste trabalho que é a detecção de embarcações.

## 2.7. DETECÇÃO DE EMBARCAÇÕES

Em grande parte dos sistemas para controle do tráfego marítimo, é comum a utilização de radares para a detecção de embarcações, uma vez que é uma tecnologia madura e de ampla utilização (PIERALICE et al., 2014; ATTEIA e COLLINS, 2014; LIU et al., 2014). Porém, existem algumas desvantagens com esse tipo de abordagem, tais como: custo financeiro elevado, as imagens não fornecem informações suficientes sobre o objeto detectado, assim como pequenas embarcações e objetos não metálicos não são de fácil detecção (WIJNHOFEM et al., 2010; AMARASINGHE et al., 2014). Por fim, as imagens de radar são muito confusas e de difícil identificação de objetos em movimento (WIJNHOFEM et al., 2010; QIUYU et al., 2013).

Por outro lado, em vários portos são utilizadas câmeras de segurança para dar visão global do ambiente aos operadores a fim de controlar no tráfego das embarcações e na salvaguarda de setores dos portos. Porém, as câmeras não são normalmente utilizadas de uma forma automática. Se utilizadas na tarefa de monitoramento automático, as câmeras podem superar as desvantagens apresentadas pelos sistemas de radares, tornando-se uma opção interessante devido ao baixo custo, facilidade de implantação em vários tipos de plataforma, flexibilidade e fácil gerenciamento (BAO et al., 2013; LI et al., 2013). Embora existam inúmeras pesquisas voltadas para a detecção de veículos em estradas, as análises de vídeos para detectar embarcações não são amplamente utilizadas e permanecem bastante instigantes devido ao fato das embarcações localizarem-se em regiões com grande diversidade de fundo, dificultando a construção de modelos correspondentes (BAO et al. 2013).

É também importante mencionar que as estratégias tradicionais para detecção de objetos por vídeo, se utilizadas sem as devidas adaptações ao ambiente marítimo, não produzem resultados satisfatórios, pois características específicas desse ambiente como intempéries, possível movimentação da plataforma onde a câmera está instalada, oclusões e inúmeras outras (MOREIRA, 2015) afetam significativamente o desempenho das técnicas.

Entretanto, a detecção é fundamental para o funcionamento adequado dos demais componentes de um sistema automático de segurança por vídeos.

Devido à diversidade dos ambientes marinhos, é comum existir uma alta taxa de falsos positivos (FP) e falsos negativos (FN). Os trabalhos mais recentes, que podem ser considerados estado-da-arte, utilizam subtração de fundo (imagem do ambiente) e funções gaussianas para modelar os valores de pixels do mar, uma vez que a estratégia de utilizar o fluxo ótico possui uma complexidade computacional elevada, dificultando a utilização do sistema em tempo real. Além disso, a diferença entre quadros pode aumentar a taxa de FN, quando a embarcação se mover em direção à câmera ou tiver um movimento muito curto (lento) (HU. et al., 2004; JING e CHAOJIAN, 2011; MOREIRA, 2014).

Neste capítulo mostrou-se a área de interesse deste trabalho: detecção de embarcação. Foram esclarecidos os conceitos principais, a fim de possibilitar a identificação de algumas características existentes nos métodos e sistemas de detecção de embarcações que serão apresentados no próximo capítulo.

### 3. TRABALHOS RELACIONADOS

Neste capítulo serão abordados os principais trabalhos relacionados ao tema desta pesquisa, ou seja, métodos de detecção de embarcação. Tais trabalhos têm como foco a proteção e o monitoramento do tráfego de embarcações em determinados ambientes a fim de agilizar o processo de monitoramento em vídeos de segurança marítimos. Porém os trabalhos aqui elencados não trabalham no ambiente alvo desta dissertação, que é o ambiente amazônico. Os trabalhos relacionados utilizam, em sua maioria, o ambiente de mar aberto o qual possui apenas a embarcação e a água, e no máximo o horizonte onde se encontra o céu e o mar.

Os trabalhos encontrados na literatura podem ser divididos em três categorias no que diz respeito à redução dos falsos positivos: (1) trabalhos que utilizam subtração de um modelo de fundo e limiarização; (2) métodos que rastreiam e analisam a permanência de um objeto em uma quantidade de quadros; e (3) detectam embarcações em uma cena marítima por meio de algoritmos de aprendizagem de máquina. Apresentam-se os trabalhos relacionados seguindo a ordem cronológica decrescente por categoria apresentada.

Primeiramente se lista os trabalhos que utilizam a subtração de fundo como forma de segmentação das cenas para detectar apenas o objeto em movimento.

Liu *et al.* (2014) apresentam em seu trabalho um método eficiente para segmentação de embarcações em imagens infravermelhas. Esse método é dividido em duas etapas. A primeira etapa é a segmentação interativa da imagem, na qual um filtro de subtração global de fundo é utilizado para retirar o fundo primário. Posteriormente, é utilizado um filtro de subtração de média de linha adaptado para realçar a embarcação e por fim, depois das aplicações interativas dos dois filtros, um limiar adequado para a segmentação da imagem é determinado.

A segunda etapa envolve a utilização de propriedades geométricas das embarcações para construir quatro características de forma: área, capacidade,



razão entre largura e altura, e razão entre parte superior e inferior. Por fim, nesta etapa é utilizado um critério de seleção para identificar a embarcação e remover as regiões que não correspondem a embarcações, levando em consideração o limiar para cada característica extraída.

As métricas utilizadas para avaliação do resultado foram: erro de classificação – valor obtido foi de 0.0259; e erro da área relativa ao alvo (*relative foreground area error*), que atingiu um valor de 0.3594 de taxa. Diferentemente da nossa pesquisa, este trabalho utiliza uma base formada por imagens infravermelhas, características que aumenta o realce das embarcações, capturadas de uma câmera instalada na proa da embarcação apontada para frente. Além disso, apenas a técnica de limiarização é utilizada para remoção de falsos positivos detectados.

Shi et al. (2013) propõem em seu trabalho um método simples e eficaz para detecção de embarcações baseado na subtração de fundo e informações da borda em duas etapas principais. Primeiramente, os pixels que estão em movimento são detectados por meio de um algoritmo de subtração de fundo da imagem no vídeo. Em seguida, as bordas são detectadas na máscara do fundo. Por fim, é realizada a transformada de distância da imagem, que é o mínimo da função de uma distância euclidiana entre um conjunto de pontos  $P$  e outro conjunto de pontos  $B$ , para cada imagem de borda. A segunda etapa principal envolve a detecção das bordas na imagem original por meio de operações morfológicas, utilizadas para realçar as bordas encontradas. Por fim, as bordas da imagem original são combinadas com a transformada de distância da imagem, preenchendo os buracos da área dos componentes das bordas. A embarcação é detectada através de uma operação de discriminação com base em características como área, largura, comprimento e razão entre largura e comprimento, segundo as Equações 15 e 16.

$$Complex = \frac{4\pi*s}{p^2}. \quad (15)$$

$$Compact = \frac{s}{w*h}. \quad (16)$$

Onde  $w$  é a largura do alvo,  $h$  representa o comprimento,  $s$  é a área e  $p$  o perímetro das regiões de bordas encontradas. Esse trabalho não utiliza

aprendizado de máquina, porém nos auxiliou na escolha de características geométricas que são invariantes à rotação.

Nos estudos de Hu *et al.* (2011), os autores apresentam um método de segurança visual para detecção e rastreamento automáticos de embarcações utilizado em gaiolas de aquicultura. O método é dividido em 4 etapas, sendo que a primeira é a detecção dos objetos em movimento, que utiliza uma subtração de fundo com atualizações no decorrer do vídeo para reduzir a influência das ondas. Posteriormente, um algoritmo adaptado de busca de 4 regiões conectadas é utilizado para reduzir ruídos das imagens, focando apenas nas áreas que possuem quantidade de pixel maior que um limiar.

Após o procedimento de encontrar os componentes conectados, a imagem ainda possui alguns ruídos produzidos pelas ondas do mar. Esses ruídos são retirados com um algoritmo de remoção de sombras. Dessa forma, o método passa a ter como foco apenas a embarcação. Por fim, é realizado o rastreamento da embarcação no vídeo com um algoritmo baseado em região. Assim como os dois trabalhos anteriores, que utilizam subtração de fundo, esse trabalho não utiliza nenhuma técnica de aprendizado de máquina para filtrar as falsas detecções.

Os próximos quatro trabalhos utilizam rastreamento e métodos de redução de FP por permanência em uma quantidade de quadros do objeto rastreado para que este possa ser considerado uma embarcação.

Marques *et al.* (2014) sugerem um algoritmo de detecção de embarcações em imagens aéreas que é dividido em três etapas: i) detecção de embarcação; ii) restrições espaciais e; iii) consistência do tempo. A detecção da embarcação é realizada primeiramente por um simples método de limiar gerando uma imagem binária, assumindo que as embarcações possuem um valor muito alto no RGB. Em seguida, nesta imagem é executada uma separação em blocos que não se sobrepõem, onde cada bloco é caracterizado por suas características de cor e classificado em embarcação ou fundo, gerando uma segunda imagem binária.

Na etapa de restrições espaciais são utilizadas características das regiões de interesse, onde para ser considerada embarcação, a área da região não deve

ultrapassar um valor máximo estabelecido por um limiar. Assume-se que a região de interesse não tocará na borda da imagem e estará isolada, ou seja, possuirá uma distância maior que a de um limiar estabelecido das outras regiões encontradas. Por fim, utiliza-se da validação temporal, que será a comparação entre o número de ocorrências de uma região de interesse de mesma área nos últimos quadros com um valor de limiar. O método obteve desempenho de precisão variando de 99,4 a 100% e valores de sensibilidade de 69,3 a 76,7%. Mesmo alcançando valores excelentes nos resultados, observa-se que por não possuir uma etapa de aprendizado dificultará esta detecção no ambiente proposto nesta dissertação, uma vez que a diferença entre um ambiente de mar aberto e os cenários amazônicos possuem uma discrepância altíssima.

Loomans *et al.* (2013) utilizam um método de detecção e rastreamento de embarcações usando câmeras ativas. Para a detecção do alvo, os autores utilizam um detector baseado em *Histogram of Oriented Gradients* (HOG) de alta complexidade computacional, cujo detector é treinado para a cabine de comando das embarcações - ponto em comum entre todos tipos de embarcação.

Em seguida, é efetuado o rastreamento com o algoritmo hierárquico Kanade–Lucas–Tomasi (KLT) para pontos característicos do HOG extraído, o qual possui uma baixa complexidade computacional. Os resultados obtidos indicam um rastreamento máximo por aproximadamente 1,5 Km durante 4 min e 37 s, sendo que o rastreamento para pôr oclusão do alvo. Contudo as características utilizadas neste trabalho são inviáveis ao método proposto, pois o ângulo de aquisição da base de vídeos é diferente, e as mesmas são extremamente variantes a rotação.

FefilatyeV et al. (2010) demonstram em seu trabalho uma abordagem para segurança visual marítima com câmera não estacionada com a utilização de um algoritmo que detecta e rastreia embarcações. Primeiro, é utilizada a transformada de Hough aplicada ao mapa de bordas para achar as possíveis linhas do horizonte. Em seguida, é usado um critério de discriminação entre duas regiões de cores homogêneas: o céu e a água. Para suavizar artefatos gerados pela compressão do formato MJpeg, é utilizado um filtro gaussiano na região do céu em que existe a possibilidade do aparecimento da parte de uma embarcação, a qual compreende a região formada pelas 100 linhas de pixels

acima da linha do horizonte. Para detectar o alvo, é utilizado um filtro de gradiente para transformar a imagem colorida em uma imagem em escala de cinza, na qual será utilizado o método Otsu. Também é realizado um refinamento para retirar alarmes falsos, no intuito de remover os objetos detectados que possuem mais de 20% de pixels da imagem e que possuem menos de 10 pixels.

Desse modo, ocorre o registro das imagens para realizar comparações entre o frame atual e o anterior para correlacionar. Também é utilizado um *framework* de rastreamento baseado em Múltiplas Hipóteses, sendo que as hipóteses são geradas pelo filtro de Kalman. O rastreamento se encerra quando o número de rastreamentos registrados é menor que a metade dos quadros seguintes do vídeo. Foi utilizada a métrica *Sequence Frame Detection Accuracy* (SFDA), cujo melhor desempenho sem a utilização do rastreamento foi de 64,97% e com rastreamento de 74,40%. A utilização deste tipo de abordagem no ambiente proposto por este trabalho é inviável pelo fato que a câmera será direcionada de cima para baixo não existindo a linha do horizonte que é um dos alicerces do trabalho de FefilatyeV et al. (2010).

Outra técnica para detecção e rastreamento automático de embarcações é apresentada por FefilatyeV e Goldgof (2008), cujo foco é utilizar vídeos de mar aberto capturados por uma câmera digital instalada em uma boia. Na primeira parte da técnica, são utilizados métodos de detecção de bordas e de detecção de horizonte. Em seguida, é efetuado um pós-processamento com a aplicação de um filtro de extração de ruídos e a utilização dos processos morfológicos de erosão e dilatação para conectar bordas de um mesmo objeto e retirar possíveis ruídos. Por fim, no processo de detecção é usado um algoritmo de componentes conectados para detectar e inserir uma caixa delimitadora na embarcação.

A etapa de detecção possui algumas falhas, tais como: (i) se a linha do horizonte não for exatamente uma reta ou (ii) quando o alvo possuir inúmeras bordas, problema ocasionado por foco incorreto ou posicionamento da câmera. Com o objeto detectado, é calculado o centro da caixa delimitadora e o processo de rastreamento é efetuado com a utilização do filtro de Kalman. Porém, o rastreamento só inicia se o objeto detectado permanecer em dois quadros consecutivos. Se o objeto detectado não for constante nos demais quadros, o rastreamento é finalizado. A métrica SFDA foi utilizada, sendo que o melhor

desempenho sem a utilização do rastreamento foi de 55,45%, enquanto que com rastreamento, o melhor desempenho foi de 91,34%.

E, finalmente, são apresentados três trabalhos que utilizam aprendizagem de máquina para a detecção da embarcação na cena.

Bechar *et al.* (2014) propõem um método de segmentação de objeto em um fundo dinâmico usando critérios fixos pixel a pixel em uma aplicação para reconhecimento de alvos marítimos. Primeiramente, a segmentação do objeto é realizada com a transformação de um vídeo 2D com movimentação da câmera e do objeto alvo em um modelo 3D, onde o objeto continua com sua movimentação e a observação da câmera torna-se estática.

Posteriormente, são empregados os critérios fixos pixel a pixel como a quantificação da rigidez da vizinhança dos pixels a partir dos modelos de primeira ordem do fluxo ótico. Por fim, utiliza-se classificação bayesiana para separar fundo do alvo e emprega-se o algoritmo de agrupamento *K-means*, o qual gera um conjunto de agrupamentos dos pixels de um quadro do vídeo para otimizar a separação da região alvo da imagem. Esta região é definida como a região com menor média e maior área.

No trabalho de Chan e Weed (2012) é apresentado uma proposta de um algoritmo de detecção de embarcação baseado em dois componentes. O primeiro é a detecção do alvo a partir da aprendizagem de máquina efetuado com o algoritmo SVM utilizando HOG retirado da base de dados VOC2010, a qual contém 600 embarcações para a base de teste positiva e imagens do ambiente marítimo sem embarcações para a base de teste negativa.

O segundo componente utiliza o *Principal Component Analysis* (PCA) para aprender, baseado na textura, a dinâmica espaço-temporal do modelo do fundo marítimo. A fusão dos dados da aprendizagem do alvo com SVM com a modelagem do fundo com o PCA faz com que a detecção da embarcação sofra uma melhoria ao diminuir a taxa de falsos positivos.

Bao *et al.* (2013) propõem em seu trabalho um método de detecção de embarcação para um sistema de vigilância. O método proposto combina modelagem baseada em contexto com análise de movimento. Para usar a modelagem contextual, é realizada uma segmentação da imagem, separando

áreas de vegetação e água, enquanto o restante é considerado desconhecido. Na sequência são extraídas características para análise das regiões, tais como: cor, textura e valores da posição vertical das regiões. Por fim, utiliza-se SVM com a abordagem de um contra todos para classificar cada uma das regiões.

Após essas etapas de classificação de regiões são utilizadas a modelagem contextual, sendo enfocados primeiramente os contextos semântico e espacial, assumindo que as embarcações viajam apenas dentro da água e não podem estar totalmente dentro do ambiente de vegetação, assim as regiões desconhecidas que preencherem estes critérios são candidatas a serem embarcações.

No contexto escalar, assume-se que a possível embarcação tem uma área maior que 600 pixels e menor que 50% da quantidade de pixels da (as) região (ões) classificadas como água. As regiões candidatas a embarcações são realmente classificadas como embarcações com a análise da ênfase do movimento pelo fluxo ótico, assumindo que as embarcações se movem significativamente em relação aos objetos próximos. O método de detecção de embarcação atingiu uma precisão entre 94% a 97% com sensibilidade de 75% a 92,7%.

### 3.1. ANÁLISE DE CONTRIBUIÇÃO DOS TRABALHOS RELACIONADOS

Como observado, existem inúmeros métodos e sistemas de detecção de embarcações em ambientes marítimos. Na Tabela 1 apresenta-se um resumo das técnicas utilizadas em cada abordagem descrita neste capítulo, destacando o processo de aquisição de imagens, as técnicas utilizadas para segmentação, as técnicas utilizadas para redução dos falsos positivos e as características utilizadas. Pode ser observado que a maioria das técnicas utiliza rastreamento de objeto em um número  $n$  de quadros da sequência dos vídeos, uma vez que esses trabalhos assumem que as câmeras apresentam uma alta taxa de deslocamento (alteração brusca da cena). Porém, neste trabalho focar-se-á na utilização da estratégia de subtração de fundo, pois se assumi que os vídeos são capturados de uma câmera fixa, facilitando a retirada de um modelo do fundo.

Pode-se observar na Tabela 1 que existem 3 maneiras utilizadas pelos métodos e sistemas de detecção de embarcação: por subtração de fundo e limiarização; por rastreamento e permanência em quadros; e por aprendizagem de máquina. Porém, nenhum trabalho utiliza técnicas combinadas. Nesse trabalho combina-se subtração de fundo com aprendizagem de máquina para melhorar a qualidade de detecção das embarcações no contexto de cenário amazônico. Além disso, utiliza-se vídeos capturados com uma câmera instalada no alto, apontada para baixo em 90 graus ou aproximado, sendo que todos os métodos listados neste capítulo foram desenvolvidos para ambientes de mar aberto e/ou com câmera apontadas para frente.

Por outro lado, o método proposto será composto por etapas básicas encontradas nos trabalhos relacionados, inclusive as características utilizadas para representar os dados. A principal diferença é a utilização em conjunto de subtração de fundo com aprendizagem de máquina, sendo que cada etapa foi ajustada ao ambiente de aplicação.

Para comparação e demonstração da eficiência do método proposto utilizar-se-á o trabalho de Marques et al. (2014), já que o mesmo se utiliza de uma base de dados mais próxima à nossa em seus experimentos, precisamente, câmera RGB localizada no alto e apontada para baixo a fim de realizar a detecção de embarcações e suas marcações.

Tabela 1 – Síntese de trabalhos relacionados e métodos utilizados em cada trabalho.

<b>Utilizam Subtração de Fundo e Limiarização</b>				
<b>Autor</b>	<b>Aquisição de Imagem</b>	<b>Técnica Utilizada Segmentação</b>	<b>Técnica de Redução de FP.</b>	<b>Características Utilizadas</b>
LIU et al., 2014.	Câmera infravermelho instalada em uma embarcação.	Combinação subtração de fundo global com subtração de fundo por linhas.	Limiarização.	Formas (área, capacidade, razão entre altura e largura, razão entre partes superior e inferior).
Shi et al., 2013.	Câmera RGB estática.	Subtração de fundo e detecção de bordas.	Limiarização.	Área, largura (L), comprimento (C), razão LxC, complexo (razão circular) e compacto.

Hu, <i>et al.</i> , 2011.	Câmera RGB instaladas em gaiolas de agricultura aquática.	Subtração de fundo, componentes conectados.	Técnicas de remoção de sombras para remover ondas e limiarização.	Área do componente conectado.
Utilizam Rastreamento e Permanência em Quadros				
Autor	Aquisição de Imagem	Técnica Utilizada Segmentação	Técnica de Redução de FP.	Características Utilizadas
Marques <i>et al.</i> , 2014.	Câmera RGB instalada em um veículo areo não tripulado.	Limiarização, validação espacial (tamanho da região encontrada (RE), RE não deve tocar nas bordas e a RE deve estar isolada).	Permanência do objeto em uma quantidade de quadros.	Componentes conectados.
Loomans, <i>et al.</i> , 2013.	Câmera RGB, tipo pan-tilt-zoom (PTZ) instalada em um porto.	-	KLT (Kanade-Lucas-Tomasi) e HOG detector.	HOG da cabine.
Fefilatyevev <i>et al.</i> , 2010.	Câmera RGB instalada em uma boia.	Transformada de Hough para detectar a linha do horizonte, filtro gaussiano, limiarização.	Permanência do objeto, rastreado pelo filtro de Kalman em um uma quantidade de quadros.	Objeto rastreado pelo Kalman.
Fefilatyevev e Goldgof., 2008.	Câmera RGB instalada em uma boia.	Combinação das técnicas Canny para detectar bordas e detector de linha do horizonte.	Permanência do objeto em uma quantidade de quadros.	Componentes conectados (área da caixa delimitadora).
Utilizam Aprendizagem de Máquina				
Autor	Aquisição de Imagem	Técnica Utilizada Segmentação	Técnica de Redução de FP.	Características Utilizadas
Bechar <i>et al.</i> , 2014.	Câmera RGB instalada em um avião.	Utiliza vetor de movimentação dos pixels (fluxo óptico) e uma classificação bayesiana.	K-Means.	Vetor de movimentação dos pixels, após a classificação bayesiana.



Chan e Weed, 2012.	Câmera RGB estática.	SVM para analisar a embarcação e PCA para analisar o ambiente.	Fusão dos resultados da segmentação	Componentes conectados e HOG.
Bao <i>et al.</i> , 2013.	Câmera RGB, do tipo PTZ instalada em um porto.	Classificação do ambiente via SVM e modelagem estatística (água, vegetação e desconhecido, modelagem de contexto (deve conter água).	Análise do movimento (assume que a embarcação se move de forma significativa em relação aos demais objetos).	Fluxo Ótico.
<b>Autor</b>	<b>Aquisição de Imagem</b>	<b>Técnica Utilizada Segmentação</b>	<b>Técnica de Redução de FP.</b>	<b>Características Utilizadas</b>
Método Proposto.	Câmera RGB instalada em um aeróstato fixo apontada a 90 grau para baixo.	Subtração de fundo e Limiarização.	Aprendizagem de Máquina.	Razão circular, Excentricidade, Razão entre Largura e Altura e Momentos Invariantes.

Fonte: O Autor, 2016.

Neste capítulo foi apresentada uma revisão da literatura sobre detecção de embarcação a partir de imagens capturadas por câmeras, uma vez que grande parte dos resultados retornados usa imagens capturadas por radares de abertura sintética. No próximo capítulo apresenta-se o método proposto para detecção de embarcação em imagens de vídeo.

## 4. MÉTODO PROPOSTO

Neste trabalho é proposto um sistema de detecção de barcos em ambiente típico dos rios amazônicos. Conforme mencionado no capítulo 1, trata-se de um cenário difícil de ser analisado, levando em conta o *background* demasiadamente diversificado.

As etapas necessárias para a operacionalização do método proposto são apresentadas neste capítulo.

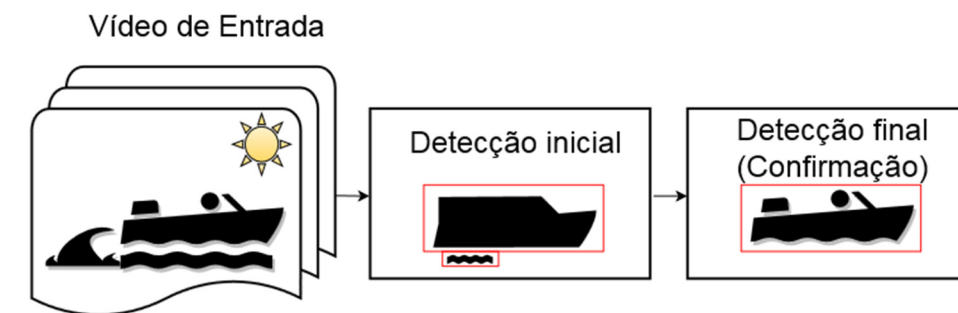
Inicialmente será discutida a arquitetura do método em alto nível. Em seguida, cada fase empregada no processo base deste trabalho de pesquisa é descrita em detalhes.

### 4.1. ARQUITETURA GERAL DO MÉTODO

A quantidade de quadros a serem analisados por um operador humano é extraordinariamente grande, levando em consideração que em apenas 5 (cinco) minutos possuem 9000 (nove mil) quadros a serem analisados, desta forma, se houver uma diminuição dos números de quadros a serem analisados, tem-se uma melhor análise do operador humano.

No intuito de se obter essa diminuição, é proposto neste trabalho um método composto de duas fases de detecção para obter um melhor resultado no processo de detectar a embarcação. A primeira fase realiza a detecção de todos os objetos em movimento da cena, pois assume-se que a embarcação se encontra em movimento. Então, na segunda fase é realizado um processo de classificação para refinar as detecções feitas na etapa anterior e para marcar nas imagens apenas as embarcações e retirar possíveis falsos positivos. A Figura 14 apresenta o fluxo de execução do método proposto, o qual é dividido em três etapas básicas: (1) aquisição (Vídeo de Entrada); (2) detecção; e (3) confirmação da detecção.

Figura 14 - Arquitetura proposta: O vídeo de entrada será capturado de uma câmera instalada no alto e apontada para baixo. Este vídeo passará pela fase de detecção inicial do movimento e, em seguida, as embarcações existentes na cena são marcadas na detecção final.



Fonte: O Autor, 2016.

Conforme mencionado no capítulo 1, a motivação deste trabalho é a detecção de embarcações a partir de imagens obtidas de câmera instalada em um aeróstato ancorado na região das reservas de Mamirauá e Amanã – Estado do Amazonas. Como o aeróstato será alçado a uma altura do solo pré-determinada e ancorado mecanicamente à base, assume-se que a câmera poderá ser instalada tanto na parte mais alta quanto ao longo do cabo de estaiamento e assumindo-se que a câmera já está devidamente calibrada. Com isso, a aquisição do vídeo será realizada a partir de uma plataforma aérea fixa, com a câmera voltada para baixo.

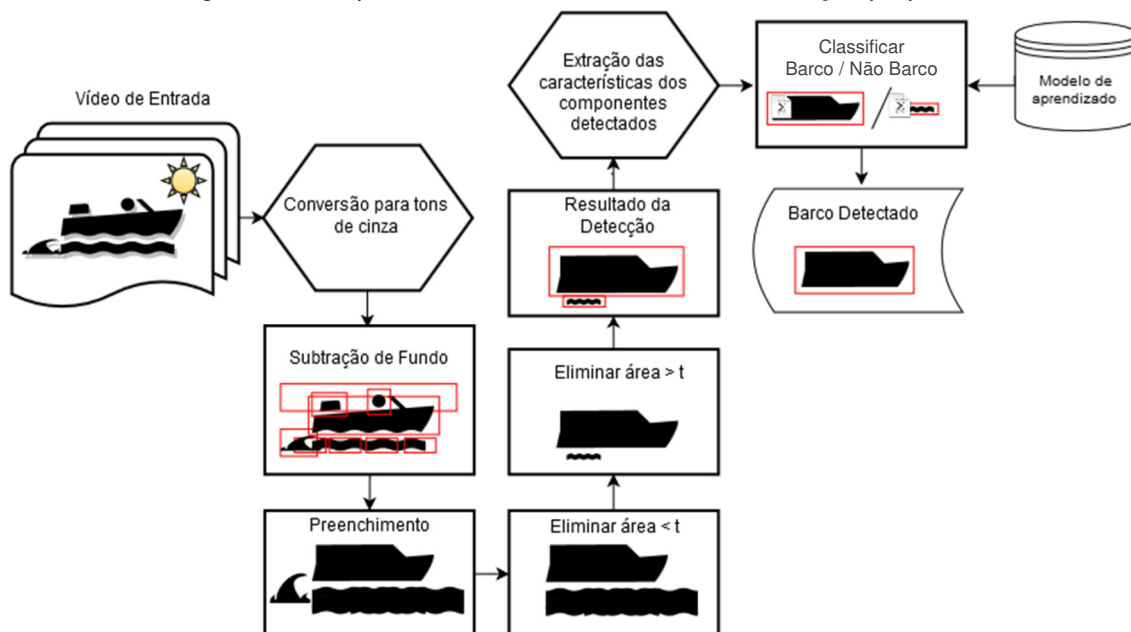
A primeira etapa da arquitetura é a detecção inicial, que é realizada através da extração dos objetos em movimento na cena, ou seja, esse método produz janelas/imagens dos objetos em movimento. Por meio da utilização de técnicas de processamento de imagens, busca-se diminuir possíveis ruídos nessa etapa.

Na segunda etapa da arquitetura, confirmação, ocorre a detecção final. Essa etapa é responsável por analisar os objetos detectados na etapa anterior e classificá-los em uma de duas classes possíveis: embarcação ou não embarcação. Esta última classe agrupa todos os demais objetos em movimento na cena (i.e., água, vegetação, etc.). Neste trabalho, a etapa de detecção final é realizada por algoritmo de aprendizagem de máquina. Portanto, as duas etapas do método proposto contemplam as três etapas básicas de um método de processamento digital de imagens: (1) pré-processamento, (2) processamento e (3) pós processamento.

A Figura 15 apresenta o método completo da proposta com todas as fases aplicadas nos vídeos adquiridos, detalhando as etapas mencionadas na

arquitetura em alto nível da Figura 14. Cada fase será descrita em mais detalhes a seguir.

Figura 15 – Arquitetura detalhada do método de detecção proposto



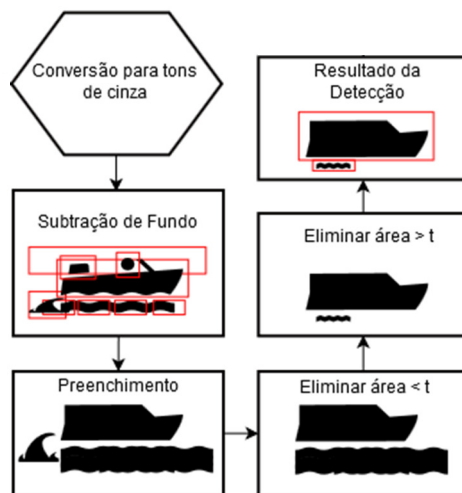
Fonte: O Autor, 2016.

Para facilitar o detalhamento de cada fase do método completo, divide-se a explicação de acordo com as duas etapas apresentadas em alto nível na Figura 14.

#### 4.2. DETECÇÃO INICIAL

A primeira parte do método que representa a detecção inicial, resultará em todos os pixels de objetos em movimento na cena de entrada considerados relevantes. A Figura 16 mostra as técnicas utilizadas na etapa de detecção inicial da embarcação na cena. O objetivo aqui é diminuir significativamente os falsos positivos.

Figura 16 - Etapa de detecção inicial: O vídeo de entrada é convertido para tons de cinza. Em seguida, os objetos em movimento são retirados da imagem por subtração de fundo e o preenchimento das regiões é utilizado para interligar partes de um mesmo objeto. Na sequência, dois limiares são utilizados para refinar os objetos encontrados, a fim de produzir um número menor de objetos a serem avaliados na próxima etapa.



Fonte: O Autor, 2016.

Considerando o fluxo da Figura 16, após a entrada dos dados (vídeo), a primeira etapa é a conversão dos frames para escala de cinza, a fim de reduzir o custo computacional do método uma vez que os dados resultantes dessa etapa serão representados por matrizes  $m \times n$ , ao invés de matrizes  $m \times n \times 3$ . Em seguida, utiliza-se um algoritmo de subtração de fundo para a retirada dos pixels em movimento da cena. Neste trabalho o algoritmo utilizado é a subtração de fundo com modelo do fundo baseado em distribuição de mistura gaussiana (JING e CHAOJIAN, 2011). Essa etapa permitirá a detecção dos objetos em movimento. O objetivo é que os objetos em movimentos sejam realçados. A nossa expectativa é que a água não seja detectada frequentemente, pois, mesmo que esteja em movimento, o seu movimento natural em um ambiente sem tempestade não é tão significativo, ocasionando a detecção de pouquíssimos pixels de água na cena. Dessa forma, serão detectados sempre os pixels que se diferem da coloração da água e com um movimento significativo, isto é, barco, planta, espuma, etc.

A saída produzida pela fase de subtração do fundo é uma máscara binária representando os objetos em movimento, informação que pode gerar inúmeros ruídos da cena, ou seja, falsos positivos detectados, tais como: água com reflexo do sol, espuma deixada por embarcações dentro ou fora da cena, etc. Para melhorar os resultados dessa detecção inicial são aplicadas técnicas de

preenchimento e eliminação de áreas. A técnica de preenchimento, vista na seção 2.4.3.5 e mostrada na Equação 7, é usada para completar uma determinada área vazada que possa existir na detecção, o que facilitará a tarefa de encontrar a quantidade de pixels nessa área. Utiliza-se essa técnica com base no trabalho de Shi et. al. (2013) e Liu et. al. (2014).

Em seguida, os falsos positivos são eliminados com base em um limiar máximo  $t_{max}$  e em um limiar mínimo  $t_{min}$  para as áreas dos objetos detectados que representam regiões de conjuntos conexos. Os limiares foram definidos, respectivamente, a partir da maior e da menor área de embarcação encontrada na base de dados completa, sendo respectivamente neste trabalho de 20000 e 900 pixels. Assim, quando os componentes conectados tiverem uma área maior que  $t_{max}$  e uma área menor que  $t_{min}$  serão eliminados. A utilização desse processo de limiarização é baseada no trabalho de Marques et al. (2014), a fim produzir uma quantidade menor de objetos de interesse para a etapa de reconhecimento de padrões utilizada na fase da detecção final.

#### 4.3. DETECÇÃO FINAL

A etapa de reconhecimento de padrões por aprendizagem de máquina utiliza as imagens das máscaras de embarcações geradas na etapa anterior e inicialmente extrai características das imagens. Dentre as diversas características encontradas na literatura, neste trabalho foram extraídas as seguintes características:

- Razão circular: razão entre a área de uma região e a área de um círculo contendo o mesmo perímetro. A utilização dessa característica foi baseada no trabalho de Shi et.al. (2013) e seu cálculo é mostrado na Equação 7.
- Excentricidade: medida do desvio em relação a uma circunferência que determina uma forma cônica, a qual pode ser calculada pela razão entre o centro da cônica pela metade do maior eixo.
- Razão entre altura e largura: essas características é inspirada no trabalho de Liu et. al. (2014), assumindo que esta razão gera uma proporção áurea para embarcações.

- Momentos Invariantes: médias ponderadas de pixels para descrever uma imagem independentemente de rotação, translação ou tamanho.

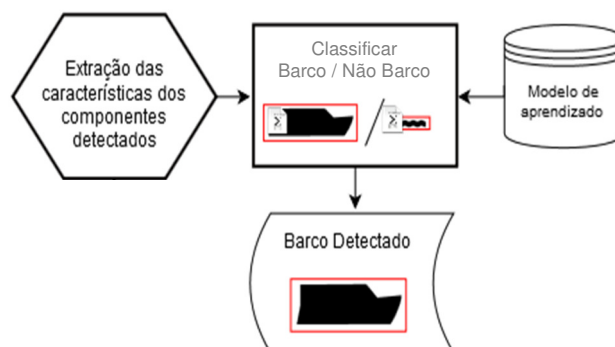
Essas características foram selecionadas por serem medidas invariantes à rotação, no intuito de melhorar o processo de reconhecimento, do método proposto.

Foram feitas combinações entre as mesmas para verificar o índice de ganho e concluiu-se que com utilização de todas elas consegue-se o maior índice de ganho.

Após a fase de extração de características, um modelo gerado por um algoritmo de aprendizagem de máquina devidamente treinado é utilizado para classificar a imagem (máscara) como barco ou não barco.

A etapa final desta fase é a confirmação do que é embarcação (detecção final) na cena. Para isso utiliza-se os objetos detectados na etapa anterior (detecção inicial) para serem classificados pelo modelo criado pelo algoritmo de aprendizagem de máquina. O classificador deverá marcar apenas os objetos considerados embarcação. A nossa hipótese é que essa segunda fase reduzirá os falsos positivos restantes, melhorando a precisão do método, como mostrado na Figura 17.

Figura 17 - Etapa de detecção Final (Confirmação). Primeiramente é realizada a extração de características dos objetos detectados na fase anterior. Em seguida, o algoritmo de classificação é executado. Este marcará apenas os objetos classificados como embarcação na cena.



Fonte: O Autor, 2016.

É importante destacar que o resultado final do método são marcações na forma de caixas delimitadoras nos objetos considerados embarcações na cena.

A fim de demonstrar a eficácia do método mostrar-se no capítulo seguinte os experimentos realizados, assim como os resultados obtidos para comprovar a capacidade de detecção da nossa proposta.



## 5. EXPERIMENTOS E RESULTADOS

Neste capítulo são descritos os experimentos realizados para ilustrar a aplicabilidade do método proposto, bem como os resultados. Inicialmente será apresentado o protocolo experimental, onde define-se como será feita a análise dos experimentos, na seção 5.1. Em seguida, a seção 5.2 apresenta detalhes sobre a base de dados utilizada nos experimentos. Por último, são apresentados os resultados obtidos para discussão considerando as métricas de avaliação investigadas.

É importante destacar que todos os experimentos realizados foram implementados e executados no MATLAB R2014a, em um sistema operacional Windows (versão 7 Pro) e um computador Intel (R) Core (TM) i5 - 2520 CPU de 2.50GHz e 2 CPU cores, com 10 GB de memória RAM.

### 5.1. PROTOCOLO EXPERIMENTAL

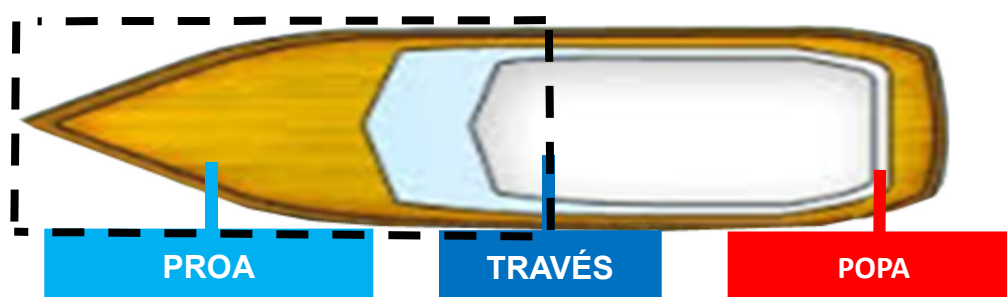
Considerando o método proposto neste trabalho e o uso de uma base de dados composta por vídeos nos experimentos realizados, cada vídeo é analisado em quadros/imagens. Logo, o método de detecção de embarcação é aplicado a cada imagem, sendo que a saída produzida são as caixas delimitadoras (*Bounding Boxes*) preditas, ou seja, todo objeto predito como embarcação é exibido no vídeo por meio de *Bounding Boxes*. No decorrer do texto, utiliza-se o termo quadros dos vídeos de imagens para facilitar o entendimento.

A avaliação dos resultados é feita em duas etapas. Na primeira etapa ocorre a análise da existência de embarcação em cada imagem. Nesse caso, para cada imagem/quadro de um vídeo retorna-se um valor verdadeiro, caso exista uma embarcação na cena, ou falso, caso não exista a embarcação. A segunda avaliação busca medir a qualidade de detecção na imagem, ou seja, a precisão da detecção por imagem.

Para avaliar a qualidade de detecção define-se a seguinte regra: para que a embarcação detectada seja considerada correta ou verdadeira positiva (VP), a imagem deve conter pelo menos a parte da embarcação correspondente à região da proa até o través, como mostrado na Figura 18, pois, assume-se que a

embarcação sempre entrará na cena a partir da proa e esta permanecerá até que a embarcação comece a sair de cena, também a partir da proa. Além disso, é importante destacar que este trabalho é componente de um projeto mais amplo que busca a classificação de embarcações. Como consequência, a imagem deve conter um percentual significativo do objeto para permitir a tarefa de identificação. Portanto, essa regra faz com que verdadeiros positivos se tornem falsos positivos (FP) caso não sigam a regra definida, isto é, composição de proa e través. O falso negativo (FN) representa a imagem que deveria ter uma caixa delimitadora, porém, não foi predita como tal.

Figura 18 – Identificação do corpo da embarcação



Fonte: O Autor, 2016.

A métricas definidas para avaliar os resultados alcançados são: precisão, revocação e medida F, na primeira fase de experimentos. Também utilizar-se a acurácia para obter uma visão geral de acertos do método, e do valor negativo predito, para avaliar quanto o método é preciso ao predizer imagens sem embarcações.

Na segunda fase de experimentos, analisa-se cada imagem que possui embarcação e possui uma caixa delimitadora predita, utiliza-se as regras adaptadas de EVERINGHAM et. al. (2010). Nesse contexto, quando a área de sobreposição ( $a_s$ ) entre a caixa delimitadora predita ( $B_p$ ) e a caixa delimitadora anotada ( $B_{gt}$ ) no conjunto de dados for maior que 90%, essa predição será considerada correta ou VP, caso contrário, será considerado um FP. A área  $a_s$  é definida como:

$$a_s = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_{gt})}. \quad (17)$$

Adicionalmente, o FN será denotado quando não existir uma caixa delimitadora em uma imagem que pela anotação deveria conter uma caixa. Assim, as métricas definidas para avaliar os resultados desta etapa são precisão e revocação. Para complementar utiliza-se medida F, que é a média harmônica entre precisão e revocação. Com isso, espera-se que os resultados alcancem valores elevados para essas métricas, ou seja, valores próximos de 1 em uma faixa de 0 a 1.

Na seção seguinte mostrar-se as especificações da base de dados e as delimitações necessárias para que sejam executados os experimentos e avaliação quantitativa dos resultados adquiridos.

## 5.2. BASE DE DADOS E DELIMITAÇÃO DOS EXPERIMENTOS

Neste trabalho foi utilizada uma base de dados própria composta por vídeos filmados pelo próprio autor por uma câmera com lente grande angular (olho de peixe) com *field of view* de aproximadamente 21.9 mm, tamanho de resolução de 1920x1080 pixels e 30 quadros por segundo, que será disponibilizada na internet. Foram utilizadas 18 sequências de vídeos adquiridas por um Veículo Aéreo Não Tripulado (VANT), sendo este um quadricóptero com módulo GPS e sistema de piloto automático, voando sobre o rio Tapajós, à margem esquerda, em frente ao porto de travessia da cidade de Itaituba no estado do Pará, com uma altitude de 30 metros e equipado com uma câmera colorida com lente grande angular. Em todas as sequências o VANT ficou em um ponto fixo no ar e filmou embarcações em movimento da entrada até sua saída na cena. As embarcações registradas na base de dados variam entre aproximadamente 6 metros a 12 metros de comprimento.

Os vídeos adquiridos foram convertidos em arquivos codificados no formato de MPEG4 com uma taxa de 29 quadros por segundo e resolução espacial de 640x480 pixels. Além disso, os vídeos foram filmados com uma câmera com lente posicionada em um ângulo reto direcionada para baixo.

Após a aquisição dos vídeos, estes foram rotulados para permitir a avaliação dos resultados obtidos tanto pelo método proposto quanto pelo

*baseline*. Os rótulos foram gerados a partir de marcações produzidas por 10 voluntários. Conforme informado na seção 5.1, para ser considerada embarcação, a imagem deverá conter a região da proa até o través do barco. Com base nessa orientação, os voluntários marcaram a imagem que consideraram atingir a métrica determinada. Assim, em cada vídeo, todas as imagens anteriores às marcações dos usuários foram rotuladas como não embarcação. Os usuários também marcaram o momento da saída do barco da cena, o qual foi definido como o instante em que a proa começa a sair de cena. Esse momento foi considerado o fim de cada vídeo.

Como 10 voluntários participaram das marcações, houve variações no resultado produzido por cada pessoa. Para adquirir apenas um valor de referência em cada sequência de vídeo é utilizada a mediana dos valores das posições das imagens marcadas pelos voluntários, tanto para a entrada como para a saída da embarcação em cena. Esses valores são apresentados na Tabela 2.

Tabela 2 – Mediana das marcações dos vídeos obtidas a partir das marcações individuais feitas pelos voluntários.

<b>Num. Víde o</b>	<b>Marcação de Entrada</b>	<b>Marcação de Saída</b>
1	84	273
2	102	270
3	1	54
4	99	213
5	96	507
6	81	219
7	63	270
8	135	547
9	45	195
10	45	180
11	60	174
12	75	165
13	96	219
14	123	417
15	72	174
16	90	204
17	48	141
18	105	354

Fonte: O Autor, 2016.;

Por uma questão de adaptação para a comparação do método proposto com o *baseline* (Marques et. al., 2014), utiliza-se uma abordagem baseada em blocos. Dessa forma, os resultados mostrados neste capítulo correspondem à predição feita de 3 em 3 quadros da sequência de cada vídeo. O total de imagens em cada sequência de vídeo a ser analisada, assim como a quantidade de imagens com e sem embarcações são mostrados na tabela 3.

Tabela 3 - Quantidade de imagens total, com e sem embarcação

<b>Núm. Vídeo</b>	<b>Total de imagens</b>	<b>Quantidade de imagens com Embarcação</b>	<b>Quantidade de imagens sem Embarcação</b>
1	92	64	28
2	91	57	34
3	18	18	0
4	72	39	33
5	170	138	32
6	74	47	27
7	91	70	21
8	183	138	45
9	66	51	15
10	61	46	15
11	59	39	20
12	56	31	25
13	74	42	32
14	140	99	41
15	59	35	24
16	69	39	30
17	48	32	16
18	119	84	35
<b>Total</b>	<b>1542</b>	<b>1069</b>	<b>473</b>

Fonte: O Autor, 2016.

Na próxima seção mostra-se os resultados obtidos com base nas métricas de avaliação investigadas para que ocorra a avaliação da eficiência do método proposto neste trabalho.

### 5.3. RESULTADOS DOS EXPERIMENTOS

Conforme mencionado na Seção 5.1, os experimentos realizados são divididos em dois grupos: (1) avaliação da precisão na detecção de embarcações na cena; e (2) análise da qualidade da marcação realizada.

A primeira série de experimentos foi realizada em relação à detecção da presença ou não de embarcações nas imagens. Para esse fim, considera-se, como informado na seção 5.2, a marcação das imagens feitas pelos voluntários. É importante destacar que a avaliação de todos os vídeos iniciou a partir da primeira imagem 1, porém, só são apresentados resultados na tela a cada três imagens analisadas, conforme também descrito anteriormente. Sendo assim, a quantidade de imagens avaliadas é mostrada na tabela 3. Às imagens que não continham embarcação, aquelas exibidas antes da marcação dos usuários, foi atribuído valor falso ou 0 (zero), enquanto que às imagens que continham embarcação, aquelas no intervalo fechado da marcação, foi atribuído valor verdadeiro ou 1 (um).

A segunda série de experimentos foi feita para aferir a qualidade da marcação realizada pelo método proposto e pelo *baseline*. Aqui, as caixas delimitadoras produzidas pelos dois métodos foram analisadas para evitar alguns erros na avaliação da primeira fase de teste, ou seja, é possível que na primeira série de experimentos os métodos marquem algum objeto na cena que não representa um barco, mas que será considerado uma marcação correta caso esteja em um intervalo rotulado como correto.

Contudo, antes da execução das duas fases de experimentos, realizou-se uma fase de seleção das melhores características da imagem e do melhor algoritmo de aprendizagem de máquina utilizado para a classificação. Essa etapa, descrita na seção a seguir, foi necessária para permitir a escolha desses dois parâmetros do método proposto.

### **5.3.1. Ajuste de parâmetros**

Em termos de características, foram utilizadas características dos trabalhos relacionados, tais como HOG, componentes conectados e valores geométricos. Quanto aos algoritmos de classificação, também foram testados algoritmos citados nos trabalhos relacionados: SVM, *k-NN* e *Random Forest*.

Dentre as características investigadas, as características geométricas produziram os melhores resultados. O método HOG gerou resultados inferiores. A razão para esse desempenho de HOG aquém do esperado é a seguinte: nossa

base de dados possui uma diferença crucial em relação às bases que utilizaram HOG na literatura, as quais foram geradas com câmeras fixas em 90 graus direcionadas para frente, enquanto que a nossa base foi filmada com uma câmera fixa em 90 graus direcionada para baixo. Além disso, esse tipo de característica é invariante à rotação, fator que dificulta a identificação de orientações relevantes para serem utilizadas.

Diante desses resultados, utiliza-se nas duas séries de experimentos características geométricas, uma vez que as mesmas são variantes à rotação. As características usadas são: valores referentes à razão circular (SHI et al, 2013), razão entre largura e altura (LIU et al, 2014) e, adicionalmente, os valores de excentricidade e momentos invariantes da imagem.

Com as características selecionadas e extraídas dos vídeos, avança-se para a seleção do classificador. O método SVM foi testado com diferentes configurações: kernels linear, polinomial e sigmoide. Também foram testados os algoritmos de classificação K-NN e Random Forest. Os testes foram feitos por meio de validação cruzada de dez rodadas, os quais alcançaram os valores de precisão, revocação e acurácia apresentados na Tabela 4.

Tabela 4 - Resultado dos algoritmos de classificação para a detecção de embarcação com as imagens da base de dados

<b>Classificadores</b>	<b>PR</b>	<b>RC</b>	<b>F</b>	<b>AC</b>
<b>SVM (Linear)</b>	<b>0,239</b>	<b>0,669</b>	<b>0,352</b>	<b>0,805</b>
<b>SVM (Polinomial)</b>	<b>0,271</b>	<b>0,675</b>	<b>0,387</b>	<b>0,81</b>
<b>SVM (Radial)</b>	<b>0,246</b>	<b>0,676</b>	<b>0,361</b>	<b>0,807</b>
<b>SVM (Sigmoide)</b>	<b>0,132</b>	<b>0,136</b>	<b>0,134</b>	<b>0,622</b>
<b>KNN (k=3)</b>	<b>0,827</b>	<b>0,785</b>	<b>0,805</b>	<b>0,911</b>
<b>KNN (k=5)</b>	<b>0,808</b>	<b>0,784</b>	<b>0,796</b>	<b>0,908</b>
<b>Random Forest (20 árvores)</b>	<b>0,764</b>	<b>0,846</b>	<b>0,803</b>	<b>0,917</b>
<b>Random Forest (50 árvores)</b>	<b>0,764</b>	<b>0,862</b>	<b>0,810</b>	<b>0,920</b>
<b>Random Forest (100 árvores)</b>	<b>0,764</b>	<b>0,871</b>	<b>0,812</b>	<b>0,925</b>

Fonte: O Autor, 2016.

Esses resultados mostraram que o método *Random Forest* obteve desempenho superior aos demais algoritmos. Com isso, *Random Forest* foi selecionado para compor o método de detecção.

Após os testes preliminares para escolha das características e do algoritmo de classificação, avança-se para a etapa de testes finais e comparação com o trabalho relacionado mais próximo ao método proposto, Marques et al.

(2014). Esse trabalho foi escolhido como *baseline* por ser um dos únicos trabalhos relacionados que utiliza imagens aéreas para compor a sua base de dados e com a câmera apontada para baixo próximo a 90 graus. Marques et. al. (2014) utilizam uma detecção inicial através de um limiar na intensidade de pixel. Em seguida, o método realiza uma validação espacial eliminando tudo o que tocar nas bordas, por meio de um limiar de área máxima e um limiar de distância entre os componentes conectados existentes na detecção inicial. Por fim, é realizada uma validação temporal com um limiar de permanência de um componente conectado na cena.

Para tornar a comparação mais justa nas duas séries de experimentos, foram feitas duas comparações com o *baseline*. Na primeira comparação, o *baseline* foi modificado para ajustá-lo melhor à nossa base, retirando a etapa de exclusão dos elementos que tocam na borda da imagem. O segundo processo de comparação utiliza a versão original do *baseline*. É importante também destacar que os resultados mostrados foram calculados considerando o total de imagens obtidas nas dezoito sequências de vídeos da base de dados.

### 5.3.2. Primeira série de experimentos

Nesta primeira série de experimentos, o método proposto foi comparado ao *baseline* em termos de desempenho de detecção. Cada vídeo foi avaliado em blocos de 03 *frames* quanto à existência ou não de embarcações na cena. Os resultados obtidos são mostrados na Tabela 5 por meio dos valores de precisão, revocação, valor negativo predito, medida F e acurácia.

Tabela 5 - Resultados obtidos na primeira série de experimentos ao comparar o método proposto com o *baseline* nas versões otimizada e original.

Métodos	PR	RC	VNP	F	AC
<b>BL. Otimizado</b>	<b>0,730</b>	<b>0,930</b>	<b>0,580</b>	<b>0,820</b>	<b>0,710</b>
<b>BL. Original</b>	<b>0,900</b>	<b>0,510</b>	<b>0,440</b>	<b>0,650</b>	<b>0,620</b>
<b>Método Proposto</b>	<b>0,910</b>	<b>0,760</b>	<b>0,610</b>	<b>0,830</b>	<b>0,790</b>

Fonte: O Autor, 2016.

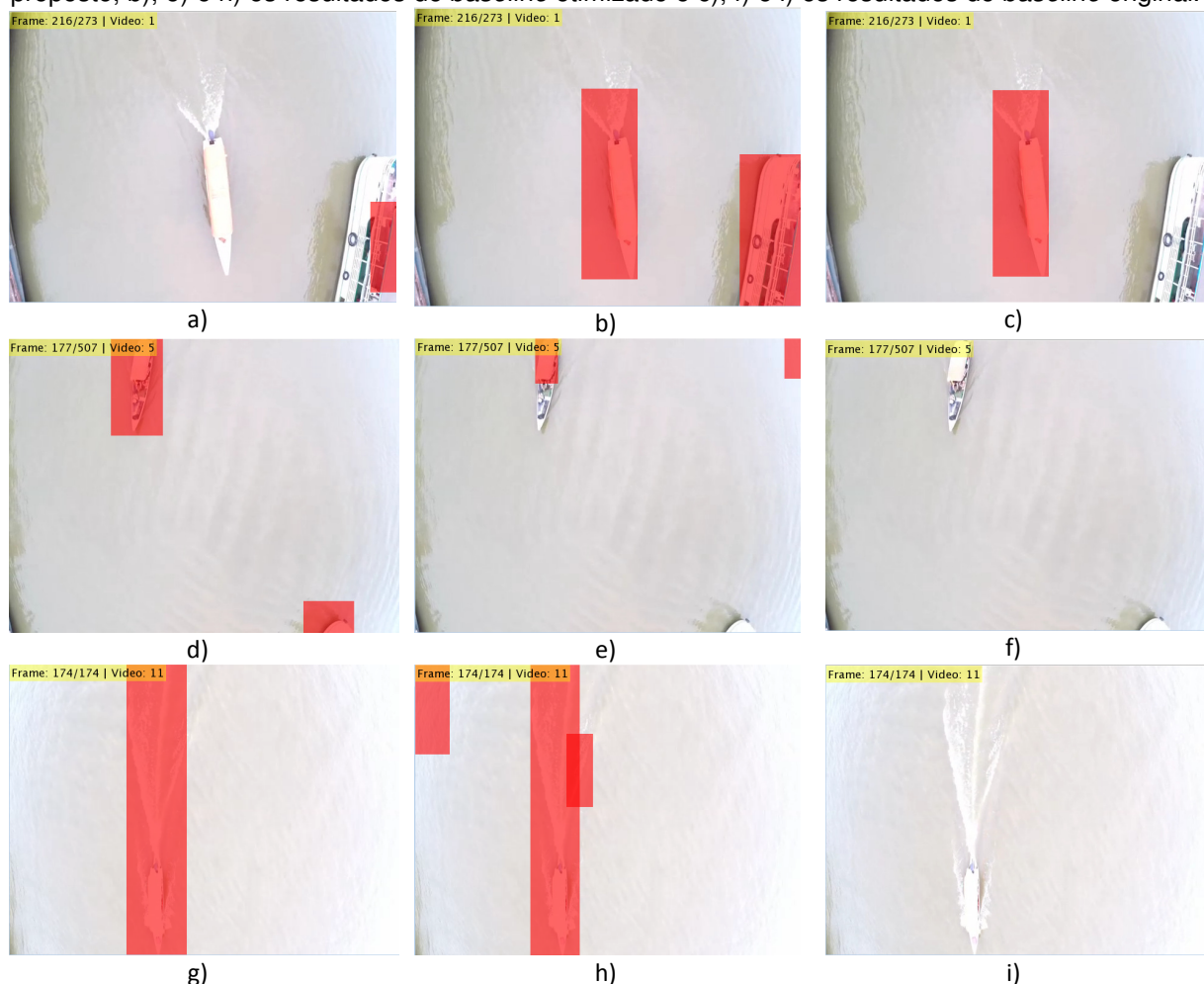
Dos resultados obtidos, observa-se que o método proposto obteve a melhor taxa de acurácia com 79% de acerto considerando todas imagens e teve



menos detecção de objetos em cada imagem. Além disso, o método proposto alcançou maior precisão, 91%, fato que mostra que o método proposto foi mais preciso ao marcar embarcações na cena.

Contudo, constatou-se que em algumas imagens que possuíam embarcações, a caixa delimitadora resultante do processo de detecção não representava realmente uma embarcação, como mostrado na Figura 19 na primeira linha e primeira coluna. Nesse caso, embora houvesse um barco na imagem, como este não foi detectado, essa marcação deveria ser avaliada como um falso positivo.

Figura 19 – Exemplos de marcações da embarcação nos quadros. A primeira linha representa o quadro 216 do vídeo 1, a segunda linha representa o quadro 177 do vídeo 5 e a última linha representa o quadro 174 do vídeo 11, onde a), d) e g), representam os resultados do método proposto, b), e) e h) os resultados do *baseline* otimizado e c), f) e i) os resultados do *baseline* original.



Fonte: O Autor, 2016.

Na Figura 19, seguindo o fluxo da esquerda para a direita e de cima para baixo, as imagens representam: a) a marcação incorreta da embarcação em

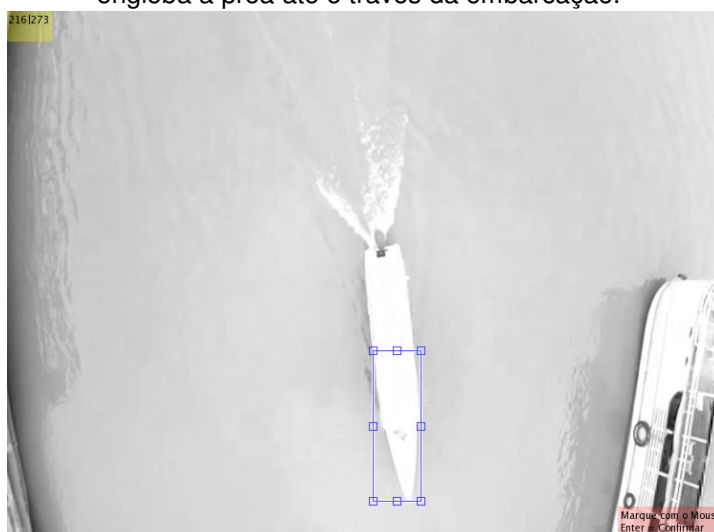
movimento, b) a marcação correta do barco em movimento mais a marcação de um falso positivo de uma embarcação que não era para ser reconhecida como embarcação, segundo o protocolo experimental, c) a marcação apenas da embarcação correta; d) idem ao b), e) a marcação de apenas parte da embarcação correta que por não seguir o protocolo experimental é considerada incorreta e uma marcação de uma campo que não é uma embarcação; f) o método deixou de marcar a embarcação, g) idem ao c), h) a marcação da embarcação correta mais duas marcações incorretas e i) idem ao f).

Diante dos resultados obtidos nesta primeira série de experimentos, tornou-se necessária a avaliação da qualidade da marcação realizada pelos métodos de detecção. Desta forma, na segunda etapa de testes as marcações foram analisadas de acordo com os critérios definidos anteriormente, isto é, para ser considerada uma embarcação, a imagem deverá mostrar a região da proa até o través da embarcação marcada. Esse passo é importante para facilitar uma futura etapa de classificação das embarcações.

### **5.3.3. Segunda série de experimentos**

A avaliação desta segunda série de experimentos foi feita com base em um *ground truth*, gerado a partir de marcações apresentadas em cada imagem dos vídeos. As imagens foram marcadas manualmente em cada frame com base na menor largura em que a embarcação se encaixasse e no seu comprimento (proa ao través) do início do vídeo até o final da cena, conforme mostrado na Figura 20. Apenas as imagens que continham embarcações foram usadas na geração do *ground truth*, compreendendo os intervalos fechados das imagens marcadas na sequência de vídeo, como apresentados na Tabela 2.

Figura 20 - Marcação do *Ground Truth* (GT). O GT compreende o menor retângulo que engloba a proa até o través da embarcação.



Fonte: O Autor, 2016.

Com esta segunda bateria de testes conseguiu-se também atingir resultados representativos para confirmação da eficácia do método. Tais resultados são mostrados na Tabela 6 em comparação com o método de Marques et al, 2014.

Tabela 6 - Resultados obtidos na segunda série de experimentos comparando método proposto com o *baseline* nas versões otimizada e original

Métodos	PR	RC	F
<b>BL. Otimizado</b>	<b>0,290</b>	<b>0,950</b>	<b>0,440</b>
<b>BL. Original</b>	<b>0,280</b>	<b>0,350</b>	<b>0,310</b>
<b>Método Proposto</b>	<b>0,700</b>	<b>0,720</b>	<b>0,710</b>

Fonte: O Autor, 2016.

Pode-se afirmar que, além de apresentar melhor acurácia e precisão, assim como ocorreu na primeira bateria de testes, esta segunda série de experimentos mostrou que o método proposto também produziu melhores resultados na qualidade das caixas delimitadoras (marcações). Pode ser observado por meio das taxas de precisão das marcações e da medida F, que o método proposto é o melhor dentre os métodos comparados.

Embora o método de Marques et al (2014) tenha alcançado uma taxa de revocação mais elevada, esse resultado não significa que o *baseline* seleciona a área correta com as caixas delimitadoras, ou seja, o método detecta a existência de um barco em cena, mas não consegue determinar a região correta no barco na cena. Na Figura 19 (e), por exemplo, observa-se que as marcações,

embora estejam na embarcação, não compreendem exatamente a região que deveria ser considerada como embarcação e a outra marcação mostrada compreende alguma saturação na água, que mesmo não sendo um barco, fez com que o método a interpretasse como barco.

O *baseline* não utiliza técnicas de aprendizagem de máquina, dificultando o refinamento dos falsos positivos, uma vez que sua abordagem é para o mar aberto, onde não se encontram embarcações ancoradas, vegetação flutuante, entre inúmeros outros objetos que poderão ser retornados como barcos sem a etapa de classificação.

Por fim, na próxima seção irá apresentar-se uma pequena discussão a respeito dos resultados obtidos nos experimentos.

#### 5.4. DISCUSSÃO

As duas séries de experimentos mostraram que os resultados obtidos pelo método proposto superaram o *baseline*. Além disso, em termos práticos, a taxa de precisão alcançada pelo método proposto (91%) é bastante elevada, fato que mostra sua viabilidade de uso em uma aplicação real. Por fim, os resultados do segundo grupo de experimentos mostraram que o método proposto marcou corretamente a região de interesse da embarcação, com taxa de 70% de precisão. Esse resultado indica que elevadas taxas de acerto poderão ser obtidas em uma posterior fase de classificação de embarcações.

Para finalizar, no próximo capítulo irá explanar-se nossas conclusões sobre o método proposto e os resultados apresentados.

## 6. CONSIDERAÇÕES FINAIS

Neste capítulo analisa-se os resultados gerais alcançados extraindo algumas conclusões, além de resumir as principais contribuições obtidas, mostra algumas futuras pesquisas que podem estender esta dissertação.

### 6.1. CONSIDERAÇÕES

A tarefa de detecção de embarcação, ou de objetos de forma geral, é uma área atualmente muito estudada devido sua diversidade de aplicações. Esse problema ainda é de difícil resolução e, normalmente, há necessidade de impor condições e restrições de escopo para que seja possível superar algumas dificuldades específicas de cada aplicação.

O objetivo desta dissertação foi desenvolver e validar um método para detecção automática de embarcação por meio da identificação de movimento em imagens e classificação no contexto de rios amazônicos, a fim de minimizar o trabalho efetuado por operadores humanos nas análises de sequências de vídeos de áreas de proteção nos rios do ambiente amazônico. A motivação deste trabalho foi um projeto mais amplo que prevê a instalação de balões ancorados, ou Aeróstatos, os quais serão utilizados como retransmissores para enlaces de comunicação, assim como ferramenta importante para o apoio nas pesquisas de preservação e monitoramento ambiental, nas reservas de Mamirauá e Amanã no Estado do Amazonas. Uma das tarefas de monitoramento prevê a detecção de barcos a partir de vídeos capturados por câmeras instaladas na parte mais alta do balão. Assume-se que as câmeras terão suas lentes direcionadas para baixo.

Para alcançar o objetivo proposto, um desafio considerável foi travado, posto que um sistema computacional para reconhecimento de padrões em imagens/vídeos engloba diversas etapas normalmente dependentes da qualidade de sua etapa anterior e resolvidas a partir da união de conhecimentos interdisciplinares advindos do processamento de imagens, aprendizagem de máquina, e outras áreas correlatas.

A revisão de literatura realizada neste trabalho permitiu destacar que as bases utilizadas nos trabalhos investigados eram privadas e criadas pelos próprios autores ou por órgãos governamentais, tornando-as indisponíveis para testes com o método proposto. Devido à essas limitações, foi criada uma base própria para ser investigada neste trabalho, a qual possui marcação de um *ground truth* com base nas restrições impostas no trabalho.

Com a criação da base constatou-se que existia alguns fatores de dificuldades inerentes ao processamento dos quadros da sequência do vídeo como rotação do objeto a ser detectado (as embarcações), a movimentação da câmera, a não uniformidade da iluminação e principalmente o aparecimento de espuma com a aceleração da embarcação. Contudo, as características extraídas das imagens, bem como as duas fases que compõem o método, precisamente fase de detecção e fase de confirmação, mostraram-se bastante adequados e permitiram superar as dificuldades inerentes ao ambiente investigado.

As taxas de precisão obtidas pelo método proposto na tarefa de detecção, assim como na avaliação da qualidade das detecções, indicam a viabilidade de uso do método proposto em aplicação real. É importante também destacar que a qualidade adequada das detecções possibilita o desenvolvimento futuro de uma etapa de classificação de embarcações, a qual terá como objetivo o reconhecimento de embarcações suspeitas na área monitorada.

Os experimentos também envolveram a comparação do método proposto ao método de Marques et al (2014), tanto em sua versão otimizada neste trabalho, quanto em sua versão original. Os resultados obtidos por meio dessas comparações mostram que o método proposto supera o *baseline* na tarefa de detecção, e produz marcações significativamente superior ao *baseline*. Esses resultados mostram que a fase de confirmação de detecção, realizada via algoritmo de aprendizagem de máquina, é determinante para o sucesso da tarefa de detecção de embarcações no ambiente amazônico investigada neste trabalho.

## 6.2. LIMITAÇÕES DO MÉTODO

Mesmo o sistema mostrando-se eficiente, possui restrições importantes, são elas:

- Não trata ruídos de ambientes de neblina ou chuva.
- O problema de oclusão não foi investigado.
- Não foram investigados problemas em ambientes noturnos.
- O problema com embarcações paradas já presentes na cena também não foi tratado.

## 6.3. CONTRIBUIÇÕES

Pode-se ressaltar que como resultado desta dissertação as seguintes contribuições:

- Descrição dos principais trabalhos relacionados referentes a métodos de detecção de embarcações em imagens e vídeos.
- Experimentos com uma combinação de características relevantes para o processo de detecção das embarcações.
- Aplicação de um método de aprendizagem de máquina, o melhor foi *Random Forest*, para melhorar a delimitação correta da embarcação.
- Criação da base de vídeos de embarcações no ambiente amazônico com as marcações de *ground truth*.

## 6.4. TRABALHOS FUTUROS

Como o resultado final e o desempenho de forma geral de um sistema de classificação podem ser modificados seja pela melhora dos resultados da fase intermediária ou pela escolha das características que definem o modelo de comparação em relação aos dados de entradas, consegue-se explorar alguns pontos, abaixo elencados, para melhorar os resultados do trabalho no futuro.

- Melhorar o processo de segmentação da imagem por meio de fusão das informações de subtração de fundo com um algoritmo de limiarização de imagem.

- Adicionar informação temporal com análise de quadros deslizantes para confirmar a presença de um objeto (embarcação) na cena.
- Adicionar características de profundidade para separar a espuma da embarcação.
- Adicionar informação de cor para extração de fundo.
- Classificar a embarcação detectada de acordo com seu tamanho.
- Estimar velocidade da embarcação em cena.



## REFERÊNCIAS

- AMARASINGHE, Sanjaya. et al. “**Location Estimation in a Maritime Environment Using a Monocular Camera**”. International Conference on Advances in ICT for Emerging Regions. Pag. 21-28, 2014.
- ANA - Agência Nacional de Águas. “**Caderno de Recursos Hídricos – Navegação Interior e sua Interface com o Setor de Recursos Hídricos**”. Brasília, ANA, 2005. Disponível em <[http://arquivos.ana.gov.br/planejamento/planos/pnrh/VF\\_Navegacao.pdf](http://arquivos.ana.gov.br/planejamento/planos/pnrh/VF_Navegacao.pdf)>. Acesso em 02/02/2015.
- ASHT, Seema; DASS, Rajeshwar. “**Pattern Recognition: A Review**”. International Journal of Computer Science and Telecommunications. Vol 3. No. 8. Pag 25-29. Aug 2012.
- ATTEIA, Ghada and COLLINS, Michael J. “**Ship Detection Performance Assessment for Simulated RCM SAR Data**”. International Geoscience and Remote Sensing Symposium (IGARSS), Pag. 553-556, 2014.
- BAO, Xinfeng. et al. “**Context Modeling Combined with Motion Analysis for Moving Ship Detection in Port Surveillance**”. Journal of Electronic Imaging. Vol. 22 Issue 4. Pag. 1-16. 2013.
- BECHAR, I. et al. “**Object Segmentation from a Dynamic Background Using a Pixelwise Rigidity Criterion and Application to Maritime Target Recognition**”. IEEE International Conference on Image Processing (ICIP). Pag. 363-367, 2014.
- BRASIL. Ministério da Educação. Secretária de Educação Básica. “**A importância do Mar na História do Brasil**”. Brasília, DF, 2006.
- CHAN, Michael T. and WEED, Christopher. “**Vessel Detection in Video with Dynamic Maritime Background**”. IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Pag. 1-5, 2012.
- CHELLAPPA, R. et al. “**Special Section on Distributed Camera Networks: Sensing, Processing, Communication, and Implementation**”. IEEE Transactions On Image Processing. Vol. 19. p. 2513-2515, 2010.
- CHEN, Hai-Wen; McGURR, Mike. “**An Automated Data Exploitation System for Airborne Sensors**”. Proceedings of SPIE – The International Society for Optical Engineering. Vol. 9089. 2014.
- CULLEN, Daniel; KONRAD, Janusz; and LITTLE, T.D.C. “**Detection and Summarization of Salient Events in Coastal Environments**”. Department of Electrical and Computer Engineering, Boston University. Boston, 2012.
- DUDA, R.; HART, P.; STORK, D. “**Pattern Classification**”. 2nd. Ed. A wiley-Interscience Publication. 2001.
- EVERINGHAM, M. et al. “**The pascal visual object classes (voc) challenge**”, International Journal of Computer Vision, 2010.
- FEFILATYEV, Sergiy. et al. “**Tracking Ships from Fast Moving Camera through Image Registration**”. International Conference on Pattern Recognition (ICPR). Pag. 3500-3503, 2010.
- FEFILATYEV, Sergiy and GOLDOF, Dmitry. “**Detection and tracking of marine vehicles in video**”. International Conference on Pattern Recognition (ICPR). Pag. 1-4, 2008.
- GONZALEZ, R. C. WOODS, R. E. “**Digital Image Processing**”. Prentice Hall. 3rd Edition, 2010.

- GUJRATHI, Poonam; PRIYA, Arokia; MALATHI, P. “**Detecting Moving Object Using Background Subtraction Algorithm in FPGA**”. International Conference on Advances in Computing and Communications. Pag. 117-120, 2014.
- HU, Weiming, et al. “**A Survey on Visual Surveillance of Object Motion and Behaviors**”. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews. VOL. 34, NO. 3, Pag. 334-352, 2004.
- HU, Wu-Chih. et al. “**Robust real-time ship detection and tracking for visual surveillance of cage aquaculture**”. Journal of Visual Communication and Image Representation. Vol. 22. Pag. 543-556, 2011.
- JING, Ge and CHAOJIAN, Shi. “**Survey on Methods of Moving Object Video Detection in Marine Environment**”. International Conference on Computer Science and Information Technology (ICCSIT), 2011.
- JOSHI, Kinjal A; THAKORE, Darshak G. “**A Survey on Moving Object Detection and Tracking in Video Surveillance System**”. International Journal of Soft Computing and Engineering (IJSCE), Vol.2, No.3, July 2012.
- KOPPARAPU, Sunil Kumar; SATISH, M. “**Identifying Optimal Gaussian Filter for Gaussian Noise Removal**”. National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics. Pag. 126-129. 2011.
- KOTSIANTIS, S. B. **Supervised machine learning: A review of classification techniques**. Informatics, vol. 31, no. 3, pp. 249-268. Slovene Society Informatika. 2007.
- LI, Zehui. *Et al.* “**A Method of Ship Detection in Optical Satellite Image Based on Saliency Map**”. ICTIS. pag. 2441-2449, 2013.
- LIU, Ge, et al. “**A New Method on Inshore Ship Detection in High-Resolution Satellite Images Using Shape and Context Information**”. IEEE Geoscience and Remote Sensing Letters. VOL. 11, No. 3. Pag. 617-621, 2014.
- LOOMANS, Marijn J.H.; Wijnhoven, Rob G.J.; With, Peter H.N. “**Robust Automatic Ship Tracking in Harbours Using Active Cameras**”. ICIP. Pag. 4117-4121. 2013.
- MAKANTASIS, Konstantinos. et al. “**Vision-based Maritime Surveillance System Using Fused Visual Attention Maps and Online Adaptable Tracker**”. International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 2013.
- MARQUES, Jorge S. et al. “**An Algorithm for the Detection of Vessels in Aerial Images**”. International Conference on Advanced Video and Signal Based Surveillance (AVSS). Pag. 295 – 300, 2014.
- MOREIRA, Rodrigo da Silva. et al., “**A Survey on Video Detection and Tracking of Maritime Vessels**”. International Journal of Recent Research and Applied Studies, Vol. 20, Issue 1, Julho, 2014.
- MOREIRA, Rodrigo da Silva. “**Um Sistema de Rastreamento de Veículos Marítimos Baseado na Rede Neural sem Peso WiSARD**”. 2015. 304f. Tese (Doutorado em Engenharia) - COPPE/ Programa de Engenharia Civil. UFRJ, Rio de Janeiro.
- PAREKH, Himani S.; THAKORE, Darshak; JALIYA, Udesang K. “**A survey on Object Detection and Tracking Methods**”. International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2. No. 2. Feb 2014.
- PIERALICE, Francesca. et al. “**An innovative methodological approach in the frame of Marine Strategy Framework Directive: A statistical model based**

- on ship detection SAR data for monitoring programmes**". Journal Marine Environmental Research. Elsevier. Vol. 102. Pag. 15-35. 2014.
- QIUYU, Zhu; YILONG, Jiang; BO, Chen. "**Design and Implementation of Video-Based Detection System for Wharf Ship**". International Conference on Smart and Sustainable City (ICSSC). Pag. 493-496. Aug 2013.
- SHI, Tingyan. et al. "**A Moving Ship Detection based on Edge Information of single image and background subtraction**". Proceedings of SPIE – Pattern Recognition and Computer Vision. Vol. 8919. Pag. 1-6, 2013.
- SOUSA, Beatriz F. s. et al. "**Avaliação de Classificadores Baseados em Aprendizado de Máquina para a Classificação do uso e Cobertura da Terra no Bioma Caatinga**". Revista Brasileira de Cartografia, Rio de Janeiro, N° 62, p. 385-399, 2010.
- THEODORIDIS, S.; KOUTROUMBAS, K. "**Pattern Recognition**". 3rd. Ed., Elsevier, 2006.
- WIJNHOVEN, Rob. et al. "**Online Learning for Ship Detection in Maritime Surveillance**". Symposium on Information Theory in the Benelux. Pag. 73-80, 2010.
- WITTEN, I. H. et al. "**Data Mining: practical learning tools and techniques**". - 3rd ed. Elsevier, 2011.