DANIEL TREVISAN BRAVO

IDENTIFICAÇÃO AUTOMÁTICA DE POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI A PARTIR DE IMAGENS AÉREAS ADQUIRIDAS POR VANTS

Tese apresentada ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento da UNINOVE como parte dos requisitos para obtenção do título de Doutor em Informática.

São Paulo

2019

UNIVERSIDADE NOVE DE JULHO – UNINOVE

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA E GESTÃO DO CONHECIMENTO

DANIEL TREVISAN BRAVO

IDENTIFICAÇÃO AUTOMÁTICA DE POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI A PARTIR DE IMAGENS AÉREAS ADQUIRIDAS POR VANTS

Tese apresentada ao Programa de Pós-Graduação em Informática e Gestão do Conhecimento da UNINOVE como parte dos requisitos para obtenção do título de Doutor em Informática.

Orientador: Prof. Dr. Sidnei Alves de Araújo

São Paulo

2019

Bravo, Daniel Trevisan.

Identificação automática de possíveis criadouros do mosquito Aedes aegypti a partir de imagens aéreas adquiridas por VANTs. / Daniel Trevisan Bravo. 2019.

150 f.

Tese (Doutorado) - Universidade Nove de Julho - UNINOVE, São Paulo, 2019.

Orientador (a): Prof. Dr. Sidnei Alves de Araújo.

1. Aedes aegypti. 2. Mapeamento automático. 3. VANT. 4. Drone. 5. Reconhecimento de padrões. 6. Visão computacional.

UNINOVE
Universidade Nove de Julho

PARECER – EXAME DE DEFESA

Parecer da Comissão Examinadora designada para o exame de defesa do Programa de Pós-Graduação em Informática e Gestão do Conhecimento, a qual se submeteu o aluno regularmente matriculado **Daniel Trevisan Bravo.**

Tendo examinado o trabalho apresentado para obtenção do título de "Doutor em Informática e Gestão do Conhecimento", com Tese intitulada "IDENTIFICAÇÃO AUTOMÁTICA DE POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI A PARTIR DE IMAGENS AÉREAS ADQUIRIDAS POR VANTS", a Comissão Examinadora considerou o trabalho:

( ) Aprovado
(X) Aprovado condicionalmente
( ) Reprovado com direito a novo exame
( ) Reprovado

Parecer: Aprovada, devendo acatar as recomendações da banca.

EXAMINADORES

Prof. Dr. Sidnei Alves de Araújo (Orientador – PPGI/UNINOVE)

Prof. Dr. Sergio Vicente Denser Pamboukian (Participante externo – PPGCAGE/MACKENZIE)

Prof. Dr. Cristiano Capellani Quaresma (Participante Externo – PPGCIS/UNINOVE)

Prof. Dr. Peterson Adriano Belan (Participante Interno – PPGI/UNINOVE)

Prof. Dr. Wonder Alexandre Luz Alves (Participante Interno – PPGI/UNINOVE)

São Paulo, 27 de novembro de 2019.

DEDICATÓRIA

Ao meu pai Samuel (in memorian)

e a minha mãe Cléa.

AGRADECIMENTOS

Primeiramente agradeço a Deus por ter me dado forças nos momentos mais difíceis. Esse período de estudos permeou várias fases boas e ruins da minha vida pessoal e profissional. Não foi nada fácil e, por isso, tenho certeza que Ele, juntamente com meu pai querido, me guiou em toda a trajetória.

Aos meus pais e, em especial, ao meu pai Samuel que sempre acreditou no meu potencial e sempre se orgulhou de mim até nos seus últimos dias de vida.

Aos meus familiares e amigos mais próximos por sempre me apoiarem.

A minha prima Vanessa pelos desabafos nos momentos críticos e pelos momentos de descontração.

Ao meu orientador Prof. Dr. Sidnei pela paciência e, sobretudo, pela compreensão em todos os momentos. Sou muito grato por todo seu esforço e dedicação para o desenvolvimento deste trabalho.

Aos amigos Charles e Stanley pela ajuda nos experimentos e pelo companheirismo.

Ao Prof. Dr. Wonder por ser sempre prestativo e por emprestar o VANT e acompanhar as aquisições das imagens.

Aos amigos que conquistei durante essa jornada e, em especial, à Dimitria e ao Nelson pelos conselhos e companheirismo.

Aos professores do PPGI por compartilharem seus conhecimentos.

Aos membros da banca, especialmente os membros externos, pelas valiosas contribuições na área de Geoprocessamento.

À Universidade Nove de Julho por ter concedido a bolsa de estudos.

A todos que, direta ou indiretamente, contribuíram para a construção e evolução deste trabalho.

RESUMO

O atual panorama de doenças causadas pelo mosquito *Aedes aegypti* no Brasil e no mundo tem motivado inúmeros esforços de pesquisa nas mais diversas áreas do conhecimento. Além das campanhas de prevenção no âmbito da saúde, a tecnologia mostra-se como uma grande aliada, a partir da utilização de veículos aéreos não tripulados (VANTs) para aquisição de imagens aéreas, facilitando o trabalho das equipes de vigilância sanitária. Contudo, tais imagens são normalmente analisadas de forma manual (visualmente), podendo demandar muito tempo dos agentes de saúde. Neste trabalho propõe-se uma abordagem de visão computacional para a identificação automática de objetos e cenários que representam potenciais criadouros do mosquito Aedes aegypti, a partir de imagens aéreas de regiões urbanas adquiridas por VANTs. A abordagem proposta contempla 4 etapas: composição de ortomosaicos, identificação de objetos e cenários suspeitos, detecção de pequenas porções d'água e geração de ortomosaicos anotados e relatórios. Para detecção de objetos e cenários suspeitos foram exploradas duas técnicas: redes neurais convolucionais – RNC e *Bag Of Visual Words – BoVW* combinada com o classificador *Support Vector Machine* – SVM (BoVW+SVM), sendo os resultados obtidos mensurados por meio da taxa *mean Average Precision* – mAP-50. Na detecção de objetos usando uma RNC modelo YOLOv3 obteve-se a taxa de 0,9610 para o mAP-50, enquanto na tarefa de detecção de cenários, na qual comparou-se os resultados da RNC tiny-YOLOv3 e de BoVW+SVM, foram obtidas as respectivas taxas de 0,9028 e 0,6453. Esses resultados sugerem que as RNCs são suficientes para identificação dos potenciais criadouros uma vez que juntas levaram à obtenção da taxa média de 0,9319 para o mAP-50. No que tange a detecção de pequenas porções d'água, nos experimentos conduzidos obteve-se o valor de 0,9757 para a medida de similaridade *Structural Similarity Index* – SSIM. Os resultados obtidos nos experimentos envolvendo as 4 etapas permitiram evidenciar que a abordagem proposta pode trazer contribuições significativas para a implementação de sistemas computacionais que visem auxiliar os agentes de saúde, no planejamento e execução de atividades de combate ao mosquito *Aedes aegypti* com o uso de VANTs.

Palavras-chave: *Aedes aegypti*, mapeamento automático, VANT, drone, reconhecimento de padrões, visão computacional, inteligência artificial.

ABSTRACT

The current panorama of diseases caused by the Aedes aegypti mosquito in Brazil and worldwide has motivated numerous research efforts in various areas of knowledge. In addition to health prevention campaigns, the technology proves to be a great ally, using unmanned aerial vehicles (UAVs) to acquire aerial images, facilitating the work of health surveillance teams. However, such images are usually analyzed manually (visually) and may require a lot of time from health agents. This work proposes a computer vision approach for the automatic identification of objects and scenarios that represent potential breeding sites of the Aedes aegypti mosquito, from aerial images of urban areas acquired by UAVs. The proposed approach includes 4 steps: composition of orthomosaics, identification of suspicious objects and scenarios, detection of small portions of water and generation of annotated orthomosaics and reports. To detect suspicious objects and scenarios, two techniques were explored: convolutional neural networks – RNC and Bag of Visual Words – BoVW combined with the Support Vector Machine classifier – SVM (BoVW + SVM), and the results obtained were measured using the mean Average Precision – mAP-50. In object detection using a YOLOv3 model RNC, we obtained the rate of 0.9610 for mAP-50, while in the scenario detection task, we compared the results of tiny-YOLOv3 RNC and BoVW + SVM, the respective rates of 0.9028 and 0.6453 were obtained. These results suggest that the RNCs are sufficient to identify potential breeding sites since together they led to the average rate of 0.9319 for mAP-50. Regarding the detection of small portions of water, the experiments conducted obtained the value of 0.9757 for the measure of similarity Structural Similarity Index – SSIM. The results obtained in the experiments involving the 4 steps showed that the proposed approach can make significant contributions to the implementation of computer systems aimed at assisting health agents in the planning and execution of activities to combat Aedes aegypti mosquito with the use of UAVs.

Keywords: Aedes aegypti, automatic mapping, UAV, drone, pattern recognition, computer vision, artificial intelligence.

Lista de Figuras

Figura 1: Exemplos de possíveis criadouros do mosquito *Aedes aegypti* 23
Figura 2: Etapas de um sistema de visão computacional..... 29
Figura 3: Exemplo para o cálculo do histograma de uma imagem colorida de 64 bins..... 33
Figura 4: Histogramas de 64 bins: (a) R, (b) G e (c) B 33
Figura 5: Ângulos utilizados para o cálculo das matrizes de coocorrência 35
Figura 6: Cálculo do valor de LBP para o pixel central..... 38
Figura 7: (a) imagem de entrada; (b) imagem LBP; (c) histograma da imagem 38
Figura 8: Funcionamento do HOG para obtenção do vetor de características de uma imagem 40
Figura 9: Histograma de gradientes bidimensional subdivido em seis intervalos 41
Figura 10: Etapas da técnica BoVW 42
Figura 11: Estratégia para a criação do dicionário de palavras visuais 43
Figura 12: Fase de codificação (coding) da técnica BoVW 44
Figura 13: Exemplo de classificação usando o SVM..... 47
Figura 14: (a) Conjunto de dados não linear; (b) Fronteira não linear no espaço de entradas; (c) Fronteira linear no espaço de características..... 48
Figura 15: Representações de algumas funções de ativação: (a) função passo, (b) função linear e (c) função sigmoidal. 52
Figura 16: Representações de uma camada da rede neural..... 53
Figura 17: Representação abreviada da rede com 3 camadas 54
Figura 18: Exemplos das operações de convolução e de subamostragem: (a) Resultado de uma convolução (direita) aplicada a uma imagem (esquerda); (b) Subamostragem com filtro de tamanho 2x2 e tamanho do passo igual a 2. 57
Figura 19: Exemplo de Arquitetura de uma RNC 58
Figura 20: Caixas delimitadoras com priorizações de dimensão e predição de localização . 60
Figura 21: Exemplo de múltiplas células classificadas como o mesmo objeto 62
Figura 22: Esquema empregado pelo YOLO 62
Figura 23: Arquitetura da RNC do YOLOv3 63
Figura 24: Gráfico comparativo do YOLOv3 com outros métodos utilizando a banco de imagens COCO 64
Figura 25: Exemplos de detecções em 3 escalas diferentes..... 65
Figura 26: Arquitetura da RNC da tiny-YOLOv3..... 66
Figura 27: Comportamento do cruzamento..... 68

Figura 28: Exemplo de mutação 68
Figura 29: Algoritmo Genético típico 69
Figura 30: Sistema de Sensoriamento Remoto 70
Figura 31: Exemplos de imagens multiespectrais no espectro visível e infravermelho 71
Figura 32: Comportamento espectral do solo, vegetação e da água 72
Figura 33: (a) imagem resultante da aplicação do IIA; (b) imagem resultante da aplicação do NDVI 73
Figura 34: (a) plano de voo de um VANT; (b) exemplo de um ortomosaico 75
Figura 35: Exemplos de imagens dos conjuntos: DS1 (a); DS2 (b); DS3 (c); DS4 (d); DS5 (e); DS6 (f); DS7 (g) 84
Figura 36: Definição da métrica IoU 87
Figura 37: Diagrama esquemático da abordagem proposta 92
Figura 38: Diagrama do funcionamento do método para geração de ortomosaicos 93
Figura 39: Arquitetura da RNC_Detec_Obj_Reserv 97
Figura 40: Diagramas do funcionamento do método para a detecção dos reservatórios d'água domésticos: (a) treinamento; (b) testes com a RNC treinada. 98
Figura 41: Arquitetura da RNC_Detec_Obj_Outros 98
Figura 42: Diagramas do funcionamento do método para a detecção de outros objetos-alvo: (a) treinamento; (b) testes com a RNC treinada. 100
Figura 43: Método BoVW+SVM utilizado para a detecção de cenários 100
Figura 44: Diagramas do funcionamento do método BoVW+SVM para a detecção de cenários: (a) treinamento; (b) testes com o classificador SVM treinado. 104
Figura 45: Arquitetura da RNC_Detec_Cenarios 105
Figura 46: Diagramas do funcionamento do método para a detecção de cenários: (a) treinamento; (b) testes com a RNC treinada. 105
Figura 47: Passos do AG desenvolvido para fornecer o IIAO 107
Figura 48: Diagrama esquemático do método proposto para reconstituição de bandas espectrais 110
Figura 49: Diagramas de funcionamento do método para a geração de ortomosaicos anotados e relatórios: (a) ortomosaicos anotados e relatórios para objetos-alvo; (b) imagens anotadas e relatórios para cenários 112
Figura 50: Ortomosaicos gerados com imagens dos conjuntos DS3 (aquisição1), DS3 (aquisição 2) e DS5 113
Figura 51: Resultados das detecções dos reservatórios d'água domésticos pela RNC_Detec_Obj_Reserv 115

Figura 52: Resultados das detecções dos reservatórios d'água domésticos pela RNC_Detec_Obj_Reserv 116
Figura 53: Resultados das detecções dos reservatórios d'água no Ortomosaico_Guaianases_1 118
Figura 54: Resultados das detecções dos reservatórios d'água no Ortomosaico_Guaianases_2: (a) ortomosaico completo; (b) subimagens do ortomosaico com os objetos-alvo detectados 119
Figura 55: Resultados das detecções dos reservatórios d'água no Ortomosaico_PortoAreia: (a) ortomosaico completo; (b) subimagem do ortomosaico com os objetos-alvo detectados 121
Figura 56: Resultados das detecções dos cenários pela RNC_Detec_Obj_Outros 123
Figura 57: Resultados do processo de classificação da SVM: (a) imagem de entrada; (b) imagem classificada 126
Figura 58: Resultados das detecções dos cenários pela RNC_Detec_Cenarios 128
Figura 59: Resultados das detecções dos cenários pela RNC_Detec_Cenarios 129
Figura 60: Resultados das detecções dos cenários pela RNC_Detec_Cenarios 130
Figura 61: Alguns resultados obtidos com o IIAO criado. (a) imagens RGB (I_VIS); (b) imagens NIR (I_NIR); (c) Imagens com os resultados esperados (imagens anotadas); (d) Imagens geradas pelo IIAO considerando as bandas NIR e B extraídas das imagens mostradas nas colunas a e b 133
Figura 62: Resultados da detecção de pequenas porções de água. (a) imagens NIR (I_NIR); (b) Imagens RGB originais (I_VIS); c) Imagens RGB reconstituídas; (d) Imagens resultantes da aplicação do IIAO (I_GRAY); (e) imagens binárias (I_BIN) 133
Figura 63: Ortomosaico_Guaianases_1 com as detecções dos objetos-alvo e as demarcações de acordo com as coordenadas georreferenciadas 136
Figura 64: Imagens classificadas pela RNC_Detec_Cenarios com detecções dos objetos-alvo e as demarcações de acordo com as coordenadas georreferenciadas 138

Lista de Tabelas

Tabela 1: APs calculadas para cada classe da RNC_Detec_Obj_Reserv..... 117
Tabela 2: APs calculadas para cada classe da RNC_Detec_Obj_Outros 122
Tabela 3: mAPs-50 calculados para cada combinação de descritores..... 125
Tabela 4: APs calculadas para cada classe do método BoVW+SVM 126
Tabela 5: APs calculadas para cada classe da RNC_Detec_Cenarios 131
Tabela 6: Resultados qualitativos obtidos de experimentos considerando imagens anotadas 134

Lista de Quadros

Quadro 1: Medidas estatísticas calculadas a partir das matrizes de coocorrência 36
Quadro 2: Composição da base de imagens utilizada neste trabalho 83
Quadro 3: Ambientes de programação usados no desenvolvimento dos principais algoritmos que contemplam as etapas da abordagem proposta..... 85
Quadro 4: Classes definidas para o treinamento da RNC_Detec_Obj_Reserv 96
Quadro 5: Classes utilizadas para o treinamento da RNC_Detec_Obj_Outros 99
Quadro 6: Classes utilizadas para o treinamento do SVM 102
Quadro 7: Hiperparâmetros otimizados para o SVM multiclasse 103
Quadro 8: Classes utilizadas para o treinamento do SVM 106
Quadro 9: Relatório dos objetos-alvo baseado no conjunto DS3 137
Quadro 10: Relatório dos cenários baseado no conjunto DS6 139

Lista de Siglas

AG	Algoritmo Genético
AP	*Average Precision*
BoVW	*Bag of Visual Words*
CLCM	*Color-Level Co-Occurence Matrices*
DSSD	*Deconvolutional Single Shot Detector*
GLCM	*Gray-Level Co-Occurence Matrices*
GSD	*Ground Sample Distance*
HOG	*Histogram of Oriented Gradients*
IIA	Índice Indicador de Água
IIAO	Índice Indicativo de Água Otimizado
LBP	*Local Binary Pattern*
MAE	*Mean Absolute Error*
mAP	*mean Average Precision*
MLP	*Multilayer Perceptron*
MPRI	*Modified Photochemical Reflectance Index*
NDVI	*Normalized Difference Vegetation Index*
NDWI	*Normalized Difference Water Index*
NIR	*Near Infrared*
NMS	*Non-maximal suppression*
OAO	*One Against One*

OAA	One Against All
OMS	Organização Mundial da Saúde
PCA	*Principal Component Analysis*
RBF	*Radial Basis Function*
R-CNN	*Region Convolutional Neural Network*
REM	Radiação Eletromagnética
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
RTK	*Real Time Kinematic*
SIFT	*Scale-Invariant Feature Transform*
SR	Sensoriamento Remoto
SSD	*Single Shot Multibox Detector*
SSIM	*Structural Similarity Index*
SURF	*Speeded-Up Robust Features*
SVC	Sistema de Visão Computacional
SOM	*Self-Organization Map*
SVM	*Support Vector Machine*
SWIR	*Short-wavelength infrared*
VANT	Veículo Aéreo Não-tripulado
VC	Visão Computacional
YOLO	*You Only Look Once*

Lista de Símbolos

ap_i	Aptidão de um cromossomo, que mede o quanto ele é adequado para satisfazer à especificação de um problema de um AG
b	Bias de uma Rede Neural Artificial
bin	Valor relacionado a um histograma, que representa o número de vezes que cada tom de cinza aparece na imagem
vl_i	Valor lógico para um pixel (0 ou 1)
$B(l, c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem na banda espectral B (*Blue*)
d	Distância entre dois pixels
c_x e c_y	Coordenadas do canto superior esquerdo da caixa delimitadora definida para as RNCs do *framework* YOLOv3
$\mathcal{D}$	Domínio de uma imagem
f_{con}	Medida de contraste de uma imagem obtida a partir das matrizes de coocorrência para o descritor CLCM
f_{corr}	Medida de correlação de uma imagem obtida a partir das matrizes de coocorrência para o descritor CLCM
f_{ent}	Medida de entropia de uma imagem obtida a partir das matrizes de coocorrência para o descritor CLCM
f_{hom}	Medida de homogeneidade de uma imagem obtida a partir das matrizes de coocorrência para o descritor CLCM
f_{ma}	Medida de momento angular de uma imagem obtida a partir das matrizes de coocorrência para o descritor CLCM
f_{var_i} e f_{var_j}	Medidas de variâncias de uma imagem obtidas a partir das matrizes de coocorrência para o descritor CLCM
$f(x)$	Função de ativação de uma Rede Neural Artificial
$Gamma$	Parâmetro utilizado no classificador SVM
gr	Iteração completa do AG que gera uma nova população
$G(l, c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem na banda espectral G (*Green*)
$\tilde{G}(l, c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem na banda espectral G (*green*) contaminada pelo uso da lente infravermelho

I	Imagem
I_COR	Imagem multibanda
I_GRAY	Imagem em tons de cinza
I_BIN	Imagem binária
I_ESP	Imagem binária do resultado esperado
I_IIAO	Imagem resultante da aplicação do $IIAO$
I_NIR	Imagem em infravermelho
I_RGB	Imagem colorida
I_VIS	Imagem no espectro visível
IW	Pesos da entrada da Rede Neural
$\mathbb{K}$	Conjunto de intensidade de cores de uma imagem
k_1 *e* k_2	Bandas selecionadas do conjunto Ω para o cômputo do $IIAO$
l e c	Coordenadas horizontal e vertical de um pixel
$LBP\ (q_c(l,c)\ )$	Valor do LBP para o pixel central na coordenada l (linha) e c (coluna)
LW	Pesos das camadas de uma Rede Neural Artificial
m	Quantidade de bandas de um sistema de cores
n	Quantidade de *bits* utilizados para quantização da imagem
$nbins$	Número de bins de um histograma
ncl	Número de classes de um classificador SVM
nc	Número de colunas de uma imagem
nl	Número de linhas de uma imagem
$NIR(l,c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem NIR
o_1 e o_2	Operações aritméticas para o cômputo do $IIAO$
$p_1, p_2, \dots, p_R$	Entradas associadas à Redes Neurais Artificiais
p_w e p_h	Largura e altura da caixa delimitadora definida para as RNCs do *framework* YOLOv3
pl_3, pl_4 e pl_5	Palavras visuais referentes à técnica BoVW

$q\ (l, c)$	Notações de pixel
p_{interp}	Precisão interpolada referente ao cálculo da AP
pr_i	Probabilidade de seleção de um AG
$R(l, c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem na banda espectral R (*Red*)
$\tilde{R}(l, c)$	Um pixel na coordenada l (linha) e c (coluna) correspondente a uma imagem na banda espectral R (*Red*) contaminada pelo uso da lente infravermelho
S	Número de neurônios de uma Rede Neural Artificial
$S(gr)$	População de cromossomos na geração gr
$S\mathrm{L}$	Medida de similaridade utilizada no AG
$t_{i,j}$	Número de transições de níveis de cinza de uma imagem
t_x, t_y, t_w e t_h	Coordenadas da caixa delimitadora definida para as RNCs do *framework* YOLOv3
T	Limiar (*threshold*) definido para a limiarização de uma imagem
$v_1\ e\ v_2$	Valores constantes para o cômputo do $IIAO$
$W_1, W_2, \dots, W_R$	Pesos associados às Redes Neurais Artificiais
y	Saída de uma Rede Neural Artificial
$\mathbb{Z}$	Conjunto dos números inteiros
Z^k	Espaço de características relacionado aos bins de um histograma
μ_j	Valor médio das distribuições marginais para as matrizes de coocorrência
β	Ângulo entre os pixels
θ	Limiar definido para um neurônio de uma Rede Neural Artificial
Ω	Conjunto de bandas espectrais para o cômputo do $IIAO$
α	Número real

Sumário

1. INTRODUÇÃO 21
1.1. CONTEXTUALIZAÇÃO DO TEMA 21
1.2. PROBLEMA DE PESQUISA 22
1.3. OBJETIVOS 24
1.3.1. OBJETIVO GERAL 24
1.3.2. OBJETIVOS ESPECÍFICOS 25
1.4. JUSTIFICATIVA, IDENTIFICAÇÃO DAS LACUNAS E QUESTÕES DE PESQUISA 25
1.5. CONTRIBUIÇÕES DA PESQUISA 26
1.6. ORGANIZAÇÃO DA PESQUISA 27
2. FUNDAMENTAÇÃO TEÓRICA 28
2.1. VISÃO COMPUTACIONAL 28
2.1.1. DESCRITORES DE IMAGENS 31
2.1.1.1. HISTOGRAMA DE CORES 32
2.1.1.2. COLOR-LEVEL CO-OCCURRENCE MATRICES (CLCM) 34
2.1.1.3. LOCAL BINARY PATTERN (LBP) 37
2.1.1.4. HISTOGRAM OF ORIENTED GRADIENTS (HOG) 39
2.1.1.5. BAG OF VISUAL WORDS (BOVW) 41
2.2. APRENDIZAGEM DE MÁQUINA 45
2.2.1. SUPPORT VECTOR MACHINE (SVM) 46
2.2.2. REDES NEURAIS ARTIFICIAIS (RNA) 49
2.2.2.1. REDES NEURAIS MULTILAYER PERCEPTRONS (MLP) 52
2.2.2.2. REDES NEURAIS CONVOLUCIONAIS 55
2.2.2.2.1. FRAMEWORK YOLOv3 59
2.2.3. ALGORITMOS GENÉTICOS 66
2.3. SENSORIAMENTO REMOTO 69
2.3.1. SENSORIAMENTO REMOTO USANDO VANTS 73

2.4. TRABALHOS ABORDANDO TEMAS CORRELATOS ENCONTRADOS NA LITERATURA 76
3. MÉTODOS E MATERIAIS 80
3.1. CARACTERIZAÇÃO DA PESQUISA 80
3.2. MATERIAIS 80
3.2.1. BASE DE IMAGENS 80
3.2.2. AMBIENTES COMPUTACIONAIS, SOFTWARES E HARDWARE EMPREGADOS NA CONDUÇÃO DOS EXPERIMENTOS 85
3.3. PROCEDIMENTO PARA CONDUÇÃO DOS EXPERIMENTOS E AVALIAÇÃO DOS ALGORITMOS E ABORDAGENS DESENVOLVIDAS 86
3.3.1. MÉTRICA PARA AVALIAÇÃO DOS MÉTODOS DE DETECÇÃO DE OBJETOS E CENÁRIOS 86
3.3.2. MÉTRICAS PARA AVALIAÇÃO DO MÉTODO DE PEQUENAS PORÇÕES DE ÁGUA 90
4. ABORDAGEM PROPOSTA E RESULTADOS EXPERIMENTAIS 92
4.1. ABORDAGEM PROPOSTA 92
4.1.1. GERAÇÃO DO ORTOMOSAICO 93
4.1.2. DETECÇÃO DE OBJETOS-ALVO E CENÁRIOS 94
4.1.2.1. DETECÇÃO DE OBJETOS-ALVO UTILIZANDO O FRAMEWORK YOLOV3 95
4.1.2.2. DETECÇÃO DE CENÁRIOS UTILIZANDO BAG OF VISUAL WORDS 100
4.1.2.3. DETECÇÃO DE CENÁRIOS UTILIZANDO A ARQUITETURA TINY-YOLOV3 104
4.1.3. DETECÇÃO DE PEQUENAS PORÇÕES DE ÁGUA 107
4.1.3.1. RECONSTITUIÇÃO DAS BANDAS ESPECTRAIS 110
4.1.4. GERAÇÃO DE ORTOMOSAICOS ANOTADOS E RELATÓRIOS COM POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI 111
4.2. EXPERIMENTOS CONDUZIDOS COM A ABORDAGEM PROPOSTA 112
4.2.1. GERAÇÃO DO ORTOMOSAICO 112
4.2.2. DETECÇÃO DE OBJETOS-ALVO E CENÁRIOS 114
4.2.2.1. DETECÇÃO DE OBJETOS-ALVO UTILIZANDO O FRAMEWORK YOLOV3 114

4.2.2.2. DETECÇÃO DE CENÁRIOS UTILIZANDO BOVW+SVM 124
4.2.2.3. DETECÇÃO DE CENÁRIOS UTILIZANDO A ARQUITETURA TINY-YOLOV3 127
4.2.3. DETECÇÃO DE PEQUENAS PORÇÕES DE ÁGUA 131
4.2.4. GERAÇÃO DE ORTOMOSAICOS ANOTADOS E RELATÓRIOS COM POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI 135
5. CONCLUSÕES E TRABALHOS FUTUROS 140
REFERÊNCIAS 143

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO DO TEMA

As epidemias de dengue, zika, chikungunya e febre amarela urbana, causadas pelo mosquito *Aedes aegypti*, vem preocupando muito as autoridades da área de saúde, não só Brasil, mas do mundo todo. De acordo com a OMS (Organização Mundial da Saúde), as doenças causadas pelo *Aedes aegypti* atingem aproximadamente 390 milhões de pessoas por ano e, só no Brasil, foram registrados mais de 440 mil casos em 2017. Não obstante, dados do Ministério da Saúde apontam que neste verão de 2019 mais de 500 cidades brasileiras correm o risco de surto dessas doenças e outras 1.881 cidades estão em sinal de alerta (OSP, 2019).

O combate ao mosquito vem sendo um dos principais desafios, pois nem sempre as medidas de prevenção, mesmo com a ampla divulgação pela mídia, são realizadas pela população de forma adequada (MS, 2019). Paralelamente às campanhas de prevenção, vários esforços vêm sendo feitos para agilizar a busca por possíveis criadouros do mosquito transmissor, principalmente nas áreas urbanas de difícil acesso pelos agentes de vigilância sanitária e moradores das habitações. Um deles é o uso de VANTs, mais conhecidos como drones, para a aquisição de imagens aéreas em locais com maior incidência da doença (DINIZ e MEDEIROS, 2018; PMS, 2019; G1-DF, 2019; ONUBR, 2019). No entanto, as imagens adquiridas são analisadas de forma manual (visualmente).

O uso de VANTs tem aumentado substancialmente nos últimos anos, principalmente em tarefas que necessitam de imagens aéreas, as quais são posteriormente processadas e analisadas (CÂNDIDO, SILVA e FILHO, 2015; CASSEMIRO e PINTO, 2014; DINIZ e MEDEIROS, 2018; JARDIM, 2018). Atualmente os VANTs são usados na automatização de tarefas em diversas áreas do conhecimento tais como agricultura de precisão (ALVES, FERREIRA e CUSTÓDIO, 2017; ZHOU et al., 2014; REINECKE, PRINSLOO e CUSTÓDIO, 2017; JARDIM, 2018); sensoriamento remoto (AGUIRRE-GÓMEZ et al., 2016, ALBUQUERQUE et al., 2017) e saúde (AGRAWAL et al., 2014; CAPOLUPO et al., 2014; MEHRA et al., 2016; PASSOS et al., 2018). Dessas, a agricultura de precisão talvez seja a área que

mais tem tirado proveito do uso de VANTs, principalmente no que tange à aquisição de imagens para práticas agrícolas e uso do solo.

Há trabalhos encontrados na literatura nos quais VANTs foram utilizados no mapeamento de focos do mosquito *Aedes aegypti,* a partir da análise automática das imagens. Todavia, as abordagens propostas nestes trabalhos são incompletas no que diz respeito à localização de objetos suspeitos de serem criadouros e à detecção de pequenas porções de água, que sinalizam potenciais proliferações de larvas do mosquito.

Tendo em vista o aumento de casos de doenças causadas pelo mosquito *Aedes aegypti* e as dificuldades encontradas em se ter um controle eficaz sobre os criadouros do mosquito, principalmente em locais de difícil acesso, torna-se importante o desenvolvimento de abordagens computacionais voltadas para a identificação automática de possíveis criadouros, a partir de imagens aéreas adquiridas por VANTs, que pode trazer benefícios para a área da saúde bem como para a população das regiões beneficiadas por tais sistemas.

1.2. PROBLEMA DE PESQUISA

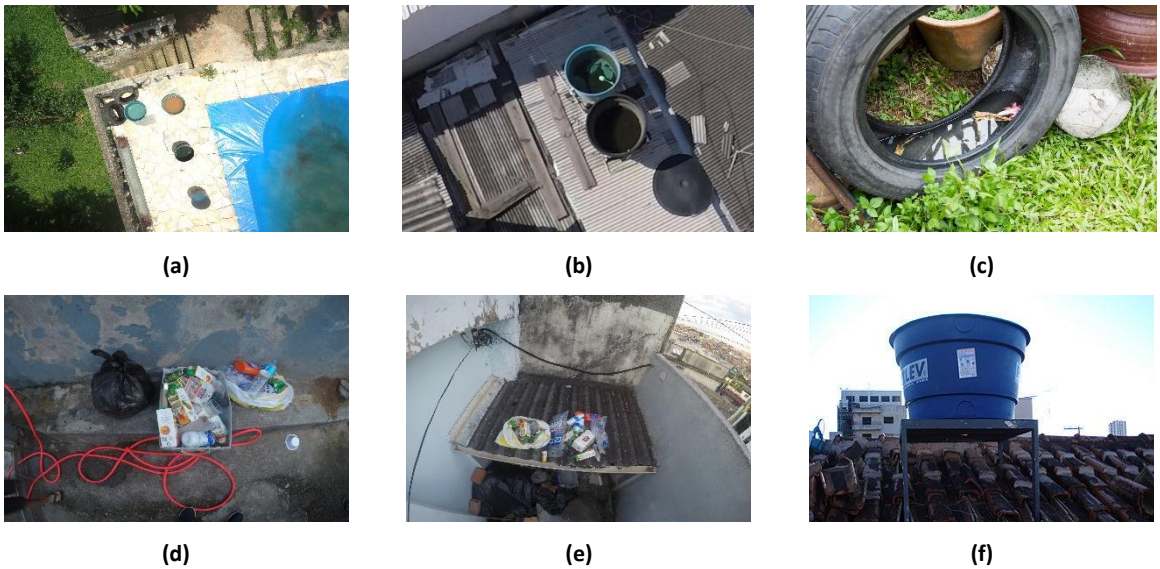
O mosquito *Aedes aegypti* macho alimenta-se exclusivamente de frutas, enquanto a fêmea necessita de sangue para o amadurecimento dos ovos que são depositados separadamente nas paredes internas dos objetos, próximos a superfícies de água limpa, local que lhes oferece melhores condições de sobrevivência (MS, 2019). Quanto à capacidade de voo, é sabido que o mosquito possui possibilidade de acesso a alturas como, por exemplo, chegar às caixas d'água, às calhas e terraços de edificações urbanas. Contudo, sua potencialidade de voo não atingiria um prédio de quatro andares. Apesar disso, ele pode chegar a alturas mais elevadas se estiver alojado em elevadores, embalagem de materiais em geral, brinquedos, caixas de ferramentas e uma infinidade de outros objetos que podem conduzi-lo até a cobertura de qualquer edifício.

São diversos os locais que podem acumular água e se tornar potenciais criadouros do *Aedes aegypti*, alguns dos quais estão ilustrados na Figura 1. O

mosquito põe seus ovos em recipientes como latas e garrafas vazias, pneus, calhas, reservatórios d'água domésticos descobertos, pratos sob vasos de plantas ou qualquer outro objeto que possa armazenar água da chuva e, portanto, podem ser caracterizados como objetos suspeitos. Se a água estiver bem tratada e com a concentração recomendada de cloro, o mosquito não se desenvolve. Já foi comprovado que a água com cloro e a água salgada funcionam como repelentes.

De acordo com o Ministério da Saúde, pesquisas realizadas em campo indicam que os reservatórios, como caixas d'água, galões e tonéis muito utilizados para armazenagem de água para uso doméstico em locais dotados de infraestrutura urbana precária são os criadouros que mais produzem *Aedes aegypti* e, portanto, os mais críticos. Soma-se a isso o fato de tais reservatórios normalmente estarem em locais de difícil acesso, sobre lajes ou telhados por exemplo, que possam ser ignorados durante as vistorias dos agentes de saúde. Em adição, em tais locais pode haver acúmulo de lixo contendo caixas, papéis, pneus e outros recipientes destampados além de outras formas de acúmulo de água como, por exemplo, em calhas ou mesmo na própria laje, como ilustrado na Figura 1. Na tentativa de mapear esses locais, aquisições de imagens vêm sendo realizadas por meio de VANTs, porém o mapeamento de criadouros pelos agentes de saúde, na maioria das vezes, é feito de forma manual (visualmente), o que pode ocasionar demora nos procedimentos de inspeção.

Figura 1: Exemplos de possíveis criadouros do mosquito *Aedes aegypti*



(a) **(b)** **(c)**

(d) **(e)** **(f)**

Além de objetos isolados há também a questão de cenários que podem ser caracterizados como potenciais criadouros do mosquito *Aedes aegypti*. Devido ao fato de os reservatórios d'água domésticos (Figuras 1b e 1f) serem os objetos que mais aparecem no topo das edificações urbanas e pelos motivos já citados, neste trabalho eles são referenciados como objetos-alvo. Além dos reservatórios, pneus velhos (Figuras 1a e 1c), calhas (Figuras 1a e 1e) e outros recipientes como os da Figura 1a que podem acumular água também são considerados objetos-alvo. Os cenários considerados neste trabalho, ilustrados nas Figuras 1d e 1e, são constituídos de lixo inorgânico a céu aberto contendo objetos pequenos que podem acumular água (por exemplo, garrafas pet e outras embalagens de plástico ou papel) ou mesmo a junção de vários objetos difíceis de serem identificados isoladamente.

Na questão da detecção de água em imagens aéreas urbanas, há vários trabalhos na literatura de sensoriamento remoto, tais como Zhou et al. (2014); Yang e Chen (2017) e Khandelwal et al. (2017), que utilizam várias bandas espectrais em sensores levados a bordo de satélites. No entanto, o uso de VANTs para a mesma tarefa torna-se mais difícil devido à baixa resolução espectral (pequena quantidade de bandas espectrais) dos sensores das câmeras que são acopladas a esses equipamentos.

Nesse contexto, surge a seguinte questão de pesquisa: como desenvolver uma abordagem computacional para identificar automaticamente, a partir de imagens aéreas adquiridas por VANTs, locais que representam potenciais criadouros do mosquito *Aedes aegypti*, levando em conta que tais imagens possuem alta complexidade de detalhes e baixa resolução espectral?

1.3. OBJETIVOS

1.3.1. OBJETIVO GERAL

Desenvolver uma abordagem de visão computacional para identificação automática de possíveis focos do mosquito *Aedes aegypti* em áreas urbanas, a partir de imagens aéreas adquiridas por VANTs.

1.3.2. OBJETIVOS ESPECÍFICOS

A abordagem proposta é composta por quatro etapas: geração de ortomosaico, detecção de objetos e cenários suspeitos, detecção de pequenas porções de água, e geração de ortomosaicos anotados e relatórios com as indicações de possíveis criadouros. Para isso, alguns objetivos específicos são elencados a seguir:

- Propor, a partir de algoritmos de reconhecimento de padrões em imagens, métodos capazes de detectar e localizar os objetos e cenários suspeitos, principalmente sobre lajes e telhados, considerados potenciais criadouros do mosquito *Aedes aegypti*.
- Propor um indicador (combinação aritmética entre duas ou mais bandas espectrais disponíveis), para evidenciar pequenas porções de água nas imagens;
- Conceber, a partir dos algoritmos de reconhecimento de padrões, um método para sinalizar locais com alguma probabilidade de ocorrência de criadouro do mosquito *Aedes aegypti*;
- Gerar ortomosaicos anotados e relatórios com as indicações de possíveis criadouros do mosquito *Aedes aegypti*.

Com base no exposto, o termo abordagem pode ser definido como o encadeamento de métodos, técnicas e algoritmos computacionais que permitem responder à questão de pesquisa formulada na seção anterior.

1.4. JUSTIFICATIVA, IDENTIFICAÇÃO DAS LACUNAS E QUESTÕES DE PESQUISA

As novas tecnologias, como os VANTs, podem significar boas alternativas para auxiliar os profissionais da área de saúde na busca por possíveis criadouros do mosquito *Aedes aegypti*, já que consistem em equipamentos de baixo custo, se comparado às aeronaves tripuladas usadas em tarefas de aquisição de imagens. Além disso, eles possibilitam pilotagem remota, voos mais próximos ao solo e a aquisição de imagens com grande nível de detalhamento e altas resoluções

temporais, que permitem a detecção de pequenos objetos na superfície terrestre e a percepção de mudanças em uma determinada região num curto espaço de tempo. Em termos de resolução temporal, o uso de VANTs permite o monitoramento de regiões em um curto espaço de tempo, o que já não é muito comum com imagens de satélite, pois tal tarefa demanda alto custo.

Dos trabalhos encontrados na literatura relatando o uso de VANTs em tarefas de identificação de criadouros do mosquito *Aedes aegypti*, apenas os trabalhos de Agrawal et al. (2014) e Mehra et al. (2016) abordaram a análise automática das imagens. Contudo, os trabalhos indicam apenas a presença de cenários suspeitos nas imagens, sem fornecer localização espacial. Em adição, os dois trabalhos abordam superficialmente a presença de água nos cenários ou objetos suspeitos, condição que aumenta a possibilidade de haver criadouro em um determinado local.

Não obstante, embora a detecção de corpos d'água (oceanos, mares e grandes lagos) em imagens aéreas seja um problema amplamente conhecido e explorado na literatura, a detecção de pequenas porções de água em imagens de satélite não é usual devido ao nível de detalhamento, que pode ser prejudicado em razão da distância em relação à superfície terrestre. Com o crescente uso de VANTs esta tarefa torna-se viável do ponto de vista do nível de detalhamento das imagens, mas fica prejudicada em virtude das baixas resoluções espectral e radiométrica. Assim, índices indicadores propostos para extração de feições d'água em imagens de satélites, como os descritos em Zhou et al. (2014); Yang e Chen (2017) e Khandelwal et al. (2017) dificilmente podem ser aplicados em imagens adquiridas por VANTs.

1.5. CONTRIBUIÇÕES DA PESQUISA

Do ponto de vista científico, a pesquisa descrita neste trabalho pode trazer contribuições tanto para a área de visão computacional, uma vez que subproblemas dessa área foram investigados e resolvidos, quanto para a área de sensoriamento remoto e outras áreas do conhecimento que tratam de assuntos correlatos. Neste sentido, pode-se destacar como principais contribuições deste trabalho:

a) Desenvolvimento de métodos computacionais, utilizando técnicas de Inteligência Artificial e de Aprendizagem de Máquina, para a detecção de objetos e cenários suspeitos de serem possíveis criadouros;
b) Desenvolvimento de um método computacional para a reconstituição de bandas espectrais. Este método é muito importante quando se necessita utilizar lentes especiais para a filtragem de alguma banda espectral específica como, por exemplo, o infravermelho próximo. Isso porque, ao se utilizar tal lente, pelo menos uma das bandas do espectro visível é perdida enquanto as demais são distorcidas;
c) Desenvolvimento de um método que emprega Algoritmo Genético para geração de indicadores que permitem identificar pequenas porções de água nas imagens, considerando o conjunto de bandas espectrais disponível;
d) Composição de uma base de imagens aéreas contendo objetos-alvo e cenários suspeitos de serem possíveis criadouros do mosquito *Aedes aegypti*. A base que contempla os cenários poderá ser disponibilizada em um *website* de domínio público para que outros pesquisadores possam utilizá-las.

1.6. ORGANIZAÇÃO DA PESQUISA

No segundo capítulo deste trabalho é apresentada a fundamentação teórica, onde são descritos os conceitos necessários para o entendimento da abordagem proposta nesta pesquisa.

O terceiro capítulo contém a caracterização da pesquisa, bem como os materiais e métodos utilizados na condução dos experimentos.

No quarto capítulo é descrita em detalhes a abordagem de visão computacional proposta e os resultados obtidos nos experimentos realizados.

Por fim, no quinto capítulo, são apresentadas as conclusões, bem as sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são explorados os tópicos relacionados ao desenvolvimento teórico deste trabalho como, por exemplo, as definições sobre sistemas de visão computacional, a utilização de imagens adquiridas por VANTs em sensoriamento remoto e as técnicas de reconhecimento de padrões e aprendizagem de máquina.

2.1. VISÃO COMPUTACIONAL

A visão computacional (VC) é o processo de modelagem e replicação da visão humana usando software e hardware. A visão computacional é uma disciplina que estuda como reconstruir, interromper e compreender uma cena 3D a partir de suas imagens 2D em termos das propriedades da estrutura presente na cena (SZELISKI, 2011). Alguns autores definem VC como uma ciência que faz com que as máquinas possam enxergar tornando possível assim a realização de tarefas como, por exemplo, o reconhecimento de objetos em imagens digitais (GONZALEZ; WOODS, 2002).

Uma imagem digital pode ser representada por uma função bidimensional $I: \mathcal{D} \rightarrow \mathbb{K}$, que mapeia uma grade retangular $\mathcal{D} \subseteq \mathbb{Z}^2$ em um conjunto de intensidade de cores $\mathbb{K} = \{0, 1, \dots, 2^n - 1\}^m$, onde n é a quantidade de *bits* utilizados para quantização da imagem e m é a quantidade de bandas de um sistema de cores (FILHO; NETO, 1999). Com base nessas definições, um pixel é um elemento do domínio da imagem, sendo denotado por $q\ (l, c) \in \mathcal{D}$ cujos valores l e c são as coordenadas horizontal e vertical de q.

Uma imagem binária é caracterizada por possuir apenas duas intensidades de cinza: preto (intensidade mínima) ou branco (intensidade máxima), representadas por 0 e 1, respectivamente. Assim, considerando $n = 1$ e $m = 1$, denota-se uma imagem binária por $I\_BIN: \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{K} = \{0, 1\}$.

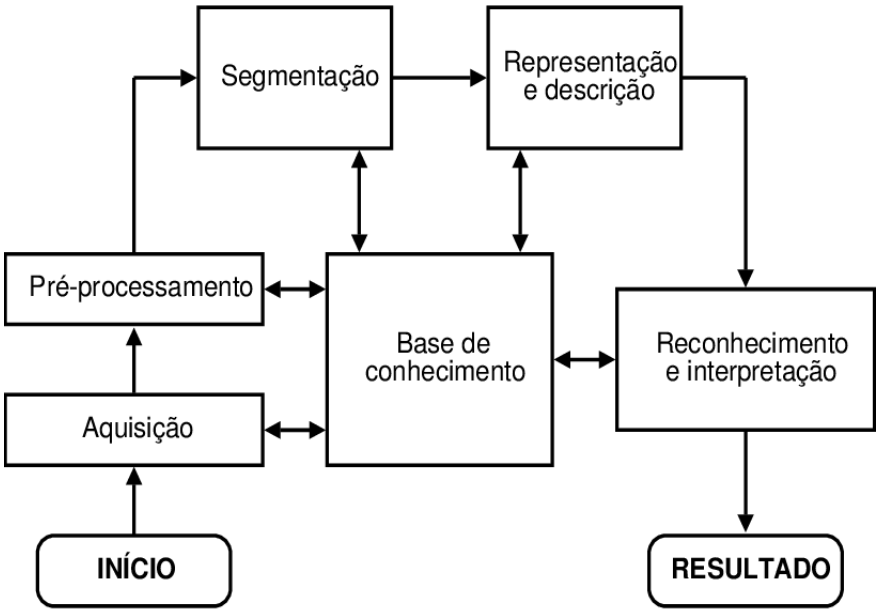
Quando $n > 1$ e $m = 1$, têm-se as imagens em níveis de cinza, na qual os pixels pode assumir valores variando de 0 até $2^n - 1$ representando as diferentes

intensidades de cinza. Nesse caso denota-se a imagem por $I\_GRAY: \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{K} = \{0, 1, \ldots, 2^n - 1\}$.

Já uma imagem multibanda, também chamada de imagem colorida, pode ser denotada por $I\_COR: \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{K} = \{0, 1, \ldots, 2^n - 1\}^n$. Assim, uma imagem RGB (*Red, Green, Blue*) pode ser denotada por $I\_RGB: \mathcal{D} \subseteq \mathbb{Z}^2 \rightarrow \mathbb{K} = \{0, 1, \ldots, 2^n - 1\}^3$, sendo cada pixel representado por três valores indicando a intensidade de vermelho (R), a intensidade de verde (G) e a intensidade de azul (B). Vale ressaltar que uma imagem multibandas pode ser entendida como um vetor de imagens de níveis de cinza, isto é, $I\_RGB = (I_R, I_G, I_B)$, sendo I_R, I_G e I_B as imagens com intensidades de vermelho, verde e azul (GONZALEZ; WOODS, 2002).

Baseado em Gonzalez e Woods (2002), um Sistema de Visão Computacional (SVC) envolve etapas, as quais são ilustradas na Figura 2.

Figura 2: Etapas de um sistema de visão computacional



Fonte: Gonzalez e Woods (2002)

Ainda segundo Gonzalez e Woods (2002), os passos mostrados na Figura 2 podem ser descritos da seguinte forma:

- Aquisição de imagens: refere-se à forma em que a imagem é adquirida, seja por meio de scanner ou uma câmera, trabalhando *on-line* ou *off-line*. Nos experimentos realizados nesta pesquisa a aquisição é feita *off-line* por um VANT.
- Pré-processamento: tem como objetivo melhorar a qualidade da imagem de forma a aumentar as chances de sucesso nas próximas etapas de processamento. Neste trabalho o método para a reconstituição de bandas espectrais, abordado na seção 4.1.3.1, representa um pré-processamento das imagens que foram submetidas à detecção de pequenas porções de água. Além disso, o escalonamento das imagens realizado no algoritmo para a detecção de objetos-alvo e cenários, detalhado na seção 2.2.2.2.1, também caracteriza um pré-processamento.
- Segmentação: é realizada para separar da imagem apenas os fragmentos ou objetos que são interessantes para a análise. Neste trabalho, a função da segmentação é separar, nas imagens, apenas os objetos-alvo e cenários que podem caracterizar possíveis criadouros do mosquito *Aedes aegypti*. Um dos métodos mais simples utilizados na fase de segmentação de imagens é a limiarização, que consiste na separação de uma imagem em diferentes regiões de acordo com a distribuição de níveis de cinza. Nesse método, o valor de cada pixel da imagem é comparado com um limiar definido por meio da análise do histograma (WEEKS, 1996). Uma vez definido o limiar, é possível determinar dois grupos: o fundo da imagem e os objetos contidos nela, onde os pontos com intensidade menor que o limiar definido são considerados partes do fundo da imagem, e os demais são considerados partes dos objetos. Matematicamente, a limiarização de uma imagem I_GRAY pode ser definida de acordo com a Equação 1.

$$I\_BIN = \begin{cases} 0, se\ q\ (l,c) > T \\ 1, se\ q\ (l,c) \leq T \end{cases} \quad (1)$$

sendo $q\ (l, c)$ um pixel pertencente à imagem I_GRAY, T o limiar selecionado e I_BIN a imagem resultante da limiarização. Os pontos rotulados em I_BIN com o valor 1 correspondem ao objeto, enquanto os rotulados com 0 correspondem ao fundo da imagem, ou vice-versa.

- Representação e descrição: envolve a extração de características dos objetos, as quais são utilizadas na etapa reconhecimento e interpretação. Neste trabalho pode-se citar, como exemplo, a utilização de descritores de texturas que têm a função de extração de características das imagens, os quais são detalhados na seção 2.1.1 a seguir;
- Reconhecimento e interpretação: o reconhecimento é o processo que atribui um rótulo ao objeto, com base em um conjunto de informações previamente fornecidas pelo descritor. A interpretação envolve também a atribuição de um significado a um conjunto de objetos reconhecidos. Neste trabalho, os algoritmos de aprendizagem de máquina empregados nas tarefas de reconhecimento e interpretação estão descritos na seção 2.2;
- Base de conhecimento: contém o conhecimento adquirido e armazenado sobre o problema. Formas, cores e dimensões podem ser citadas como exemplos de conhecimento prévio nas tarefas de identificação de objetos-alvo e cenários que possam estar relacionados a possíveis criadouros do mosquito.

2.1.1. DESCRITORES DE IMAGENS

O descritor tem o objetivo de retornar importantes características (atributos) da imagem por meio de um conjunto de valores. Este conjunto é chamado “vetor de características” do objeto e é utilizado por algoritmos de classificação para separar os objetos contidos na imagem analisada em suas respectivas classes (SILVA et al., 2013).

É possível descrever uma região contida em uma imagem por meio da extração de características estatísticas dessa imagem utilizando, por exemplo, a análise de sua textura. Esta é uma abordagem natural, pois é uma característica utilizada para interpretar informações visuais. A textura contém informações sobre a distribuição

espacial, variação de luminosidade, suavidade, rugosidade, regularidade e descreve o arranjo estrutural das superfícies e as relações entre regiões vizinhas (PEDRINI e SCHWARTZ, 2008).

Na literatura existem diversos descritores de imagens. Nas seções 2.1.1.1 a 2.1.1.5 são apresentados aqueles empregados neste trabalho para detecção de cenários.

2.1.1.1. HISTOGRAMA DE CORES

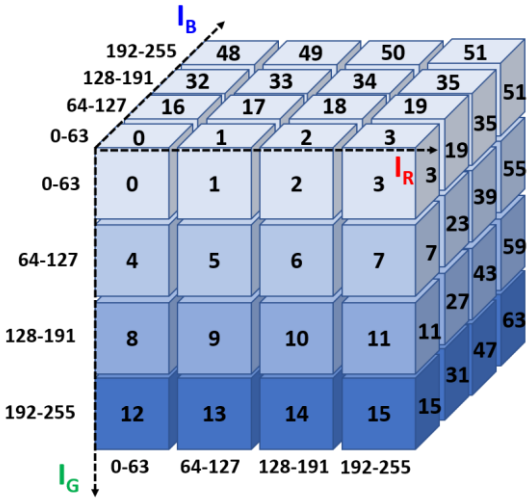
De forma geral, o histograma de uma imagem corresponde à distribuição dos seus níveis de cinza, e pode ser representado por um gráfico indicando o número de pixels na imagem para cada nível de cinza (PEDRINI e SCHWARTZ, 2008). No caso de imagens multiespectrais, cada banda é requantizada em um certo número de intervalos, de forma que o espaço de características Z^k é dividido em hipercubos (bins do histograma). Por exemplo, a partir de uma imagem colorida I_RGB, onde $I\_RGB = (I_R, I_G, I_B)$ com 8 bits para cada componente I_R, I_G, I_B, é possível dividir cada eixo do Z^3 em 4 intervalos (bins): *[0,63]*, *[64,127]*, *[128,191]* e *[192,255]*.

A contagem de cores em cada bin é usada no cálculo do histograma. Assim, para cada bin, analisa-se os níveis de cinza das 3 bandas da imagem colorida (RGB). No cubo, cada bin é representado por um número. Para o pixel da imagem RGB pertencer ao bin 0, o valor de cada banda R (eixo x), G (eixo y) e B (eixo z) tem que estar no intervalo *[0,63]*. Para estar no bin 18, o valor de R tem que estar em *[128,191]*, G em *[0,63]* e B em *[64,127]*. Dessa forma, o cálculo de qual bin um pixel vai estar, é recuperado por meio da Equação 2. Sabendo os bins de cada pixel da imagem RGB, é possível calcular o histograma da imagem.

$$bin = \frac{R}{nbins} + G * \left(\frac{4}{nbins}\right) + \left(\frac{B * 16}{nbins}\right) \quad (2)$$

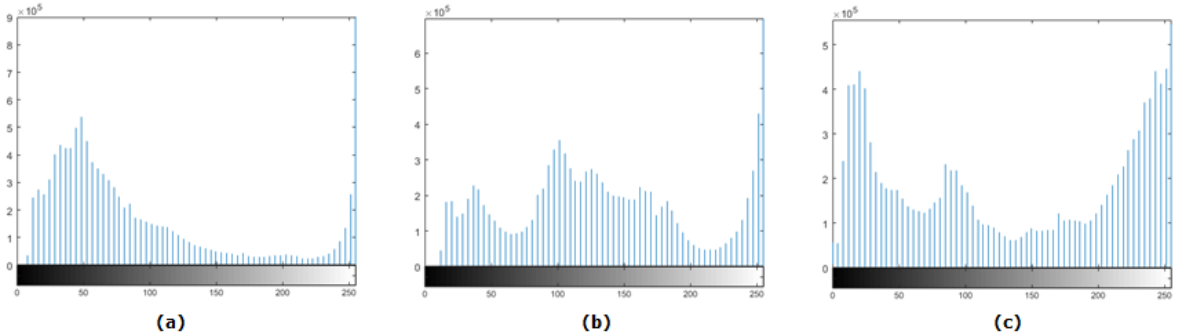
Na Figura 3 é ilustrado um exemplo para o cálculo do histograma de uma imagem colorida de 64 bins. A imagem tem 256 níveis de cinza para cada canal, de modo que eles são divididos em 4 intervalos: *[0,63]*, *[64,127]*, *[128,191]*, *[192,255]*, resultando em 64 bins.

Figura 3: Exemplo para o cálculo do histograma de uma imagem colorida de 64 bins



Na Figura 4 são mostrados exemplos de histogramas de 64 bins para os canais de cor de uma imagem, sendo a Figura 4a ilustrando o histograma para a banda *Red* (R), a Figura 4b para a banda *Green* (G) e a Figura 4c para a banda *Blue* (B).

Figura 4: Histogramas de 64 bins: (a) R, (b) G e (c) B



Devido a sua simplicidade, baixo custo de processamento, eficiência e propriedades e invariância a rotação e translação, histogramas de cores são amplamente utilizados em visão computacional para extrair características de baixo nível. Neste trabalho histogramas de cores de 128 bins foram empregados na detecção de cenários nas imagens.

2.1.1.2. COLOR-LEVEL CO-OCCURRENCE MATRICES (CLCM)

Uma das abordagens utilizadas para adquirir informações sobre transições de níveis de cinza entre dois pixels é aquela obtida por meio da construção da matriz de coocorrência, baseada na ocorrência repetida de configuração de alguns níveis de cinza na textura, variando rapidamente segundo a frequência espacial em texturas finas e lentamente em texturas ásperas. Os elementos da matriz descrevem a frequência com que ocorrem as transições de nível de cinza entre pares de pixels (PEDRINI e SCHWARTZ, 2008).

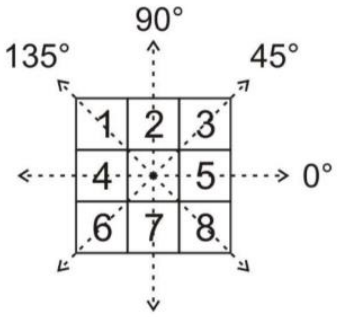
Efetuando-se variações na relação espacial, por meio de alterações na orientação e na distância entre as coordenadas dos pixels, podem ser obtidas diversas matrizes de coocorrência, a partir das quais são extraídas medidas que podem compor um vetor de características para descrição de textura de uma imagem.

Para determinar quais pixels e quais transições de níveis de cinza serão considerados, cada elemento é composto de dois pares ordenados denotando as coordenadas de cada pixel envolvido na relação. Uma vez determinado o número de ocorrências de cada uma das transições de níveis de cinza, basta acrescentar o número de transições na i-ésima linha e j-ésima coluna da matriz, obtendo-se, então, a matriz de coocorrência. Dessa maneira, pode-se definir arbitrariamente a distância e o ângulo entre os pixels para os quais serão computadas as transições apenas efetuando alterações nesse conjunto, entretanto, distâncias e ângulos distintos acabam sendo incluídos em uma mesma matriz.

Haralick et al. (1973) definem de modo mais específico quais transições devem ser utilizadas para criação de cada matriz de coocorrência por meio da utilização de dois parâmetros adicionais ($d\ e\ \beta$), exercendo controle sobre a distância e o ângulo entre os pixels, respectivamente. Dessa forma, diversas matrizes podem ser criadas,

proporcionando a obtenção de um maior número de características. Na Figura 5 são ilustrados quatro possibilidades de ângulos para o parâmetro β, os quais indicam como deve ser o relacionamento entre dois pixels. As transições para cada ângulo são computadas sempre em relação ao pixel localizado na posição central.

Figura 5: Ângulos utilizados para o cálculo das matrizes de coocorrência



Fonte: Pedrini e Scharwtz (2008)

Considerando $d = 1$, por exemplo, para cada um desses ângulos será computada uma matriz de coocorrência, que representa as transições de níveis de cinza entre pixels dispostos nessa orientação específica.

Matriz de coocorrência de níveis de cinza (*Gray Level Co-occurrence Matrices – GLCM*) é um método estatístico para descrição de texturas. A GLCM armazena a probabilidade de dois valores de intensidade de cinza estarem envolvidos por uma determinada relação espacial. A partir desta matriz de probabilidades, diferentes medidas estatísticas podem ser extraídas a fim de caracterizar a textura presente na imagem. Com o objetivo de descrever as propriedades contidas nas texturas, Haralick et al. (1973) propuseram 14 medidas estatísticas calculadas a partir das matrizes de coocorrência, sendo que 6 delas apresentam maior relevância (Quadro 1). Nas equações, $t_{i,j}$ denota o número de transições na i-ésima linha e j-ésima coluna e μ_j o valor médio das distribuições marginais para $\mu_i = \sum_{i,j}^{H_g} i t_{i,j}$ e $\mu_j = \sum_{i,j}^{H_g} j t_{i,j}$.

Quadro 1: Medidas estatísticas calculadas a partir das matrizes de coocorrência

Medida estatística	Equação
Momento angular	$f_{ma} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} t_{i,j}^2$
Entropia	$f_{ent} = -\sum_{i=0}^{H_g} \sum_{j=0}^{H_g} t_{i,j} \log(t_{i,j})$
Contraste	$f_{con} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i - \mu_i)^2 \, t_{i,j}$
Variância	$f_{var_i} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i - \mu_i)^2 \, t_{i,j}$ $\quad$ $f_{var_j} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (j - \mu_j)^2 \, t$
Correlação	$f_{corr} = \frac{1}{\sigma_x \sigma_y} \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} (i - \mu_i)(j - \mu_j) t_{i,j}$
Homogeneidade	$f_{hom} = \sum_{i=0}^{H_g} \sum_{j=0}^{H_g} \frac{1}{1 + (i-j)^2} t_{i,j}$

A principal diferença entre o cálculo das GLCMs (para imagens níveis de cinza) e as CLCMs (para imagens coloridas) é que na primeira as matrizes são calculadas individualmente para cada componente no espaço de cores e, na segunda, elas são geradas baseadas na relação entre as componentes. Dessa forma, utilizando o parâmetro $d = 1$ e três combinações de componentes, uma imagem gera 39 matrizes CLCM (3 combinações de componentes * 13 CLCMs), além das medidas estatísticas apresentadas anteriormente.

2.1.1.3. LOCAL BINARY PATTERN (LBP)

O descritor LBP original, proposto por Ojala et al. (1996), vem sendo muito utilizado em várias aplicações em visão computacional, reconhecimento de padrões e processamento de imagens. É um descritor local de textura baseado na suposição de que a informação de uma textura é dividida em dois aspectos complementares: padrão e intensidade.

O LBP é formado pela comparação sequencial da intensidade dos pixels vizinhos com a do pixel central em uma janela de dimensão 3×3 (versão original), que pode ser estendida para 5×5, 7×7, etc. Para cada pixel $q(l, c)$ de uma janela é feita a comparação com o pixel central $q_c(l, c)$ e o valor 1 é atribuído para o pixel se ele for maior ou igual ao pixel central ou 0 caso contrário (Equação 3).

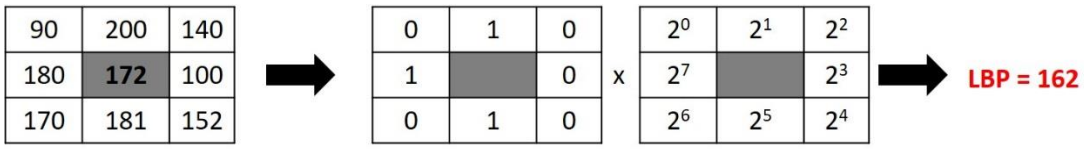
$$vl_i = \begin{cases} 0, & se\ q(l,c) < q_c(l,c), \\ 1, & se\ q(l,c) \geq q_c(l,c) \end{cases} \quad (3)$$

A sequência de bits é lida sequencialmente e mapeada para um número decimal (usando a base 2) como o valor da feição atribuído ao pixel central. Esses valores de feição agregados caracterizam a textura local na imagem. Assim, o LBP para o pixel central $q_c(l, c)$ dentro de uma janela 3×3 pode ser representado de acordo com a Equação 4.

$$LBP\ (q_c(l,c)\ ) = \sum_{i=1}^{8} 2^{i-1} vl_i \quad (4)$$

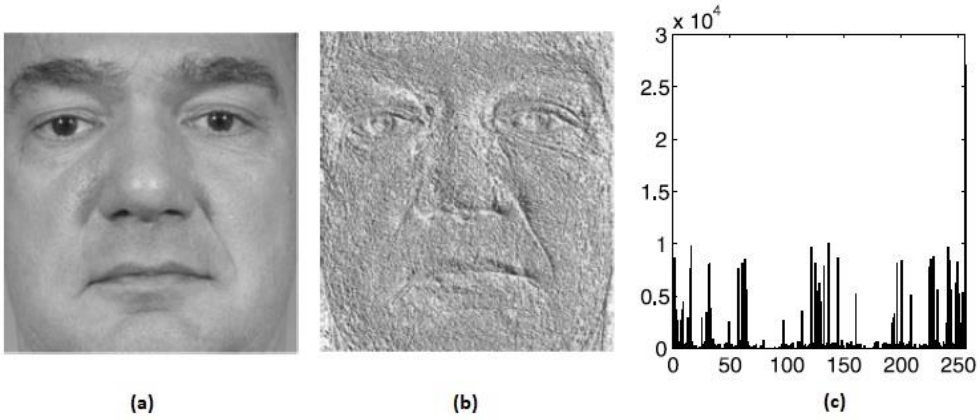
Na Figura 6 pode-se observar um exemplo do cálculo do valor LBP para o pixel q_c assinalado na primeira matriz. Aplicando-se as Equações 3 e 4, obtém-se o valor 162. Esse procedimento deve ser feito para toda a imagem.

Figura 6: Cálculo do valor de LBP para o pixel central



Por fim, um histograma dos rótulos de 256 partições é computado e é usado como um descritor de textura. Cada partição do histograma codifica primitivas locais, as quais incluem diferentes tipos de bordas, curvas, manchas, áreas planas, etc. Na Figura 7a é ilustrada a imagem original que, com a aplicação do descritor LBP, gerou a imagem da Figura 7b na qual pode-se observar que as feições faciais estão bem realçadas. Na Figura 7c é ilustrado o histograma da imagem LBP, que é utilizado como o descritor da imagem.

Figura 7: (a) imagem de entrada; (b) imagem LBP; (c) histograma da imagem



Para gerar o vetor de características para esse descritor, é necessário definir o tamanho das células. Dessa forma, a imagem é particionada em células não sobrepostas. Para coletar informações em regiões maiores, deve-se configurar tamanhos de célula maiores. No entanto, quando se aumenta o tamanho da célula, perde-se os detalhes locais.

2.1.1.4. HISTOGRAM OF ORIENTED GRADIENTS (HOG)

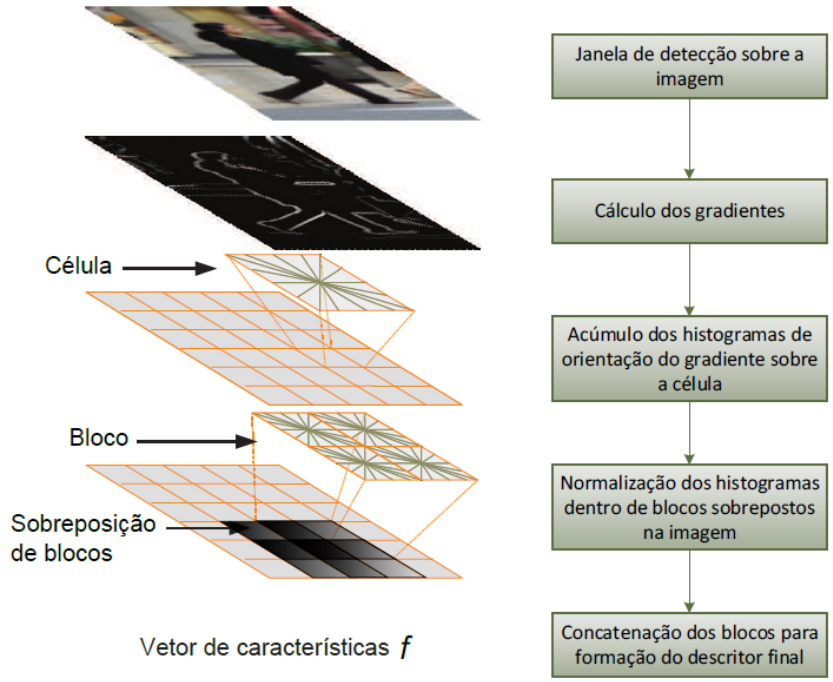
A técnica de extração de características HOG foi proposta por Dalal e Triggs (2005). Histogramas de gradientes orientados têm como base o trabalho de Lowe (2004), denominado SIFT (*Scale Invariant Feature Transform*). Porém, ao contrário do SIFT, que calcula os histogramas de gradientes em volta de *keypoints* invariáveis a escala, o descritor HOG calcula os histogramas de gradientes em uma densa e sobreposta grade de células uniformemente espaçadas na imagem.

A ideia básica por traz da utilização do descritor HOG é a de que o contorno e aparência locais de objetos pode muitas vezes ser bem caracterizada pela distribuição local das orientações do gradiente da imagem, mesmo sem o conhecimento exato da posição do gradiente.

Na prática, intuitivamente, o descritor tenta capturar a forma das estruturas na região, adquirindo informações sobre gradientes. Ele é implementado dividindo a imagem em pequenas regiões espaciais, denominadas células (geralmente 8x8 pixels), e acumulando, para cada célula, um histograma 1-D de orientações do gradiente de cada pixel presente na célula. A combinação dos histogramas forma a representação dos contornos e aparências locais.

Para diminuir a variância à iluminação e sombreamento, é útil que se faça uma normalização local dos histogramas. Essa normalização é feita acumulando uma medida da energia local dos histogramas sobre regiões maiores, denominadas blocos (geralmente blocos de células 4×4), e usando essa medida para normalizar as células dentro de um determinado bloco. O descritor HOG é formado pela concatenação dos diversos blocos presentes em uma janela de detecção percorrida sobre a imagem (COSMO, 2014). Na Figura 8 é mostrado o funcionamento do HOG para a extração do vetor de características.

Figura 8: Funcionamento do HOG para obtenção do vetor de características de uma imagem

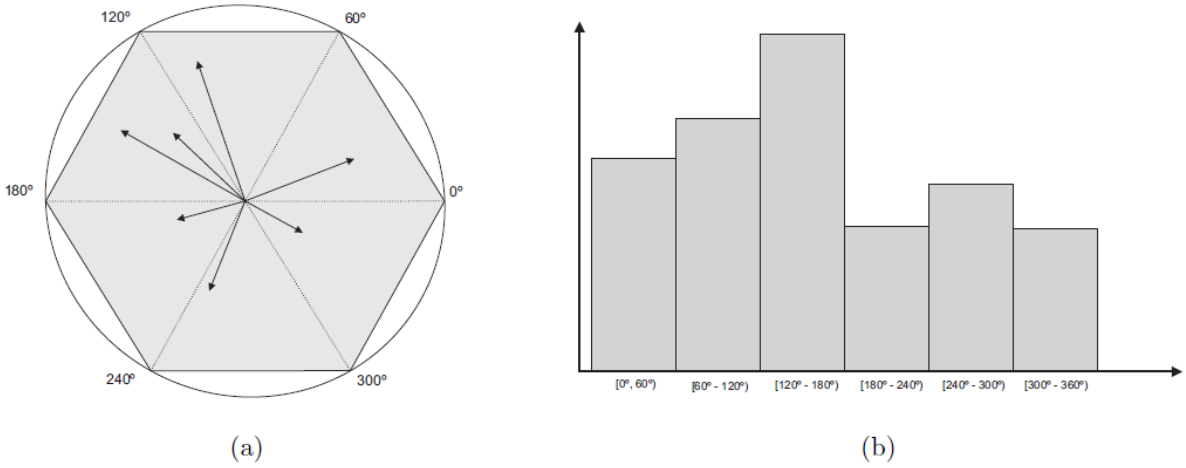


Fonte: COSMO, 2014

Na Figura 9b é ilustrado um exemplo de um histograma de gradientes bidimensional subdivido em seis intervalos de orientação, os quais estão definidos na Figura 9a no sentido anti-horário. Cada intervalo guarda a soma das magnitudes de todos os vetores pertencentes ao mesmo. Por exemplo, a frequência em [120 graus; 180 graus) e a soma das magnitudes dos dois vetores desse intervalo. De fato, um histograma bidimensional pode ser visto como uma aproximação de um círculo por um polígono, onde cada lado do polígono corresponde a um intervalo de classe do histograma. Isso pode ser estendido para o caso tridimensional aproximando-se uma esfera por poliedros.

As principais vantagens do descritor HOG são: a captura de informações de contorno locais através da codificação da orientação do gradiente em histogramas, redução da variância espacial através do acúmulo local desses histogramas sobre regiões da imagem e uma redução da variância à iluminação através da normalização local desses histogramas.

Figura 9: Histograma de gradientes bidimensional subdivido em seis intervalos



(a) (b)

2.1.1.5. BAG OF VISUAL WORDS (BOVW)

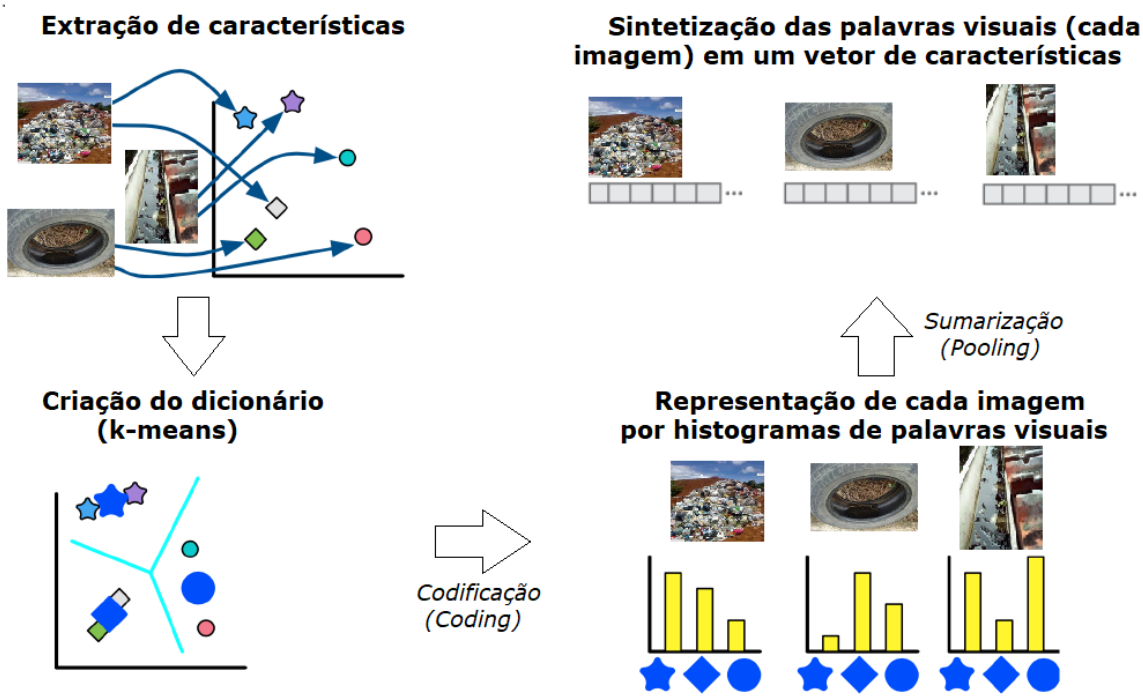
A técnica BoVW, também conhecida como *Bag of Visual Features* (BoVF), é utilizada para combinar, de maneira robusta, as características locais extraídas de uma imagem em um único vetor de características. É uma técnica bastante popular devido a sua simplicidade, pois é baseada na representação não-ordenada de descritores locais aplicados em uma imagem e são, portanto, conceitualmente e computacionalmente mais simples do que muitos métodos alternativos.

Na Figura 10 pode-se observar as seguintes etapas da técnica BoVW: (i) extração das características das imagens; (ii) geração de um dicionário de palavras visuais por meio de um algoritmo de agrupamento (por exemplo, o *k-means*); (iii) contagem da ocorrência (frequência) de cada palavra visual contida na imagem para a criação de histogramas; (iv) sumarização das informações, que consiste em sintetizar as palavras visuais da imagem em um vetor de características.

Para combinar características locais de uma imagem a fim de convertê-las em um único vetor de características, é necessária a criação de um dicionário visual no qual cada característica local extraída da imagem é tratada como uma palavra visual pertencente a ele. Por meio do mapeamento dessas características locais a sua respectiva palavra visual no dicionário, pode-se descrever a imagem em um único vetor de características. Tal procedimento é conhecido como técnica de *pooling* e o

mais tradicional na literatura é o histograma de palavras visuais, o qual consiste em uma contagem simples da frequência com que cada palavra visual aparece na imagem.

Figura 10: Etapas da técnica BoVW



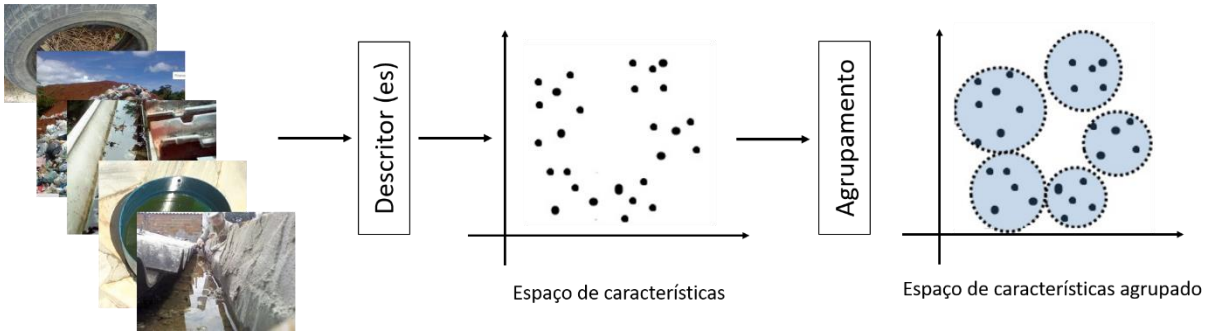
Fonte: adaptado de Mehra et al. (2016)

O dicionário de palavras visuais é o responsável por determinar quais são as características e padrões que representam a estrutura de uma imagem. Vale lembrar que não se pode generalizar um dicionário universal, válido para todos os tipos de aplicações, pois diferentes domínios necessitam de dicionários específicos, uma vez que a representação pode mudar. Nesse caso, o desafio é como determinar qual é o dicionário ideal para cada tipo de aplicação de acordo com o tipo de representação utilizada para descrever as características.

A maioria dos trabalhos na literatura apresentam a mesma estratégia para a geração do dicionário de palavras visuais. Na Figura 11 é mostrada tal estratégia, a qual é definida pelos seguintes passos: (i) um subconjunto de imagens do banco de imagens é escolhido; (ii) para cada imagem, suas regiões de interesse são detectadas

e descritas utilizando um descritor (ou vários descritores), gerando vetores de características; (iii) é realizado um agrupamento dos dados desse espaço de características utilizando algum algoritmo de agrupamento, sendo o centroide de cada grupo considerado como uma palavra visual do dicionário.

Figura 11: Estratégia para a criação do dicionário de palavras visuais



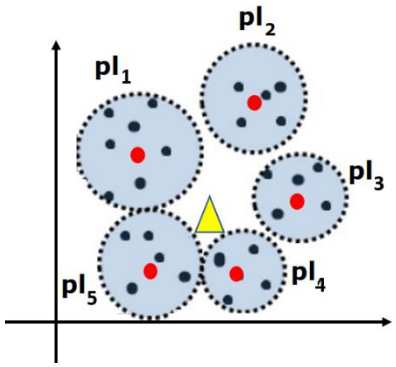
O algoritmo de agrupamento é um fator que interfere diretamente na qualidade e no desempenho da geração de um dicionário visual. Diferentes algoritmos podem levar a diferentes dicionários (JAIN et al., 1999), gerando resultados mais próximos das expectativas do usuário. Além disso, a quantidade de palavras visuais do dicionário é outro fator que influencia diretamente na qualidade da representação da imagem. Um dicionário pequeno tem pouco poder discriminativo, uma vez que dois agrupamentos podem ser atribuídos a mesma palavra visual. Em contrapartida, a quantidade de palavras visuais é uma informação definida empiricamente e que pode variar entre diferentes bases de dados.

O agrupamento pode ser realizado, por exemplo, por meio do algoritmo *k-means*, o qual consiste no particionamento de um conjunto de pontos entre k subconjuntos disjuntos visando a minimizar a distância intra-grupos e maximizar a distância inter-grupos. A quantidade de palavras visuais do dicionário, na maioria das vezes escolhida empiricamente, é representada pelos *k-clusters* obtidos por meio da execução do *k-means*.

Uma vez definido o dicionário de palavras visuais, cada imagem, representada por um vetor de características, será associada a uma palavra visual. Essa fase é denominada *assignment* ou *coding*. Na Figura 12 é ilustrada a fase de *coding*, sendo

uma palavra visual o resultado de um agrupamento do espaço de características. Uma maneira de realizar a fase de codificação é utilizar a estratégia *multiple assignment*, que foi concebida para mitigar o impacto negativo quando um grande número de vetores de características de uma imagem está perto de uma fronteira de dois ou mais agrupamentos (triângulo amarelo ilustrado na Figura 12). A estratégia *multiple assignment* (JEGOU et al., 2007) atribui ao vetor de características todas as palavras visuais mais próximas. Nesse caso, adiciona-se uma unidade ao bin correspondente a cada palavra visual mais próxima. Supondo que existam 3 palavras visuais mais próximas, o triângulo amarelo seria atribuído às palavras pl_3, pl_4 e pl_5.

Figura 12: Fase de codificação (coding) da técnica BoVW



A fase de sumarização (*pooling*) é a etapa responsável por sintetizar, em um único vetor de características, a representação final da imagem. Basicamente, existem três técnicas de pooling mais tradicionais: *sum-pooling*, *average-pooling* e *max-pooling*. Todas elas fornecem uma representação de tamanho fixo e são baseadas na contagem da ocorrência não-ordenada das palavras visuais no espaço da imagem. As técnicas *sum-pooling* e *average-pooling* são semelhantes, ou seja, cada uma gera um histograma de palavras visuais, porém a diferença entre elas está no fato de que a primeira representa a soma das ocorrências de cada palavra visual na imagem, enquanto a segunda se refere à média, gerando um histograma normalizado. Já a técnica de *max-pooling* gera um vetor binário que indica a presença ou ausência de uma palavra visual na imagem.

Vale dizer que as operações de *coding* e *pooling* permitem a redução de características obtido a partir de um descritor de imagens ou de uma combinação de

descritores. Após a etapa de sumarização e de posse dos vetores de características das imagens, é possível utilizá-los como conjunto de treinamento para um algoritmo de classificação como, por exemplo, o *Support Vector Machine* (SVM).

2.2. APRENDIZAGEM DE MÁQUINA

As dificuldades enfrentadas pelos sistemas que dependem do conhecimento codificado sugerem que os sistemas de inteligência artificial precisam ter a capacidade de adquirir seus próprios conhecimentos, extraindo padrões de dados brutos. Esse recurso é conhecido como aprendizagem de máquina, que é um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, reconhecer padrões e tomar decisões com o mínimo de intervenção humana (GOODFELLOW, BENGIO e COURVILLE, 2016). Segundo Mitchell (1997), o campo da aprendizagem de máquina está preocupado com a questão de como construir programas de computador que melhoram automaticamente com a experiência.

O reconhecimento de padrões estuda uma maneira de as máquinas também poderem observar o ambiente, aprender a distinguir padrões de interesse e tomar decisões sobre as categorias de padrões (JAIN et al., 2000). No reconhecimento de padrões em imagens digitais, algumas das características mais empregadas são: a cor e a textura. A boa quantificação destas características permitirá a identificação e a classificação de padrões.

Para serem reconhecidos e classificados por um sistema automático, os padrões devem ser descritos por um conjunto de características mensuráveis, as quais são extraídas de um objeto ou entidade de interesse em uma imagem, por exemplo. Quando essas características são similares dentro de um grupo de padrões, diz-se que esses padrões pertencem a uma mesma classe. O objetivo dos sistemas de reconhecimento é determinar, com base nas informações disponíveis, a classe de padrões responsável pela produção de um conjunto de medidas similar ao do padrão sob análise. No entanto, o reconhecimento correto depende da quantidade de informação discriminante contida nas características extraídas e da utilização efetiva dessas informações (TOU e GONZALEZ, 1974).

O reconhecimento/classificação de padrões pode consistir em uma das duas tarefas (COSTA e CESAR, 2000; JAIN et. al., 2000):

- classificação supervisionada (análise discriminante): o padrão de entrada é identificado como um membro de uma classe pré-definida. Um ou mais exemplos de cada classe previamente conhecida são utilizados como modelo (protótipos) para a classificação de novos objetos. Esse tipo de classificação frequentemente envolve dois estágios: aprendizado, quando os critérios e métodos são testados nos modelos, e reconhecimento, quando o sistema treinado é usado para classificar novos objetos;
- classificação não-supervisionada (*clustering* ou agrupamento): o padrão é associado a uma classe até então desconhecida. Esse é o caso em que se tem um conjunto de objetos e tenta-se encontrar a classe mais adequada, sem modelo específico ou características e critérios disponíveis. Nesse caso, a busca por um critério de classificação e características apropriadas caracteriza-se como um processo de descoberta, através do qual são criados novos conceitos e identificados relacionamentos entre os objetos.

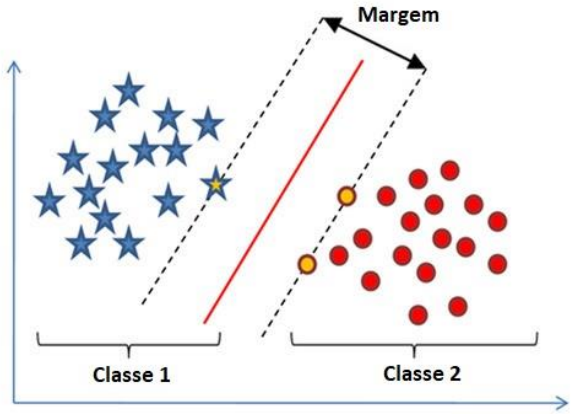
Neste trabalho são utilizadas algumas técnicas de aprendizagem de máquina para reconhecimento e interpretação de padrões em imagens, as quais são descritas a seguir.

2.2.1. SUPPORT VECTOR MACHINE (SVM)

Máquina de Vetores de Suporte, ou *Support Vector Machine* (SVM), é um método de aprendizagem supervisionada utilizado, principalmente, para classificação e regressão. SVM foi introduzido por Cortes e Vapnik (1995) e é empregado para estimar uma função que classifique dados em duas classes. O conceito básico das SVMs compreende a construção de um hiperplano como superfície de decisão de forma que seja máxima a margem de separação entre as classes. O objetivo do treinamento por meio das SVMs é a obtenção de hiperplanos que dividam a amostra de tal maneira que sejam otimizados os limites de generalização, sendo os pontos localizados próximos destes limites chamados de vetores de suporte.

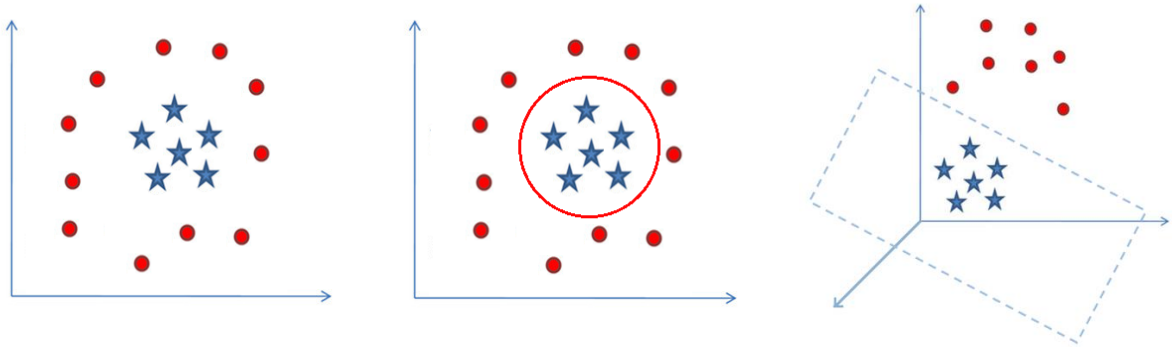
Na Figura 13 é ilustrado um exemplo de classificação de dados, no qual o método busca determinar os parâmetros de uma linha (em outras palavras, um separador linear ótimo) a fim de maximizar a distância dos vetores de suporte é sempre maior, ou seja, os melhores vetores de suporte que maximizam a margem. Entretanto, há situações onde isso não é possível, e neste caso, permite-se que a SVM classifique alguns dados incorretamente, até um certo grau.

Figura 13: Exemplo de classificação usando o SVM



Há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano. Um exemplo é apresentado na Figura 14, em que o uso de uma fronteira curva seria mais adequado na separação das classes. Na Figura 14a pode-se observar um caso onde os dados estão dispostos de forma não-linear. Com o intuito de separá-los em duas classes distintas, foi definida uma fronteira curva (círculo vermelho na Figura 14b) que, na sequência, permitiu a divisão linear por meio da definição de um espaço de características (Figura 14c). As SVMs lidam com problemas não lineares mapeando o conjunto de treinamento de seu espaço original, referenciado como de entradas, para um novo espaço de maior dimensão, denominado espaço de características.

Figura 14: (a) Conjunto de dados não linear; (b) Fronteira não linear no espaço de entradas; (c) Fronteira linear no espaço de características



(a) **(b)** **(c)**

As SVMs possuem diferentes *kernels* que são utilizados na resolução de problemas de espaços não-lineares. Os mais utilizados são: Polinomial (que manipula uma função polinomial cujo grau pode ser definido durante os treinamentos); Sigmoidal (permite que a SVM se comporte de maneira similar à rede MLP); e Gaussiano (a SVM se comporta como uma rede RBF).

Existem alguns parâmetros de regularização usados na SVM. O fator de penalidade do modelo é um parâmetro que evita uma classificação errada ou um *overfiting*. Para um valor elevado deste fator, a otimização deverá escolher um hiperplano de separação de margem pequena e ao contrário, um hiperplano de separação de margem grande. O parâmetro $Gamma$ indica quais pontos serão considerados em relação à fronteira de separação, ou seja, com um valor baixo, pontos distantes da fronteira são considerados e com um valor alto o oposto é verdadeiro, ou seja, pontos mais próximos serão avaliados.

Para problemas que possuem mais de duas classes (multiclasse), o conjunto de dados de treinamento deve ser combinado para formar problemas de duas classes (BISOGNIN, 2007). A seguir são descritas as duas principais estratégias para esta finalidade: Um Contra Um (em inglês, *One Against One*, OAO) e Um Contra Todos (em inglês, *One Against All*, OAA).

A estratégia Um contra Um (Um x Um) é um método simples e eficiente para a resolução de problemas Multiclasses. Supondo um problema com ncl classes, para

cada par dessas ncl classes é construído um classificador binário. Cada classificador é construído utilizando elementos das duas classes envolvidas, obtendo um total de $ncl\ (ncl\ -\ 1)\ /\ 2$ classificadores. Portanto os conjuntos de treinamento devem ser rotulados novamente para cada par de entradas.

Supondo que um problema tem ncl classes, o método Um contra Todos (Um x Todos) consiste em particionar estas ncl classes em dois grupos. Um grupo é formado por uma classe e o outro é formado pelas classes restantes. Um classificador binário é treinado para esses dois grupos e este procedimento é repetido para cada uma das ncl classes. Uma vantagem desse método é o número reduzido de classificadores, comparado ao método Um Contra Um, o que torna a classificação mais rápida em casos de poucas classes. Uma desvantagem é que cada classificador utiliza todas as classes, sendo assim o desempenho depende do número de classes.

Neste trabalho o classificador SVM multiclasse com a estratégia Um x Todos foi empregada na detecção de cenários nas imagens.

2.2.2. REDES NEURAIS ARTIFICIAIS (RNA)

As Redes Neurais Artificiais (RNAs) têm sido bastante estudadas desde a sua redescoberta como um paradigma de reconhecimento de padrões no final dos anos 80. Elas representam uma ferramenta de grande valor em várias aplicações que são vistas como difíceis, particularmente, em reconhecimento de padrões visuais e de fala (HAYKIN, 2003).

As RNAs baseiam-se na combinação de processadores simples (neurônios), cada qual com um número de entradas e gerando uma única saída. Os neurônios calculam determinadas funções matemáticas, normalmente não-lineares, as quais podem ser discretas ou contínuas dependendo do tipo de rede em uso. Além disso, os neurônios são dispostos em uma ou mais camadas interligadas por um grande número de conexões, geralmente unidirecionais (a saída de um neurônio pode conectar-se à entrada de um ou mais neurônios). Na maioria dos modelos essas conexões estão associadas a pesos, os quais armazenam o conhecimento

representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O funcionamento de toda essa estrutura foi inspirado no cérebro humano com o intuito de imitá-lo (SONKA, HLAVAC e BOYLE, 1999).

As RNAs apresentam soluções bastante atrativas, pois devido à representação interna e ao paralelismo natural inerente à arquitetura é possível alcançar com elas um desempenho superior a alguns modelos convencionais (BRAGA et. al., 2000). A ideia central das redes neurais é que tais parâmetros possam ser ajustados para que elas possam apresentar o comportamento desejado. Dessa maneira, a rede pode ser treinada para resolver um problema particular ajustando os parâmetros (pesos ou bias).

Inicialmente, as RNAs passam por uma fase de aprendizagem (procedimento usual), na qual um conjunto de exemplos é apresentado para a rede. Nessa etapa são extraídas automaticamente as características necessárias para representar a informação fornecida, as quais serão utilizadas no momento de gerar as respostas a um determinado problema. Um simples neurônio, derivado do trabalho pioneiro em simulação neural conduzido por McCulloch e Pitts (1943). O modelo por eles proposto (modelo MCP) é uma simplificação do que se sabia até aquele momento sobre um neurônio biológico. Nesse modelo, as entradas são denotadas por $p_1, p_2, \dots, p_R$ e os pesos associados à elas por $W_1, W_2, \dots, W_R$, sendo R o número de entradas. Os neurônios também possuem um escalar bias b, que é adicionado ao produto $W.p$. O bias é como um peso, exceto pelo fato de que ele tem uma entrada constante de $b = 1$.

O neurônio biológico dispara uma saída quando a soma dos impulsos que ele recebe excede o seu limiar de excitação. Esse mecanismo é emulado em RNAs de um simples modo: realiza-se a soma dos valores $W_i\, p_i$ recebidos (Equação 5) e depois compara-se o valor dessa soma a um limiar θ, definido para o neurônio, para saber se ativa ou não a saída.

$$\sum_{i=1}^{R} W_i\, p_i + b \qquad (5)$$

No modelo MCP, a ativação do neurônio é obtida por uma função de ativação $f(x)$, que é associada ao neurônio com o objetivo de decidir se ativa ou não a saída com base no resultado da soma ponderada. Os valores produzidos como saída são 0 ou 1, mas podem ser outros valores dependendo da função de ativação selecionada. Diferentes funções de ativação foram derivadas a partir do modelo proposto por McCulloch e Pitts (1943), dentre elas a função passo, a função linear e a função sigmoidal (BRAGA et. al., 2000; JAIN et. al., 1996). A função passo, ilustrada pela Figura 15a, produz saídas iguais a 0 ou 1, conforme definido na Equação 6.

$$f(x) = \begin{cases} 0, se\ x \leq 0 \\ 1, se\ x > 0 \end{cases} \quad (6)$$

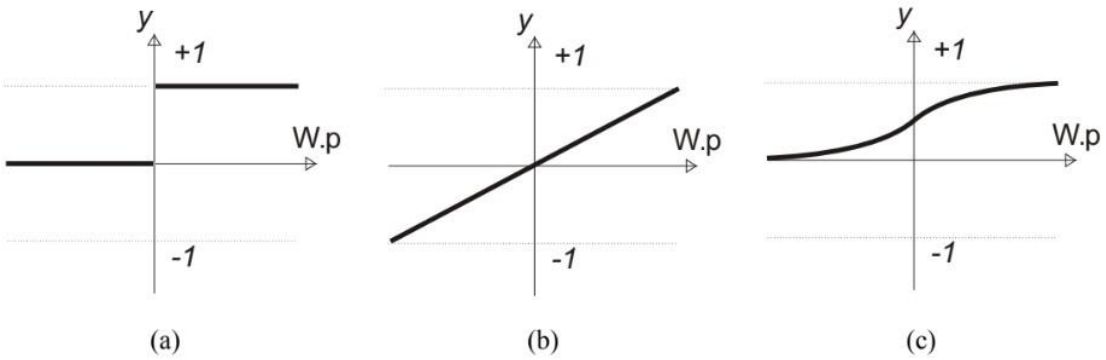
A função linear, é definida pela Equação 7, sendo α um número real que define a saída linear $y = f(x)$ para os valores de entrada x. Sua representação é mostrada na Figura 15b.

$$f(x) = \alpha x \quad (7)$$

A função sigmoidal, mostrada na Figura 15c, é uma função semilinear, limitada e monotônica (BRAGA et. al., 2000). Uma das funções mais importantes é a função logística definida na Equação 8.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Figura 15: Representações de algumas funções de ativação: (a) função passo, (b) função linear e (c) função sigmoidal.



Vale ressaltar que o modelo MCP é limitado à resolução de problemas nos quais os dados são linearmente separáveis.

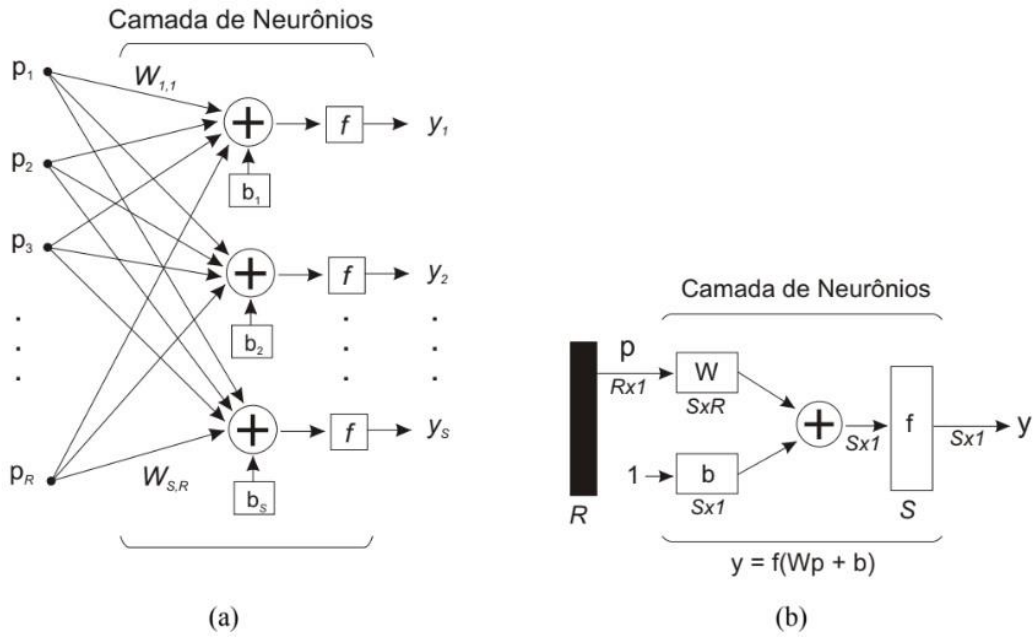
2.2.2.1. REDES NEURAIS MULTILAYER PERCEPTRONS (MLP)

Uma característica importante das redes MLP é que elas podem tratar com dados que não são linearmente separáveis. Para tanto, o modelo MLP contempla, além da camada de entrada e de saída (como ocorre no modelo MCP), uma ou mais camadas ocultas (intermediárias) que possibilitam a rede mapear padrões de entrada com estruturas similares, para saídas diferentes (HAYKIN, 2003). Em uma rede desse tipo, o processamento realizado pelos neurônios é definido pela combinação dos processamentos realizados pelos neurônios da camada anterior (HAYKIN, 2003).

Na Figura 16a é exemplificado o funcionamento de uma camada de neurônios de uma rede MLP. Nessa camada o vetor de entradas, representado por p é conectado a cada um dos neurônios por meio da matriz de pesos W, sendo $W_{1,1}$ a conexão entre o primeiro neurônio da camada e a primeira entrada da rede e $W_{S,R}$ a conexão entre o último neurônio e a última entrada. O número de entradas, representado por R, não necessita ser igual ao número de neurônios, representado por S. O bias é representado por b e a função de ativação por f. Uma representação abreviada pode ser usada para ilustrar a camada de uma rede, conforme mostrada na

Figura 16b. Nessa figura, as dimensões de cada componente são mostradas abaixo delas, por exemplo a matriz de pesos W que possui dimensão $S \times R$ (DEMUTH e BEALE, 2003).

Figura 16: Representações de uma camada da rede neural

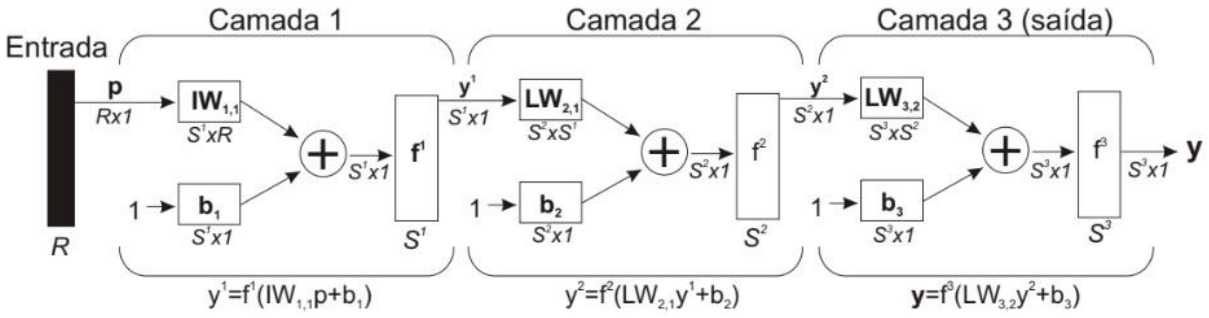


A representação de uma rede com três camadas é mostrada na Figura 17, considerando a representação abreviada. A terceira camada é considerada a camada de saída. Alguns autores referem-se à entrada como uma quarta camada, mas essa designação não será aplicada aqui. Nessa figura, a matriz de pesos da entrada é representada por IW (*input weights*), enquanto as demais matrizes de pesos são representadas por LW (*layer weights*). As saídas y^n das camadas intermediárias são as entradas para as camadas seguintes.

Para uma rede com pelo menos uma camada intermediária, pode-se dizer que o processamento ocorre da seguinte maneira (BRAGA et. al., 2000): (i) na primeira camada cada neurônio traça retas no espaço de padrões de treinamento, (ii) na segunda camada, cada neurônio combina as retas traçadas pelos neurônios da camada anterior conectados a ele, formando regiões convexas e (iii) na terceira

camada (saída), cada neurônio forma regiões que são combinações das regiões convexas definidas pelos neurônios da camada anterior conectados a ele.

Figura 17: Representação abreviada da rede com 3 camadas



Fonte: (Demuth e Beale, 2003)

A definição do número de neurônios em cada uma das camadas intermediárias é empírica, dependendo fortemente da distribuição dos padrões de treinamento e da validade da rede.

Dentre os algoritmos de treinamento, o mais popular é o algoritmo *backpropagation* (RUMELHART e MCCLELLAND, 1986), o qual foi um dos principais responsáveis pelo ressurgimento do interesse por redes neurais artificiais. O algoritmo *backpropagation* é um algoritmo de aprendizado supervisionado, que utiliza um mecanismo de correção de erros para ajustar os pesos considerando os pares (entrada x saída desejada) (BRAGA et. al., 2000). Esse algoritmo trabalha em duas fases: (i) a fase *forward*, que define a saída da rede para uma determinada entrada, e (ii) a fase *backward*, que utiliza a saída obtida pela rede e a saída desejada para ajustar os pesos entre as camadas. Algumas variações foram propostas visando a acelerar o tempo de treinamento, como a *backpropagation* com momentum (RUMELHART e MCCLELLAND, 1986), a *Levenberg-Marquardt* (HAGAN e MENHAJ, 1994) dentre outras.

A capacidade das RNAs não se resume apenas em mapear as relações de entrada e saída. Sua capacidade de aprender por meio de exemplos e de generalizar a informação aprendida é o principal atrativo para sua aplicação. As famílias de redes neurais comumente usadas em reconhecimento de padrões são as redes *feed-*

forward. Outra família popular de rede é a *Self-Organization Map* (SOM), ou rede de *Kohonen*, de aprendizado não-supervisionado, que é usada principalmente para agrupamento de dados e mapeamento de características (JAIN et. al., 2000, 1996). Aplicações de redes neurais em processamento de imagens, especialmente redes *feed-forward*, *Kohonen* e *Hopfield* são discutidas por Egmont-Petersen et. al. (2002).

Neste trabalho foi utilizada uma RNA do tipo MLP nos experimentos para reconstituição de bandas espectrais, na etapa de detecção de pequenas porções de água.

2.2.2.2. REDES NEURAIS CONVOLUCIONAIS

Aprendizagem profunda (Deep Learning) é o termo usado para denotar o problema de treinar redes neurais artificiais que realizam o aprendizado de características de forma hierárquica, de tal forma que características nos níveis mais altos da hierarquia sejam formadas pela combinação de características de mais baixo nível (BEZERRA, 2016). Técnicas de aprendizado profundo oferecem atualmente um importante conjunto de métodos para analisar sinais como: áudio e fala, conteúdos visuais, incluindo imagens e vídeos, e ainda conteúdo textual. Entretanto, esses métodos incluem diversos modelos, componentes e algoritmos.

Métodos que utilizam *Deep Learning* buscam descobrir um modelo (por exemplo, regras, parâmetros) utilizando um conjunto de dados (exemplos) e um método para guiar o aprendizado do modelo a partir desses exemplos. Ao final do processo de aprendizado tem-se uma função capaz de receber por entrada os dados brutos e fornecer como saída uma representação adequada para o problema em questão (PONTI e COSTA, 2017).

Redes Neurais Convolucionais (RNCs) são provavelmente o modelo de rede *Deep Learning* mais conhecido e utilizado atualmente. O que caracteriza esse tipo de rede é ser composta basicamente de camadas convolucionais, que processa as entradas considerando campos receptivos locais. Adicionalmente, inclui operações conhecidas como *pooling*, responsáveis por reduzir a dimensionalidade espacial das representações. A principal aplicação das RNCs é para o processamento de

informações visuais, em particular imagens, pois a convolução permite filtrar as imagens considerando sua estrutura bidimensional (espacial).

Em vez da conectividade global, uma rede convolucional utiliza conectividade local. Por exemplo, considere a primeira camada intermediária de uma RNC. Cada unidade desta camada intermediária está conectada a uma quantidade restrita de unidades localizada em uma região contígua da camada de entrada, em vez de estar conectada a todas as unidades. As unidades da camada de entrada conectadas a uma unidade da camada intermediária formam o campo receptivo local dessa unidade. Por meio de seu campo receptivo local, cada unidade camada intermediária pode detectar características visuais elementares, tais como arestas orientadas, extremidades, cantos. Essas características podem então ser combinadas pelas camadas subsequentes para detecção de características mais complexas (por exemplo, olhos, bicos, rodas, etc.).

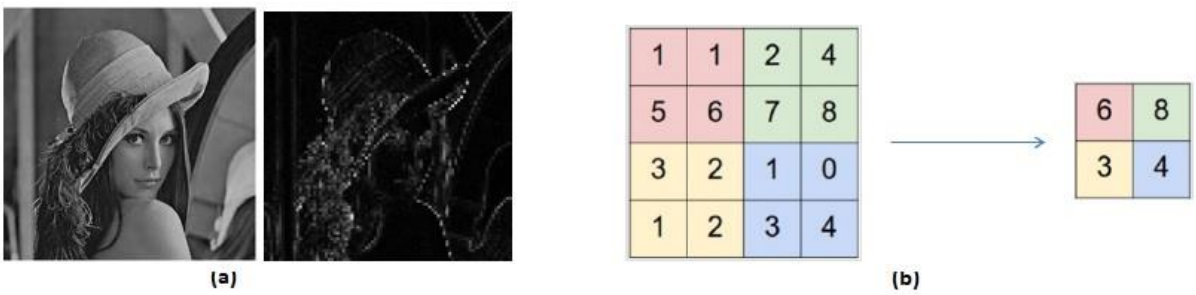
Por outro lado, é provável que um determinado detector de alguma característica elementar seja útil em diferentes regiões da imagem de entrada. Para levar isso em consideração, em uma RNC as unidades de uma determinada camada são organizadas em conjuntos disjuntos, cada um dos quais é denominado um mapa de característica (*feature map*), também conhecido como filtro. As unidades contidas em um mapa de características são únicas na medida em que cada uma delas está ligada a um conjunto de unidades (isto é, ao seu campo receptivo) diferente na camada anterior. Além disso, todas as unidades de um mapa compartilham os mesmos parâmetros. O resultado disso é que essas unidades dentro de um mapa servem como detectores de uma mesma característica, mas cada uma delas está conectada a uma região diferente da imagem. Portanto, em uma RNC, uma camada oculta é segmentada em diversos mapas de características (BEZERRA, 2016).

Cada unidade em um mapa de característica realiza uma operação denominada convolução. Outra operação importante utilizada em uma RNC é a subamostragem. Em processamento de imagens, a subamostragem de uma imagem envolve reduzir a sua resolução, sem, no entanto, alterar significativamente o seu aspecto. No contexto de uma RNC, a subamostragem reduz a dimensionalidade de um mapa de característica fornecido como entrada e produz outro mapa de característica, uma espécie de resumo do primeiro.

Há várias formas de subamostragem aplicáveis a um mapa de característica: selecionar o valor máximo (*max pooling*), a média (*average pooling*) ou a norma do conjunto, entre outras. O uso da subamostragem faz com que uma RNC seja robusta em relação às localizações exatas das características na imagem, uma propriedade que permite a esse tipo de rede aprender representações invariantes com relação a pequenas translações. Na Figura 18 são mostrados exemplos da operação de convolução e de subamostragem, sendo a Figura 18a ilustrando uma imagem de entrada com sua respectiva imagem de saída resultante da aplicação da operação de convolução, permitindo a subamostragem com filtro (máscara) de tamanho 2×2, com passo igual a 2 (aplicação do filtro a cada dois pixels), ilustrados na Figura 18b.

Em geral, a arquitetura de uma RNC possui diversos tipos de camadas: camadas de convolução, camadas de subamostragem, camadas de normalização de contraste e camadas completamente conectadas. Na forma mais comum de arquitetar uma RNC, a rede é organizada em estágios. Cada estágio é composto por uma ou mais camadas de convolução em sequência, seguidas por uma camada de subamostragem, que por sua vez é seguida (opcionalmente) por uma camada de normalização. Uma RNC pode conter vários estágios empilhados após a camada de entrada (que corresponde à imagem). Após o estágio final da rede, são adicionadas uma ou mais camadas completamente conectadas.

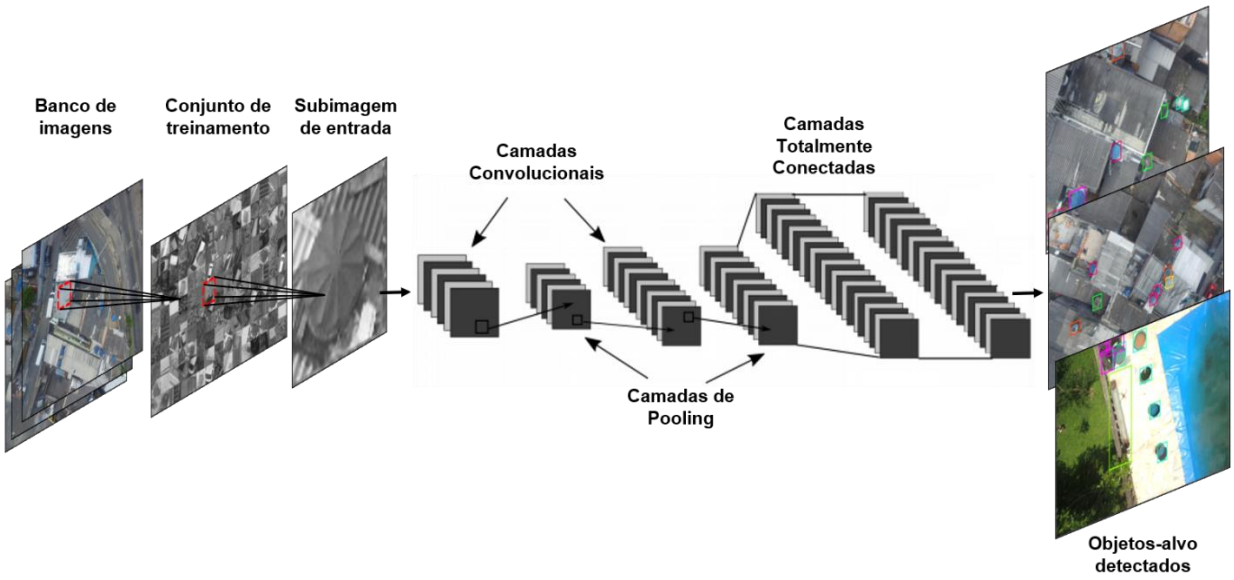
Figura 18: Exemplos das operações de convolução e de subamostragem: (a) Resultado de uma convolução (direita) aplicada a uma imagem (esquerda); (b) Subamostragem com filtro de tamanho 2x2 e tamanho do passo igual a 2.



Fonte: Bezerra (2016)

Na Figura 19 é apresentado um exemplo de arquitetura de uma RNC com camadas convolucionais e duas completamente conectadas. Após a camada de entrada (que corresponde aos pixels da imagem), tem-se as camadas convolucionais compostas por mapas de características (representados como planos na Figura 19), intercaladas por camadas de subamostragens e, por fim, as camadas completamente conectadas. Neste exemplo, a camada de saída é composta por valores que correspondem às classes.

Figura 19: Exemplo de Arquitetura de uma RNC



A computação realizada em cada mapa de característica de uma camada de convolução envolve a aplicação de várias operações de convolução, uma para cada unidade contida no mapa, sendo que cada unidade do mapa de características realiza uma convolução usando a matriz de pesos. A matriz de pesos corresponde ao núcleo dessa convolução, e a matriz sobre a qual a convolução é aplicada correspondente ao campo receptivo local. A quantidade de unidades é a mesma em cada mapa de característica e depende do tamanho do campo receptivo das unidades e do denominado tamanho do passo (*stride size*). Esse segundo hiperparâmetro é um inteiro positivo (normalmente igual a 1) que define o quanto de sobreposição há entre os campos receptivos locais de duas unidades vizinhas em um mapa de características.

Outro tipo de camada em uma RNC é a denominada camada de subamostragem (*subsampling layer* ou *pooling layer*). O objetivo dessa camada, que também é composta por um conjunto de mapas de características, é realizar a agregação das saídas (ativações) de um conjunto de unidades da camada anterior. É possível mostrar que camadas de subamostragem resultam em uma rede mais robusta a transformações espaciais.

Já as camadas completamente conectadas, que geralmente são encontradas antes da camada de saída, geram descritores de características da imagem que podem ser mais facilmente classificados pela camada de saída.

2.2.2.2.1. FRAMEWORK YOLOv3

Nas abordagens tradicionais de visão computacional, uma janela deslizante é usada para procurar objetos em diferentes locais e escalas, sendo essa operação muito cara computacionalmente. Esse tipo de processamento é comum com o uso de RNCs, como no caso da arquitetura LeNet-5, idealizada por LeCun et al. (1998). Os algoritmos de detecção de objetos baseados no *Early Deep Learning*, como o R-CNN (*Region Convolutional Neural Network*) e o *Faster R-CNN* (*Faster Region Convolutional Neural Network*), usam um método chamado Busca seletiva, que visa a redução do número de caixas delimitadoras que o algoritmo tem que testar por meio de agrupamento hierárquico, de regiões semelhantes da imagem, com base na compatibilidade de cor, textura, tamanho e forma.

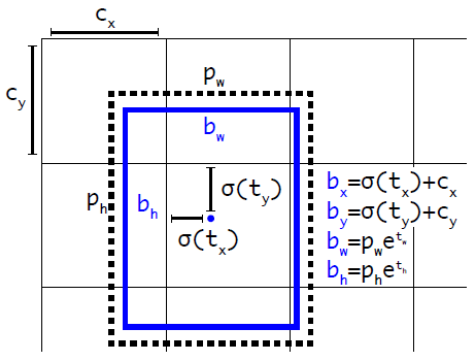
YOLO é um *framework* composto por RNCs projetadas especialmente para detecção de objetos. O *"YOLO - You Only Look Once"*, concebido por Redmon et al. (2016), tem essa denominação, pois refere-se ao fato de que as RNCs implementadas no *framework* processam a imagem inteira uma única vez ao mesmo tempo, gerando as predições dos objetos. As arquiteturas das RNCs que compõem o *framework* têm a capacidade de reconhecer 80 objetos diferentes em imagens e vídeos, em tempo real. O YOLO superou métodos populares e robustos como o *Faster R-CNN* com o RestNet (HE et al., 2016) e o SSD (LIU et al., 2016), apresentando resultados competitivos e sendo mais rápidos.

O YOLO aborda o problema de detecção de objetos de uma maneira completamente diferente dos métodos baseados em janelas deslizantes. A imagem inteira é utilizada apenas uma vez como entrada na rede. Primeiramente, a imagem é dividida em uma grade de células. O tamanho dessas células varia dependendo do tamanho da dimensão da imagem entrada. Por exemplo, para uma imagem de entrada de 416×416 pixels, considerando a dimensão igual a 13, o tamanho de cada célula é de 32×32 pixels.

Há um número fixo de “caixas de ancoragem” para cada célula, o que corresponde a formas de objetos pré-definidas calculadas previamente de acordo com os objetos gerais do conjunto de dados. Dessa forma, seguindo o mesmo exemplo supracitado, com 169 células (13×13), onde cada célula possui cinco âncoras, obtém-se 845 possíveis predições das *bounding boxes* (caixas delimitadoras).

A caixa delimitadora é definida como duas coordenadas relativas à matriz da imagem, correspondentes à posição central do objeto, e as duas dimensões de largura e altura. Em geral, a rede prediz 5 caixas delimitadoras em cada célula no mapa de características de saída. Ela ainda prediz 4 coordenadas para cada caixa delimitadora, t_x, t_y, t_w e t_h, além de um valor, que é a medida *Intersection over Union* (IoU) calculada entre a predição e o conjunto *Ground truth* (estado "verdadeiro" do objeto). Se a célula estiver deslocada do canto superior esquerdo da imagem por $(c_x; c_y)$ e a caixa delimitadora anterior tiver largura e altura p_w, p_h, então a predição é realizada conforme ilustrada na Figura 20.

Figura 20: Caixas delimitadoras com priorizações de dimensão e predição de localização



Fonte: Redmon et al. (2016)

Células e âncoras, em tais regiões, irão predizer objetos em áreas específicas da imagem. Cada caixa delimitadora será acompanhada por uma "pontuação de objetividade" que definirá a confiança do modelo em relação a essa caixa delimitadora que contém um objeto. Além disso, para cada classe de objeto possível, haverá um escore de probabilidade independente, que juntos devem somar 100%. Essa medida é calculada por $\Pr(Objeto) * IoU_{predict}^{truth}$. A função de perda na rede leva em conta o escore de objetividade, a classificação das categorias de objetos e a regressão das coordenadas / dimensões da caixa delimitadora.

A probabilidade condicional de cada classe é dada por $\Pr(Classe_i|Objeto)$. Um conjunto de probabilidades condicionais são calculadas por célula, a despeito da quantidade de caixas delimitadoras. Por fim, a taxa de confiança para uma classe específica é calculada pela multiplicação entre as probabilidades condicionais de classe e de cada caixa delimitadora, como expresso pela Equação 9. Essa taxa indica a probabilidade de uma classe ocorrer em uma região da imagem, bem como o quão acurada é a caixa delimitadora gerada (REDMON et al., 2016).

$$\Pr(Classe_i|Objeto) * \Pr(Objeto) * IoU_{predict}^{truth} = \Pr(Classe_i) * IoU_{predict}^{truth} \quad (9)$$

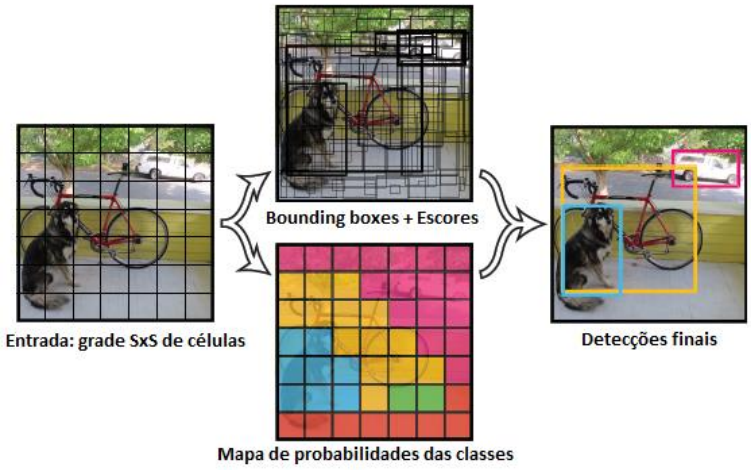
Na Figura 21 é ilustrado um exemplo no qual várias células são resultantes da classificação de um mesmo objeto, acarretando sobreposições de caixas delimitadoras. A maioria dessas delas é eliminada por uma pontuação de confiança mínima (limite), que é definida por padrão em 30%, ou porque elas estão suprimindo o mesmo objeto que outra caixa delimitadora com pontuação de confiança muito alta. Essa técnica é chamada de *Non-maximal Suppression (NMS)*.

Figura 21: Exemplo de múltiplas células classificadas como o mesmo objeto

Fonte: Redmon et al. (2016)

Além disso, predições que correspondam ao mesmo objeto *Ground truth* serão comparadas, e apenas aquelas com maior confiança serão mantidas. Na Figura 22 é ilustrado o esquema empregado pelo YOLO para a predição das caixas delimitadoras.

Figura 22: Esquema empregado pelo YOLO

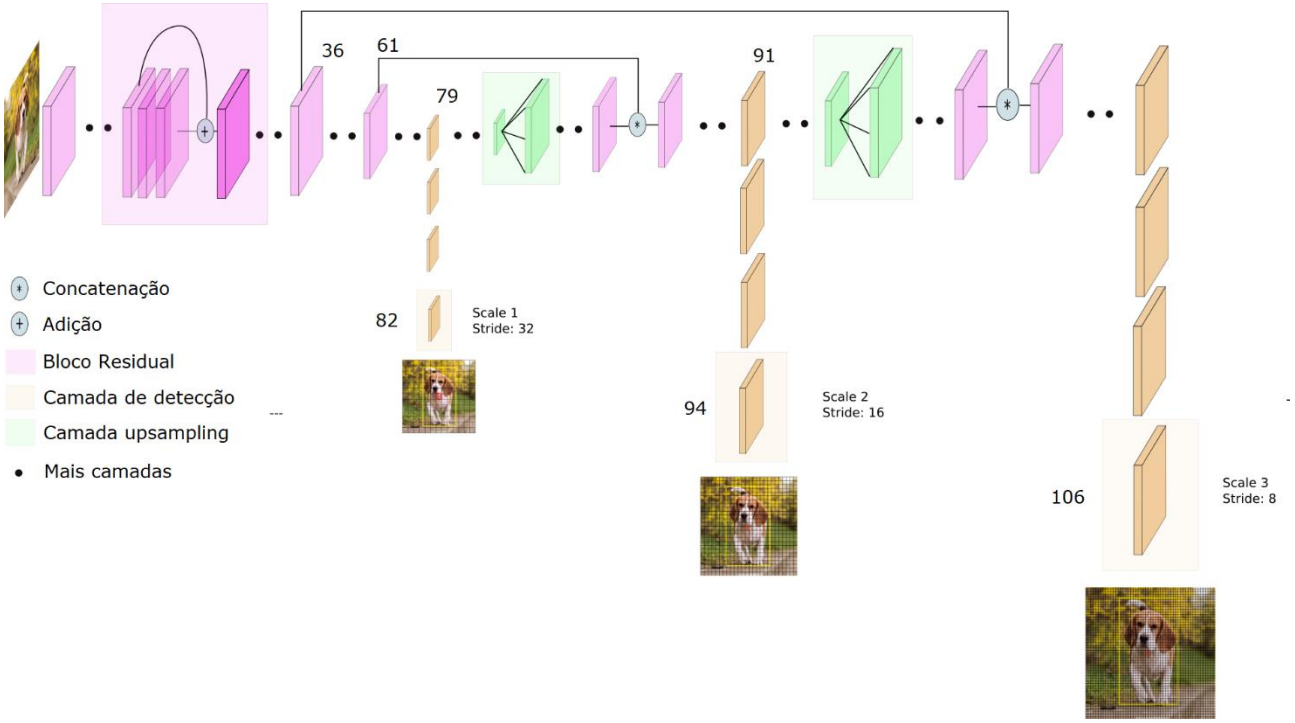


Fonte: Redmon et al. (2016)

Redmon e Farhadi (2018) lançaram o YOLOv3 com uma rede composta de 106 camadas, 53 para o backbone (“darknet-53”) e os outros 53 para a tarefa de detecção de objetos, ainda sendo um sistema composto por rede neural totalmente convolucional.

Na Figura 23 é ilustrada a arquitetura da RNC da YOLOv3 com 106 camadas totalmente convolucionais.

Figura 23: Arquitetura da RNC do YOLOv3



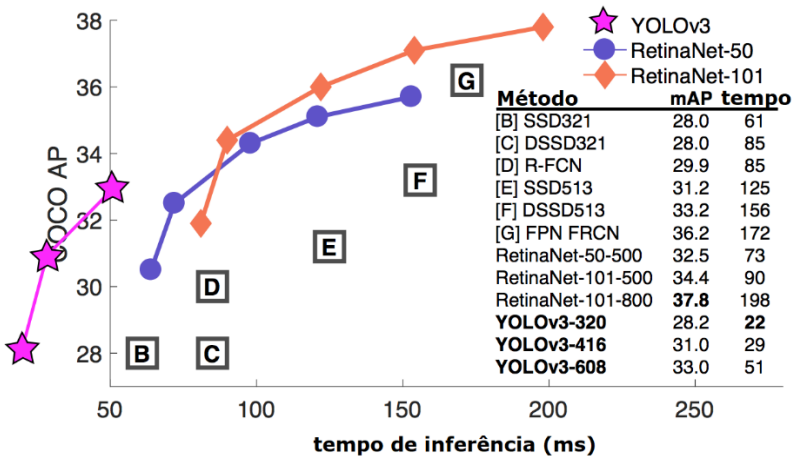
Fonte: adaptado de Redmon e Farhadi (2018)

Dentre os parâmetros que devem ser configurados para a rede, destacam-se os seguintes:

- *Batch*: definido por um valor que corresponde à quantidade de imagens selecionadas para cada iteração no treinamento da rede;
- *Subdivision*: definido por um valor que divide a quantidade de *batches* em mini-batches. Por exemplo, se *Batch=64* e *Subdivision=8*, então $64/8 = 8$ imagens serão submetidas para cada iteração;
- Pesos pré-treinados: correspondem aos valores de pesos, os quais são definidos aleatoriamente de acordo com a arquitetura da RNC.

Pode-se dizer que o YOLOv3 roda significantemente mais rápido do que outros métodos de detecção de objetos com comparável desempenho. O ótimo desempenho das RNCs do framework justifica seu uso nas tarefas de detecção de objetos-alvo e cenários neste trabalho. Na Figura 24 é ilustrado o gráfico que compara o desempenho do YOLOv3 com relação a outros métodos para a detecção de objetos em imagens.

Figura 24: Gráfico comparativo do YOLOv3 com outros métodos utilizando a banco de imagens COCO

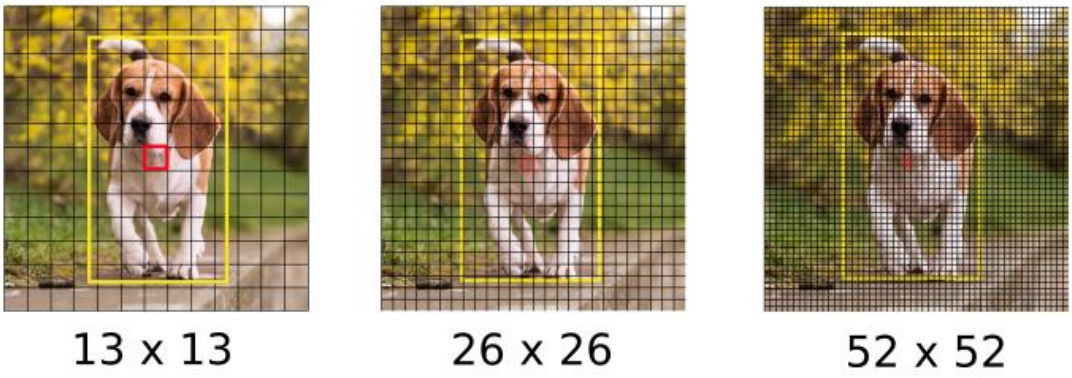


Fonte: adaptado de Redmon e Farhadi (2018)

Vale ressaltar que o YOLOv3 aprender a predizer bem em uma variedade de dimensões de entrada. Isso significa que a mesma rede pode prever detecções em diferentes resoluções. Assim, o YOLOv3 faz predições em 3 escalas diferentes. Isto significa que, com uma entrada de 416×416, são realizadas detecções nas escalas 13×13, 26×26 e 52×52, as quais são exemplificadas na Figura 25.

A rede reduz a imagem de entrada até a primeira camada de detecção, onde é feita uma detecção usando mapas de características de uma camada com passo 32. Além disso, as camadas são ampliadas por um fator de 2 e concatenadas com mapas de características de camadas anteriores com mapa de características de tamanhos idênticos. Outra detecção é feita agora na camada com *stride* 16. O mesmo procedimento de *upsampling* é repetido, e uma detecção final é feita na camada de *stride* 8.

Figura 25: Exemplos de detecções em 3 escalas diferentes



Fonte: Redmon et al. (2016)

Em cada escala, cada célula prevê 3 caixas delimitadoras usando 3 âncoras, perfazendo o número total de 9 âncoras usadas.

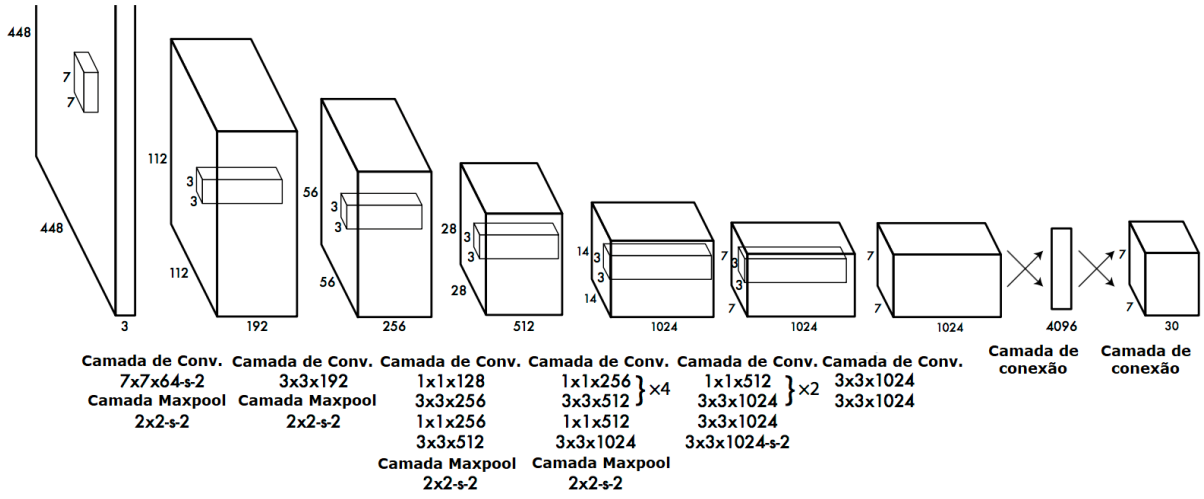
Um dos critérios de parada que deve ser considerado para a interrupção do treinamento da rede é o cálculo de perda, mais conhecido como *sum-square error*, que consiste em uma adição simples de diferenças, incluindo erros de coordenadas, erros de IoU e erro de classificação.

O YOLOv3 foi avaliado na tarefa de detecção de objetos no banco de imagens denominado COCO (80 classes) para duas métricas diferentes: *Mean Average Precision* (mAP-70), correspondendo a uma métrica de exatidão mais rigorosa e mAP-50 como sendo mais tolerante a menores qualidades de detecção que são convencionais na literatura. No *benchmark* do mAP-50, a YOLOv3 alcançou resultados semelhantes aos concorrentes como RetinaNet concebido por Lin et al. (2017) e o DSSD proposto por Fu et al. (2017), sendo mais rápido. No entanto, no benchmark que requer uma maior qualidade de predição (mAP-70), o YOLO tem uma diminuição significativa na precisão.

Para cada uma das três versões da YOLO, os autores também lançaram a respectiva variante “tiny-YOLOv3”. Essa versão é muito menor em comparação com as demais (apenas 9 camadas) e se concentra em ambientes restritos. Apesar de ser uma rede mais superficial, alcança bons resultados, além de ser mais de 10 vezes

mais rápida, ou seja, cerca de 220 FP. Na Figura 26 é ilustrada a arquitetura da RNC da tiny-YOLOv3

Figura 26: Arquitetura da RNC da tiny-YOLOv3



Fonte: adaptado de Redmon e Farhadi (2018)

2.2.3. ALGORITMOS GENÉTICOS

Algoritmo Genético (AG) consiste em um método de otimização inspirado nos mecanismos de evolução de populações de seres vivos. O AG foi introduzido por John Holland e popularizado por um dos seus alunos, David Goldberg (MITCHELL, 1997).

Otimização pode ser definida como a busca da melhor solução entre todas as possíveis para um dado problema. Um exemplo simples de otimização é a melhoria da imagem das televisões com antena acoplada no próprio aparelho. Por meio do ajuste manual da antena, várias soluções são testadas, guiadas pela qualidade de imagem obtida na TV, até a obtenção de uma resposta ótima ou subótima, ou seja, uma imagem de boa qualidade.

Geralmente, as técnicas de busca e otimização apresentam:

- Um espaço de busca, onde estão todas as possíveis soluções do problema;
- Uma função objetivo (função de aptidão), que é utilizada para avaliar as soluções produzidas, associando a cada uma delas uma nota.

Um AG processa populações de cromossomos. Um cromossomo é uma estrutura de dados, geralmente representada por vetor ou cadeia de bits, que representa uma possível solução do problema a ser otimizado. Em geral, um cromossomo representa um conjunto de parâmetros da função objetivo cuja resposta será maximizada ou minimizada. O conjunto de todas as configurações que o cromossomo pode assumir forma o seu espaço de busca. Se o cromossomo representa n parâmetros de uma função, então o espaço de busca é um espaço com n dimensões.

Inspirado no processo de seleção natural de seres vivos, o AG seleciona os melhores cromossomos da população inicial (aqueles de alta aptidão) para gerar cromossomos filhos por meio dos operadores de crossover e mutação. Uma população intermediária (*mating pool*) é utilizada para alocar os cromossomos pais selecionados. Geralmente, os pais são selecionados com probabilidade proporcional à sua aptidão. Portanto, a probabilidade de seleção pr_i de um cromossomo com aptidão ap_i é dada pela Equação 10.

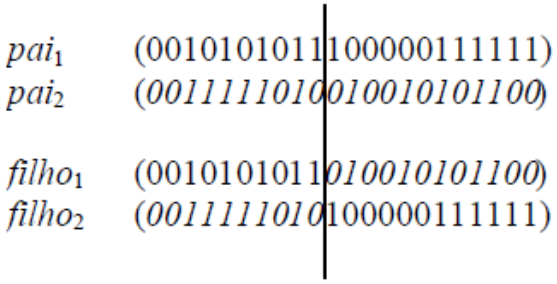
$$pr_i = \frac{ap_i}{\sum_{i=1}^{N} ap_i} \tag{10}$$

Os operadores de cruzamento e mutação são os principais mecanismos de busca dos AGs para explorar regiões desconhecidas do espaço de busca. O operador de cruzamento é aplicado a um par de cromossomos retirados da população intermediária, gerando dois cromossomos filhos. Cada um dos cromossomos pais tem sua cadeia de bits cortada em uma posição aleatória, produzindo duas cabeças e duas caudas. As caudas são trocadas, gerando dois novos cromossomos. Na Figura 27 é mostrado o comportamento do cruzamento.

O cruzamento é aplicado com uma dada probabilidade a cada par de cromossomos selecionados. Na prática, esta probabilidade, denominada de taxa de cruzamento, varia entre 60% e 90%. Não ocorrendo o cruzamento, os filhos serão iguais aos pais. Isto pode ser implementado, gerando números pseudoaleatórios no

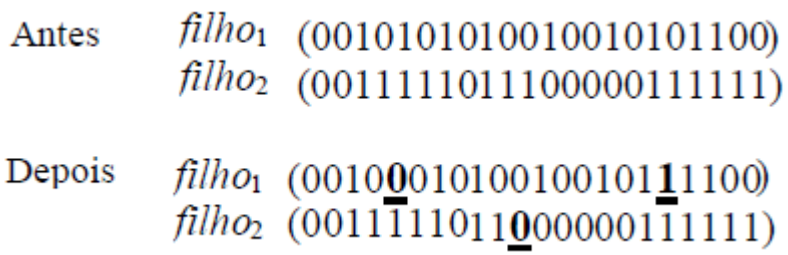
intervalo [0,1]. Assim, o cruzamento só é aplicado se o número gerado for menor que a taxa de cruzamento.

Figura 27: Comportamento do cruzamento



Após a operação de cruzamento, o operador de mutação é aplicado, com dada probabilidade, em cada bit dos dois filhos. O operador de mutação inverte *s* valores de bits, ou seja, muda o valor de um dado bit de 1 para 0 ou de 0 para 1. A mutação melhora a diversidade dos cromossomos na população, no entanto por outro lado, destrói informação contida no cromossomo, logo, deve ser utilizada uma taxa de mutação pequena (normalmente entre 0,1% a 5%), mas suficiente para assegurar a diversidade. Na Figura 28 é mostrado um exemplo em que dois bits do primeiro filho e um bit do segundo sofrem mutação.

Figura 28: Exemplo de mutação



Após a definição da primeira população, o procedimento se repete por um dado número de gerações. Quando se conhece a resposta máxima da função objetivo, pode-se utilizar este valor como critério de parada do AG. No entanto, não há um

critério exato para terminar a execução do AG, porém com 95% dos cromossomos representando o mesmo valor, é possível dizer que o algoritmo convergiu.

Vale ressaltar que o melhor cromossomo pode ser perdido de uma geração parar outra devido ao corte do crossover ou à ocorrência de mutação. Dessa forma, é interessante transferir o melhor cromossomo de uma geração para outra sem alterações. A esta estratégia dá-se o nome de elitismo.

O Algoritmo mostrado na Figura 29 ilustra o funcionamento do AG, considerando $S(gr)$ a população de cromossomos na geração gr.

Figura 29: Algoritmo Genético típico

```
gr ← 1
Inicializar S(gr)
Avaliar S(gr)
enquanto o critério de parada não for satisfeito faça
        gr ← gr + 1
        selecionar S(gr) a partir de S(gr – 1)
        aplicar crossover sobre S(gr)
        aplicar mutação sobre S(gr)
        avaliar S(gr)
fim enquanto
```

Neste trabalho o AG foi empregado na etapa de detecção de pequenas porções de água nas imagens para gerar um indicador a partir da combinação entre duas ou mais bandas espectrais.

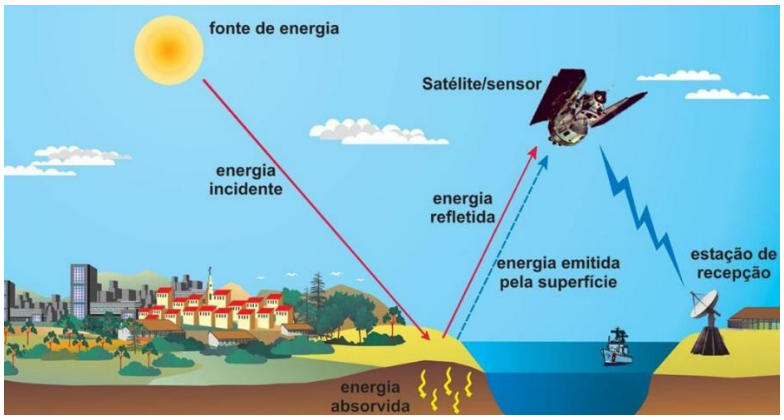
2.3. SENSORIAMENTO REMOTO

O termo sensoriamento remoto (SR) refere-se a um conjunto de técnicas destinado à obtenção de informação sobre objetos, sem que haja contato físico com eles (REEVES, 1975). Para compreender melhor a definição dada acima, é

necessário identificar os quatro elementos fundamentais do sensoriamento remoto, que são constituídos pelo sensor, a fonte, o alvo e a radiação eletromagnética (REM).

Na Figura 30 é ilustrado como funciona um sistema de sensoriamento remoto, que se dá de forma simples: a radiação eletromagnética que é emitida pelo sol e refletida pelo objeto terrestre é captada pelo sensor e convertida em um sinal que possui a capacidade de ser registrado, sendo posteriormente apresentado de uma forma adequada, na qual é possível extrair as informações que se procura, seja em forma de valores ou de imagens.

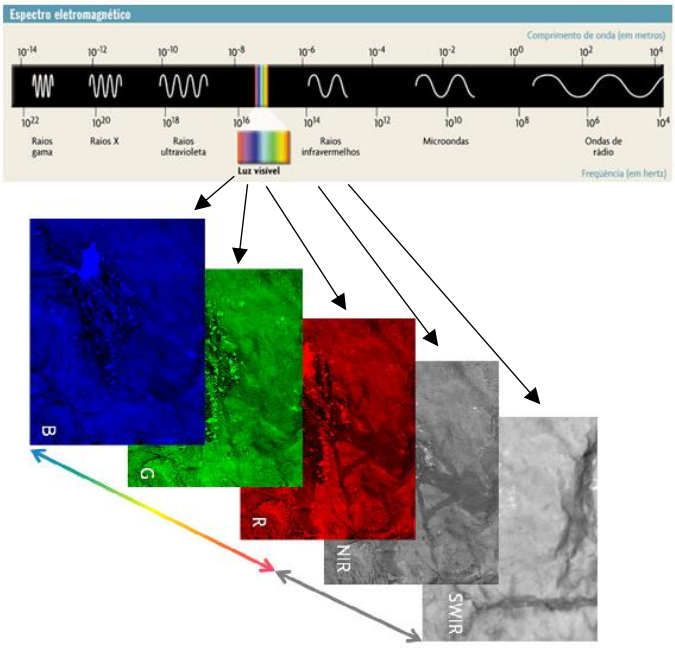
Figura 30: Sistema de Sensoriamento Remoto



Fonte: http://parquedaciencia.blogspot.com/2013/07/como-funciona-e-para-que-serve-o.html

Um sensor remoto é um dispositivo que detecta energia eletromagnética, quantifica e geralmente grava no formato analógico ou digital. Dentre os sensores destacam-se os satélites de observação da Terra, que tem como instrumento principal um sistema sensor capaz de produzir imagens da superfície da Terra em várias bandas simultâneas; neste caso, o imageador orbital funciona basicamente como uma câmera digital com as adaptações necessárias para gerar imagens em muitas bandas espectrais, os quais são denominados sensores multiespectrais. Na Figura 31 são mostrados exemplos de imagens multiespectrais nas 3 bandas (R, G e B) do espectro visível e nas bandas NIR (infravermelho-próximo) e $SWIR$ (infravermelho médio).

Figura 31: Exemplos de imagens multiespectrais no espectro visível e infravermelho



Fonte: adaptado de López et al. (2013)

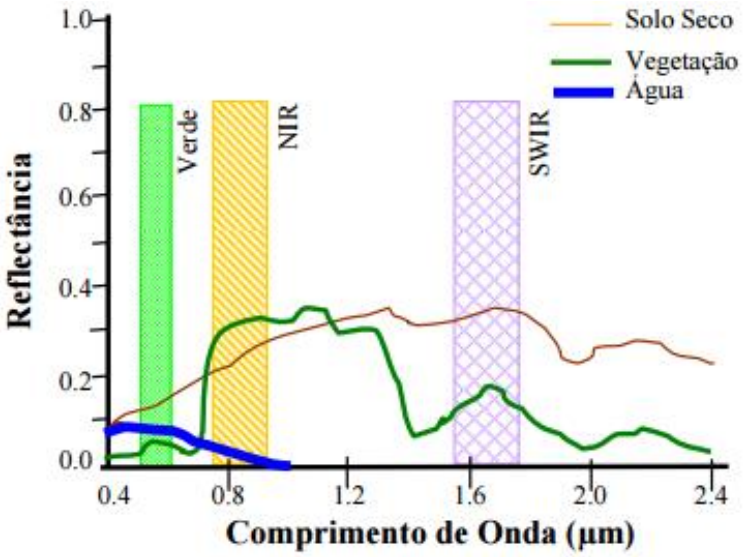
A radiação solar incidente na superfície terrestre interage de modo diferente com cada tipo de alvo. Esta diferença é determinada principalmente pelas diferentes composições físico-químicas dos objetos ou feições terrestres. Estes fatores fazem com que cada alvo terrestre tenha sua própria assinatura espectral. Em outras palavras, cada alvo absorve ou reflete de modo diferente cada uma das faixas do espectro da luz incidente.

Como pode ser observado na Figura 32, o comportamento espectral de feições que representam solo seco e rochas apresenta altas reflectâncias, principalmente no NIR e $SWIR$. Já, em feições que representam vegetação, existem altas reflectâncias no infravermelho-próximo. De acordo com Polidorio et al. (2004), a água tem a característica de refletir uma parcela muito pequena da luminosa incidida, pois a maior parte da energia luminosa incidente é transmitida, absorvida e dispersada pela água. O espectro da radiação refletida pela água ocupa, em geral, a faixa de comprimentos de onda entre 400-900nm, o que equivale à faixa do visível e o infravermelho-próximo.

No contexto de SR, corpo d'água é qualquer acumulação significativa de água, usualmente cobrindo a Terra ou outro planeta, tais como oceanos, mares e lagos.

Corpos d’água mais puros são mais evidentes por apresentarem baixa reflectância, principalmente nas faixas espectrais iguais ou superiores ao infravermelho-próximo.

Figura 32: Comportamento espectral do solo, vegetação e da água

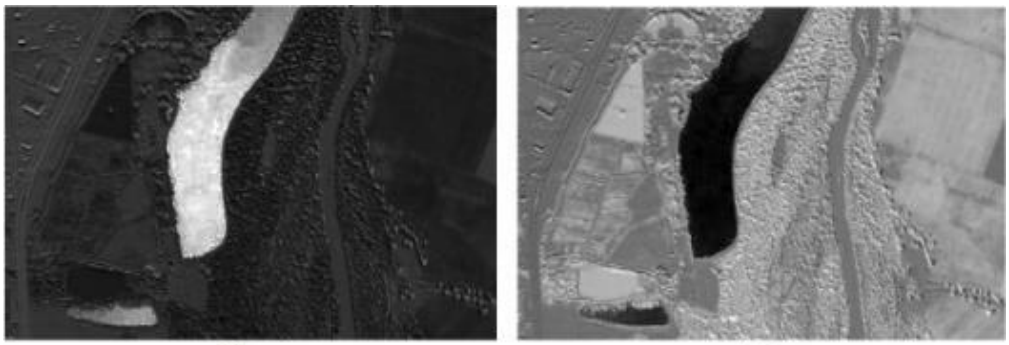


Fonte: adaptado de Polidorio et al. (2004)

Para a detecção de corpos d’água e de outras feições em imagens de satélites, índices indicadores geralmente são calculados a partir de bandas espectrais específicas. Dentre os mais utilizados destacam-se o NDVI – *Normalized Difference Vegetation Index* (ROUSE et al., 1974), NDWI – *Normalized Difference Water Index* (GAO, 1996) e IIA – Índice Indicador de Água (POLIDORIO et al., 2004).

Na Figura 33 é possível observar a detecção de um corpo d’água utilizando o $NDVI$ e o IIA. Para o cômputo do $NDVI$, as áreas que contém vegetação apresentarão valores de índice inferiores ou próximos de 1 (um) e as áreas sem vegetação apresentarão valores de índice próximos de -1; no $NDWI$, valores maiores ou iguais a 1 indicam a presença de corpos d’água e no IIA, valores próximos de *0* para os corpos d’água e para outras feições valores próximos de –1. Assim, os corpos d’água são identificados na Figura 33a com níveis de cinza mais próximos do branco e, na Figura 33b, mais próximos do preto.

Figura 33: (a) imagem resultante da aplicação do IIA; (b) imagem resultante da aplicação do NDVI



Fonte: Polidorio et al. (2017)

O reconhecimento de corpos d'água a partir de imagens de sensoriamento remoto tem sido amplamente explorado. A partir da utilização de índices indicadores e imagens de satélites, é possível reconhecer os diferentes tipos de feição na superfície terrestre, incluindo os corpos d'agua. Dentre eles, podemos citar os trabalhos de Zhou et al. (2014), Yang e Chen (2017) e Khandelwal et al. (2017). Contudo, os índices indicadores propostos para extração de feições d'água em imagens de satélites, em geral, não reproduzem os mesmos resultados quando aplicados em imagens adquiridas por VANTs, visto que nessas as bandas espectrais não necessariamente têm as mesmas larguras e estão nas mesmas faixas das bandas espectrais das imagens adquiridas por satélites

2.3.1. SENSORIAMENTO REMOTO USANDO VANTS

Os veículos aéreos não tripulados (VANTs), também conhecidos como drones, têm sido utilizados como plataformas para o sensoriamento remoto e envolvem várias aplicações, tais como cadastro de propriedades, segurança, monitoramento de obras, agricultura de precisão, mineração, monitoramento ambiental, entre outras (CANDIDO, SILVA e FILHO, 2015).

Drones comumente utilizados em sensoriamento remoto são os multi-rotores e os de asa fixa. Ambos são utilizados para levantamentos de dados geoespaciais, cada um com sua capacidade de acordo com sua arquitetura.

Os sensores de imageamento utilizados neste tipo de veículo podem ser sensores na faixa do visível (RGB), infravermelhos (NIR), multiespectrais e sensores hiperespectrais.

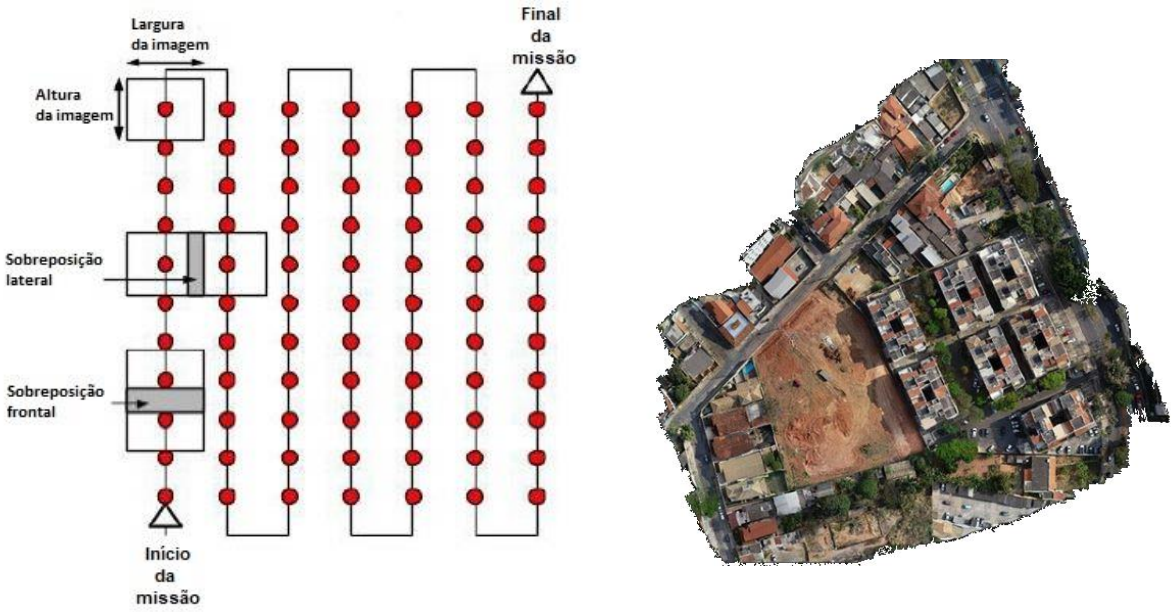
Para realizar a aquisição de imagens usando um VANT é necessário que seja feito o planejamento da missão, visando selecionar a área a ser mapeada. Essa área é delimitada por meio da escolha de alguns pontos, os quais devem ser marcados em mapas de satélites. Na Figura 34a é possível observar como ocorre a missão para que se alcance a sobreposição das imagens para a geração do mosaico, sendo que os pontos vermelhos indicam as coordenadas das imagens que devem ser adquiridas. Quanto mais próximos forem os pontos de aquisição, melhor será a sobreposição das imagens.

A partir das imagens adquiridas em uma determinada área, é possível gerar um mosaico ou ortomosaico (Figura 34b), o qual é construído por meio de procedimentos de calibração radiométrica, de alinhamento e de ortorretificação das imagens, além da busca de pontos homólogos entre duas ou mais imagens sobrepostas entre si. É importante dizer que quanto maior for a sobreposição das imagens adquiridas em uma missão pelo VANT, melhor será a qualidade do ortomosaico.

Na aerofotogrametria, o GSD (*Ground Sample Distance*) é uma das variáveis mais importantes e é a primeira que deverá ser definida, pois ela garante a resolução espacial do mapeamento, ou seja, o nível de detalhamento. A escolha do GSD influencia diretamente na nitidez do mapeamento, pois para aumentar o nível de detalhamento deve-se sobrevoar mais baixo, com isso uma porção menor do terreno será coberta, consequentemente uma área menor será mapeada. Para aumentar a capacidade de mapeamento deve-se aumentar o tamanho do GSD e como consequência se perderá detalhamento (nitidez). Assim, chega-se a uma relação: o tamanho do GSD é inversamente proporcional ao nível de detalhamento, ou seja, quanto maior o GSD, menor é o nível detalhamento e quanto menor for o GSD, maior será o nível de detalhamento.

Imagens aéreas adquiridas por VANTs vem trazendo um nível de detalhamento inovador para o sensoriamento remoto, o que permite avanços significativos na qualidade de algumas aplicações em várias áreas. Aguirre-Gómez et al. (2017), por exemplo, demonstraram a análise, em escala temporal e espacial, da extensão e distribuição de cianobactérias nos lagos de Chapultepec, situado na região oeste da cidade do México, por meio de imagens adquiridas por um VANT.

Figura 34: (a) plano de voo de um VANT; (b) exemplo de um ortomosaico



(a) **(b)**

Fonte: adaptado de Cassemiro e Pinto (2014)

Hardy et al. (2017) propuseram um método para o mapeamento de corpos d'água naturais (lagoas e rios peri urbanos) e arrozais irrigados e não irrigados, que podem caracterizar habitats do vetor da Malária, em sete locais na principal ilha do arquipélago de Zanzibar, por meio de imagens adquiridas por um drone.

Albuquerque et al. (2017) abordaram a aplicação de um conjunto de VANTs e sensores para a aquisição de imagens, no município de Lençóis Paulista-SP, com a finalidade de avaliar processos de Restauração Florestal. A avaliação é feita por meio do cálculo do índice $MPRI$ (*Modified Photochemical Reflectance Index*), o qual mostrou bom potencial para monitorar o padrão de vegetação.

2.4. TRABALHOS ABORDANDO TEMAS CORRELATOS ENCONTRADOS NA LITERATURA

Nesta seção é feita uma breve descrição de cada um dos trabalhos encontrados na literatura acerca das seguintes temáticas que permeiam este trabalho: detecção e localização de objetos, detecção de porções de água e identificação de possíveis criadouros do mosquito *Aedes aegypti*. Cabe enfatizar que foram considerados apenas os trabalhos dos últimos 5 anos que exploram tais temáticas com o uso de imagens aéreas adquiridas por VANTs.

Com relação à detecção de objetos em imagens adquiridas por VANTs, Xu et al. (2017) propuseram um *framework* para detecção de carros utilizando uma RNC, denominada *Faster R-CNN*, em imagens capturadas de vídeos em baixa altitude, as quais foram adquiridas em cruzamentos sinalizados. Nos experimentos realizados foi atingida a completude de 96,40% e a corretude de 98,43% com 2,10 f/s de velocidade de detecção em tempo real. Ammour et al. (2017) também desenvolveram um método para a detecção e contagem de carros utilizando uma RNC combinada com o classificador SVM (*Support Vector Machines*), que atingiu uma acurácia de 93,6% em experimentos realizados. Porém, há a ocorrência de muitos falsos-positivos nos resultados da classificação.

Com o intuito de auxiliar operações de busca e salvamento em regiões com risco de avalanches, Bejiga et al. (2017) desenvolveram um método para extrair descritores de imagens de detritos dessas regiões por meio de uma RNC e um classificador SVM para a detecção de objetos de interesse tais como esquis ou possíveis vítimas. Tal método atingiu uma acurácia de 97,59% na classificação dessas imagens, porém foi constatado que há uma melhora no processo de classificação quando a resolução das imagens é maior, prejudicando o tempo de processamento.

Yi et al. (2019) desenvolveram um método, que atingiu a taxa de 0,74 para o AP (average precision), para detecção de pedestres utilizando uma das arquiteturas de RNCs do YOLOv3, denominada tiny-YOLOv3, em conjunto com o algoritmo *k-means* para filtrar as melhores características do conjunto de treinamento. Benjdira et al. (2019) realizaram um estudo comparativo do método Faster R-CNN com o YOLOv3. Para a detecção de carros utilizando VANTs, eles demostraram que o

YOLOv3 é melhor do que o Faster R-CNN, pois atingiu uma acurácia de 99,07%. Tian et al. (2019) desenvolveram um método, utilizando o YOLOv3, para a detecção em tempo real (tempo médio de detecção=0.304 segundos por *frame*) de maçãs em pomares a fim de avaliar as fases de crescimento das maçãs e estimar o rendimento. Experimentos demostraram que o modelo YOLOv3-denso (versão modificada) proposto é superior ao modelo YOLO-v3 original e ao modelo de rede R-CNN com VGG16.

No que tange a detecção de porções de água em imagens áreas adquiridas por VANTs, apenas alguns trabalhos foram encontrados na literatura. Dentre eles destaca-se um método proposto por Colet, Braun e Manssour (2016) para identificação automática de superfícies de águas turvas navegáveis, baseada em técnicas de visão computacional. Redes neurais artificiais (RNAs) também foram usadas para construir um classificador projetado para gerar um mapa de navegação, e a análise de componentes principais (PCA) foi realizada para comprimir as informações extraídas usadas como entrada para a RNA. Em experimentos realizados em imagens de 3 cenários diferentes, foi obtida uma acurácia entre 91,21% e 95,85%, no entanto ainda há a presença de falsos-positivos. Aguirre-Gómez et al. (2016) realizaram análises, em escala espacial e temporal, da extensão e distribuição de cianobactéria, em imagens adquiridas por VANTs, nos lagos Chapultepec localizados no México. As análises demostraram que o uso de VANTs em medições reais caracteriza um método preciso, flexível, barato e rápido para detectar e prever a eutrofização e, portanto, o florescimento de cianobactérias em reservatórios de água. No entanto, nos experimentos conduzidos nesses trabalhos apenas grandes porções de água (corpos d'água) como rios, lagos e grandes poças foram consideradas. Desta forma, a detecção de pequenas porções de água (por exemplo em caixas d'agua e outros recipientes destampados) continua sendo um grande desafio.

No que se refere ao mapeamento de possíveis criadouros do mosquito *Aedes aegypti*, de um modo geral, foram encontrados 6 trabalhos, considerando o período de 2014 a 2018, os quais são descritos a seguir.

Agrawal et al. (2014) desenvolveram um método que visa a detecção e visualização de possíveis criadouros do mosquito, com base em 500 imagens georreferenciadas obtidas da Internet. O método envolveu três etapas: avaliação da qualidade das imagens; classificação das imagens utilizando *bag of visual words*

(sacola de características visuais) com o descritor SIFT e o classificador SVM; e a visualização dos criadouros usando mapas de calor onde são apontadas as regiões com mais riscos de incidência de habitats do mosquito. Na etapa de classificação foi obtida uma acurácia em torno de 82%, porém a questão da detecção de água parada nesses criadouros não foi bem explorada.

No trabalho desenvolvido por Mehra et al. (2016) foi proposto um *framework* para a detecção dos criadouros e de água parada utilizando imagens do Google e de diversos dispositivos (câmeras digitais, celulares e drones). Para a extração de características também foi utilizada a técnica *bag of visual words* com o descritor SURF e a classificação por meio de classificadores bayesianos, atingindo uma acurácia de 90% em experimentos realizados. Imagens termográficas também foram utilizadas para compor os conjuntos de treinamento utilizados.

Fornace et al. (2014) utilizaram-se de imagens adquiridas por VANTs para um estudo de caso a fim de realizar um monitoramento epidemiológico da Malária em regiões da Malásia e Philipinas. Pela análise visual das imagens, foi possível monitorar mudanças em habitats de vetores da doença e em reservatórios de animais selvagens. As imagens adquiridas por esses veículos neste trabalho proporcionaram uma análise visual mais acurada devido ao nível de detalhamento ser maior do que as adquiridas por satélites, porém tal análise ainda foi feita de forma manual. No trabalho desenvolvido por Diniz e Medeiros (2018) foram adquiridas imagens por um VANT de um bairro do município de Caicó-RN e, a partir de um ortomosaico gerado por essas imagens, foi possível conceber um sistema para realizar o mapeamento manual (visualmente) de possíveis criadouros (objetos e cenários) do mosquito *Aedes aegypti*, sem considerar a possibilidade de acúmulo de água ou não. Embora tais imagens utilizadas neste trabalho permitissem um mapeamento mais detalhado dos criadouros, a marcação e identificação de objetos e cenários também foi feita manualmente.

Prasad et al. (2015) propuseram o uso de um quadcopter (tipo específico de VANT) para inspecionar imagens e vídeos adquiridos em áreas urbanas tais como terraços, construções e estações de bombeamento afim de identificar acúmulo de água parada que podem ser possíveis criadouros do mosquito. A identificação foi feita por meio da utilização do classificador SVM e de outro classificador de fluxo óptico.

Neste trabalho foi explorada a natureza especular da água no processo de detecção, porém não há métrica que indique o quão acurada ela é.

Passos et al. (2018) realizaram um estudo de técnicas de Aprendizagem de Máquina para a detecção automática de objetos em imagens. Para este fim, foi criada uma base de dados própria contendo vídeos com diversos recipientes que acumulam água limpa espalhados em diversos cenários. Antes de iniciar a aquisição das imagens, os parâmetros da câmera são ajustados manualmente e um procedimento de calibração dela é realizado. Todos os vídeos e imagens são manualmente anotados e podem compor conjuntos de treinamento e validação de algoritmos detectores de objetos. Neste trabalho é considerado apenas a detecção de alguns objetos suspeitos sem considerar a presença de cenários.

Nos trabalhos relacionados à detecção de objetos em imagens adquiridas por VANTs não há menção de abordagens para a detecção de objetos específicos tais como os reservatórios d'água domésticos, que podem caracterizar um dos potenciais criadouros do mosquito *Aedes aegypti* que mais aparecem em áreas urbanas, principalmente nas regiões mais periféricas das grandes cidades.

Na maioria dos trabalhos diretamente relacionados à temática investigada neste trabalho, análise das imagens é feita manualmente (análise visual). Os poucos trabalhos que tratam do mapeamento automático de possíveis criadouros do mosquito consideram somente a existência ou não de objetos e/ou cenários suspeitos, sem fornecer localização espacial.

Por fim, as escassas abordagens encontradas na literatura para a detecção de pequenas porções de água não tratam da identificação de acúmulo de água em reservatórios domésticos destampados que podem vir a ser criadouros do mosquito *Aedes aegypti*.

3. MÉTODOS E MATERIAIS

3.1. CARACTERIZAÇÃO DA PESQUISA

Do ponto de vista de sua natureza, esta pesquisa pode ser classificada como aplicada, pois caracteriza-se pelo seu interesse prático, ou seja, há a pretensão que os resultados sejam aplicados ou utilizados imediatamente na solução de problemas que ocorrem na realidade. Já do ponto de vista de seus objetivos, ela se caracteriza como exploratória, uma vez que visa maior familiaridade com o problema investigado no intuito de torná-lo mais explícito ou construir hipóteses (APPOLINÁRIO, 2006).

Quanto aos objetivos, a pesquisa pode ser classificada como explicativa, pois os processos que estão sendo estudados são descritos em sua totalidade, com todas as etapas e processos utilizados na resolução do problema (GIL, 2002).

Tendo em vista que o mapeamento automático é realizado a partir de experimentos aplicados sobre as imagens adquiridas aéreas adquiridas pelos VANTs, podemos caracterizar o método de pesquisa como experimental.

3.2. MATERIAIS

3.2.1. BASE DE IMAGENS

As imagens utilizadas neste trabalho foram adquiridas de acordo com o seguinte protocolo:

- Na aquisição de imagens aéreas de áreas urbanas com o uso de VANTs a distância máxima foi de 70 m acima do solo, obedecendo a distância máxima de 120 m indicada nas especificações da ANAC (Agência Nacional de Aviação Civil - RBAC-E nº 94/2017);
- As imagens foram adquiridas em dias com sol;
- As imagens foram adquiridas em apenas ambientes externos (edificações urbanas e ambientes simulados).

A maioria das imagens que compõe a base de imagens desenvolvida neste trabalho foi adquirida com o uso de 3 VANTs: DJI Phantom 3 *professional* equipado com uma câmera RGB Sony EXMOR 12.4 MP (modelo FC300X; 1/2.3” CMOS; FOV 94° 20 mm; f/2.8 de abertura), DJI Phantom 4 *advanced* com uma câmera RGB DJI 20 MP (modelo FC330; 1/2.3” CMOS; FOV 94° 20 mm; f/2.8 de abertura); e DJI Spark Combo com uma câmera RGB DJI 12 MP (modelo FC220; 1/2.3” CMOS; FOV 81.9° 25 mm; f/2.6 de abertura). Em todos os casos as bandas espectrais variaram de 370 a 750 nm (espectro visível).

Foram adquiridas imagens com resolução 4000×3000 pixels, as quais constituem os conjuntos denominados DS1, DS2, DS3, DS4, DS5 e DS6. A base contempla ainda outro conjunto de imagens que foram adquiridas sem uso de VANT, denominado DS7, o qual é composto por imagens de 3000×2250 pixels. Vale dizer que os locais onde foram feitas as aquisições foram selecionados pela conveniência de pessoas que os habitam, pelo suporte prestado pela ONG Teto[1] (como no caso da comunidade Porto de Areia) ou por serem espaços públicos.

O conjunto de imagens DS1 contém 76 imagens, contendo piscinas e pequenos lagos, adquiridas no dia 22/05/2016, em um dia ensolarado com temperatura de 26,5 graus, em uma região com chácaras localizada na cidade de Mairiporã/SP (coordenada central com latitude -23.23531 e longitude -46.61574). Na câmera RGB acoplada ao drone, foi empregada uma lente especial para a filtragem da banda espectral infravermelho próximo. A lente possui as seguintes especificações: GoPro Hero modelo GP33728 16MP RGN (Red+Green+NIR), com as bandas espectrais na faixa de 400 a 1.000 nm; comprimento focal de 3,37 mm e f/2.8 de abertura. O voo foi realizado a 50 m do solo, com GSD (*Ground Sample Distance*) de 2,16 cm/px (centímetros por pixel).

O conjunto DS2 é composto por 92 imagens adquiridas em uma área da Universidade de São Paulo – USP, no dia 04/09/2016, em um dia com sol entre nuvens com temperatura na faixa de 21,8 graus, contendo piscinas e pequenas fontes

[1] https://www.techo.org/brasil/?gclid=CjwKCAjwxOvsBRAjEiwAuY7L8to5QHP37rAAGw4kpk7F3kyhpT1FLF4gSm0keZddnBwlAkD6AIl7qBoCQJwQAvD_BwE

(coordenada central com latitude -23,5614311 e longitude -46,7198984). Para a composição desse conjunto, as imagens foram adquiridas em duas missões: a primeira usando apenas a câmera RGB acoplada ao drone e a segunda considerando a adaptação da mesma lente para a filtragem do infravermelho próximo. A exemplo das aquisições realizadas para compor o DS1, o voo foi realizado a 50 m do solo, com GSD de 2,16 cm/px.

As 142 imagens que compõem o conjunto DS3 foram adquiridas, no dia 30/04/2017, em um dia ensolarado com temperatura na faixa de 22,3 graus, em dois distritos localizados no extremo leste do município de São Paulo: Guaianases e Ferraz de Vasconcelos (coordenada central com latitude -23,5298832 e longitude -46,3993575). As imagens desse conjunto, na grande maioria, possuem reservatórios de água para uso doméstico (caixas d'água) de vários modelos e foram adquiridas em uma única missão para cada localidade, usando apenas a câmera RGB. O voo foi realizado a uma distância de 30 m do solo (aquisição 1) e de 40 m (aquisição 2), com GSD de 1,30 e 1,73 cm/px, respectivamente.

Para compor o conjunto DS4 foram adquiridas 60 imagens em uma chácara particular localizada na cidade de Mairiporã/SP (coordenada central com latitude -23,23531 e longitude -46,61574). As imagens foram adquiridas, no dia 24/02/2019, em um dia ensolarado com temperatura na faixa de 32,5 graus, em duas missões: a primeira usando apenas a câmera RGB acoplada ao drone, considerando quatro distâncias acima do solo: 1, 2, 5 e 7 m, cujo GSD variou de 0,04 a 0,30 cm/px; e a segunda com a adaptação da lente infravermelho próximo, considerando as mesmas distâncias. As imagens pertencentes a esse conjunto possuem cenários simulados com seis recipientes contendo pequenas porções de água, com quantidade variada, em diversas condições: água limpa, água suja e água com limo. Além disso, uma calha e pneus velhos contendo água também fizeram parte dos cenários imageados.

O conjunto DS5 foi cedido pela ONG Teto. Ele é composto por 101 imagens de 4000×3000 pixels adquiridas em 2017, em uma comunidade denominada Porto de Areia, localizada no distrito de Carapicuíba/SP (coordenada central com latitude -23.520632 e longitude -46.827239). As imagens desse conjunto foram adquiridas por meio de uma câmera RGB a uma distância de 70 m do solo, com GSD de 2,46 cm/px.

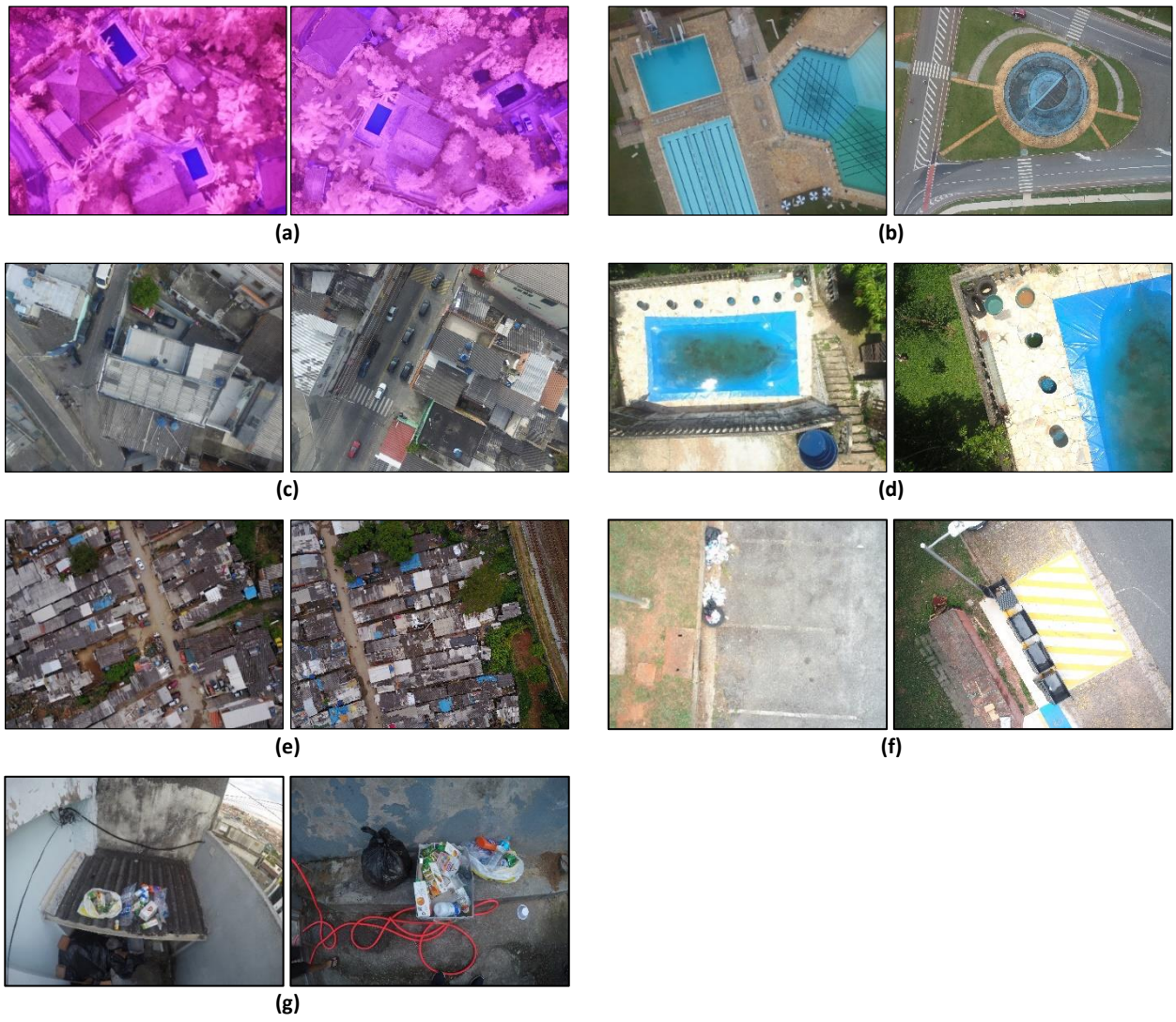
Com o intuito de simular cenários que podem tornar-se potenciais criadouros do mosquito, foram feitas duas aquisições: em uma área da Universidade de São Paulo – USP (coordenada central com latitude -23,5614311 e longitude -46,7198984) no dia 26/10/2019, em um dia ensolarado com temperatura na faixa de 29 graus, e em uma residência particular no bairro de Guaianases (coordenada central com latitude -23,5298832 e longitude -46,3993575) no dia 27/10/2019, em um dia ensolarado com temperatura na faixa de 32,6 graus. Para tanto, foram adquiridas 119 imagens com o VANT DJI Phantom 4 *advanced* e 111 imagens com uma câmera GoPro HERO4 Silver (1/2.3" CMOS; FOV 90° 3.0mm; f/2.8 de abertura), respectivamente. As 119 imagens da primeira aquisição (conjunto DS6) foram obtidas a três distâncias acima do solo: 7, 10 e 13 m, cujo GSD variou de 0,30 a 0,56 cm/px. Já as 111 imagens do conjunto DS7 foram adquiridas a uma distância de 3 m (± 0,5 m) acima do solo. A maioria das imagens possui lixo a céu aberto com pneus e pequenos recipientes que podem acumular água em diversas situações.

O Quadro 2 sintetiza a composição da base de imagens desenvolvida neste trabalho e que foi utilizada na realização e validação dos experimentos, enquanto na Figura 35 são ilustrados exemplos de imagens pertencentes aos sete conjuntos (DS1 a DS7).

Quadro 2: Composição da base de imagens utilizada neste trabalho

Conjunto	Número de imagens	Resolução (em pixels)	Equipamento usado na aquisição	Altura (s) do solo (em m)
DS1	76	4000×3000	DJI Phantom 3 *professional*	50
DS2	92	4000×3000	DJI Phantom 3 *professional*	50
DS3	142	4000×3000	DJI Phantom 3 *professional*	30 e 40
DS4	60	4000×3000	DJI Phantom 4 *advanced*	1, 2, 5 e 7
DS5	101	4000×3000	DJI Spark Combo	70
DS6	119	4000×3000	DJI Phantom 4 *advanced*	7, 10 e 13
DS7	111	3000×2250	Câmera GoPro HERO4 Silver	3 (± 0,5)

Figura 35: Exemplos de imagens dos conjuntos: DS1 (a); DS2 (b); DS3 (c); DS4 (d); DS5 (e); DS6 (f); DS7 (g)



Para o planejamento das missões dos VANTs, usando um *smartphone*, foi utilizado o aplicativo *Map Made Easy - Map Pilot*, que permite indicar a área onde as aquisições são realizadas, bem como ajustar o valor de sobreposição das imagens, altura do drone, dentre outras configurações.

Como pode ser visto, a base desenvolvida e usada nos experimentos é composta por 701 imagens que contemplam os objetos e cenários descritos na seção 1.2, a qual apresenta o problema de pesquisa investigado neste trabalho.

3.2.2. AMBIENTES COMPUTACIONAIS, SOFTWARES E HARDWARE EMPREGADOS NA CONDUÇÃO DOS EXPERIMENTOS

A maioria dos algoritmos que constituem a abordagem proposta neste trabalho foi implementada em linguagem C/C++ com o uso da plataforma Microsoft® Visual Studio™ 2015, a IDE (*Integrated Development Environment*) DevC++, OpenCV[2], GAlib[3] e Darknet[4], as quais são compostas por rotinas e algoritmos de processamento de imagens, visão computacional, algoritmos genéticos e *deep learning*. Além disso, conforme consta no Quadro 3, alguns experimentos foram realizados no software Matlab 2018. Vale ressaltar que, para os experimentos utilizando a biblioteca Darknet, foi utilizada a plataforma CUDA para acessar os recursos de processamento da GPU (*Graphics Processing Unit*).

Quadro 3: Ambientes de programação usados no desenvolvimento dos principais algoritmos que contemplam as etapas da abordagem proposta.

Algoritmo(s)	**Ambiente(s) de programação**
Detecção de objetos-alvo	Microsoft® Visual Studio™ 2015 (linguagem C/C++) com as bibliotecas OpenCV e Darknet (YOLOv3)
Detecção de cenários	Microsoft® Visual Studio™ 2015 (linguagem C/C++) com as bibliotecas OpenCV e Darknet (Tiny YOLO) Matlab 2018 (BoVW+SVM)
Detecção de pequenas porções de água	DevC++ (linguagem C/C++) com as bibliotecas OpenCV e GAlib

[2] OpenCV (Open Source Computer Vision Library) – https://www.opencv.org/

[3] http://lancet.mit.edu/ga/

[4] https://github.com/AlexeyAB/darknet

O hardware usado para os experimentos foi um processador Core i7-6500U, com 16 GB de memória RAM, com 2,5GHz de velocidade e placa de vídeo NVIDIA GeForce 930M com 4 GB.

3.3. PROCEDIMENTO PARA CONDUÇÃO DOS EXPERIMENTOS E AVALIAÇÃO DOS ALGORITMOS E ABORDAGENS DESENVOLVIDAS

Para a detecção de possíveis criadouros do mosquito como, por exemplo, os reservatórios d'água domésticos, foram utilizadas as imagens que fazem parte dos conjuntos DS3 e DS5. Para tanto, foram feitos experimentos com uma arquitetura de RNC do *framework* YOLOv3.

As imagens pertencentes aos conjuntos DS4, DS6 e DS7 foram utilizadas nos experimentos para a detecção de cenários (simulados) suspeitos de serem possíveis criadouros do mosquito por meio das técnicas RNC (*framework* YOLO modelo tiny) e BoVW+SVM.

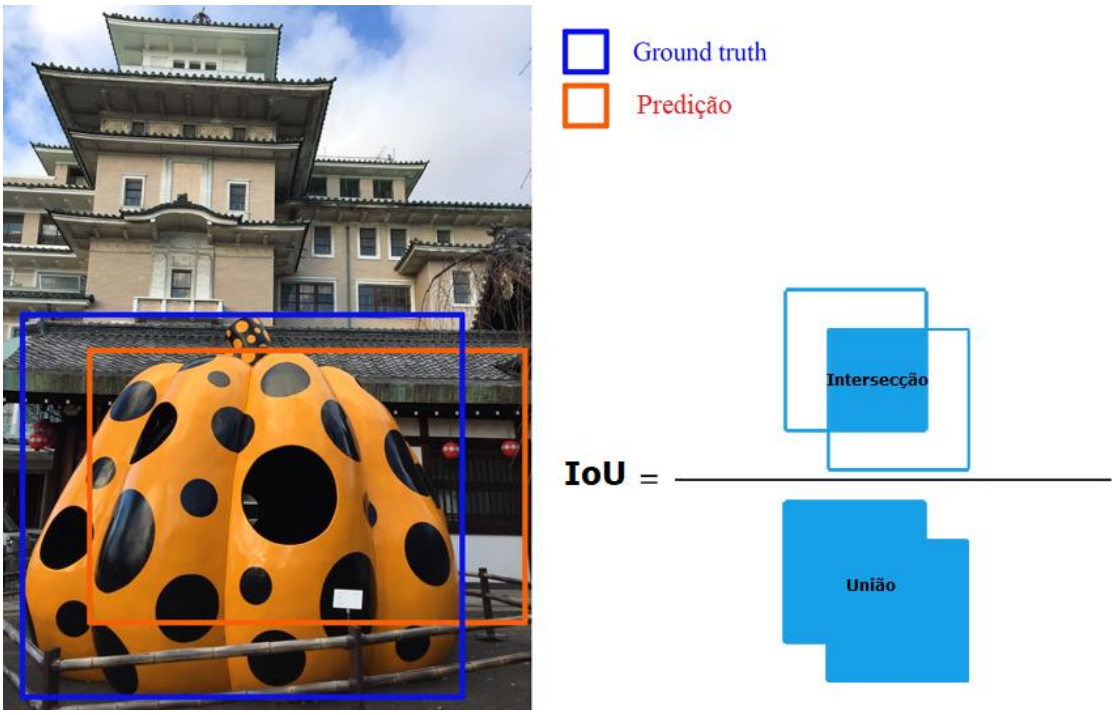
Para a condução dos experimentos acerca da detecção de pequenas porções de água foram utilizadas as imagens pertencentes ao conjunto DS2. Um AG foi empregado com a finalidade de fornecer a criação de um índice indicador de água com base em tais imagens. É importante dizer que, neste trabalho, pequenas porções de água referem-se a volumes de água bem menores quando comparados aos corpos d'água mencionados em trabalhos de sensoriamento envolvendo imagens de satélites.

3.3.1. MÉTRICA PARA AVALIAÇÃO DOS MÉTODOS DE DETECÇÃO DE OBJETOS E CENÁRIOS

A avaliação do desempenho de um método para detecção de objetos de interesse em imagens é feita com base em alguma métrica. Uma delas é a *Intersection-over-Union* (IoU), também conhecida como Índice de *Jaccard*. Ela mede quão semelhante espacialmente é uma caixa delimitadora predita a uma dada *Ground-Truth* (Rahman e Wang, 2016). Na Figura 36 são ilustradas a caixa predita

(identificada com a cor laranja) e a *Ground-Truth* (identificada com a cor azul). A Equação 11 demonstra o cálculo para a IoU, que define a proporção entre a interseção de ambas as caixas delimitadoras sobre sua união.

Figura 36: Definição da métrica IoU



$$IoU = \frac{região\ de\ intersecção}{região\ de\ união} \tag{11}$$

Há termos comuns no treinamento e teste de algoritmos de aprendizagem de máquina aplicados em tarefas de classificação, os quais se referem às amostras de aprendizado e às predições. São eles:

- *Ground truth* (GT): também conhecido como Real Positivo (RP) na Detecção de Objetos, corresponde à caixa delimitadora na qual o modelo será treinado e representa o resultado perfeito que o modelo deve almejar durante o treinamento. As métricas usam essas “anotações” do GT no conjunto de dados como o referencial.

- Verdadeiro Positivo (VP): refere-se a predições que foram feitas corretamente e estão vinculadas aos GTs.
- Falso positivo (FP): é geralmente relacionado a detecção de um objeto em um local inválido na imagem, ou o encontro correto do local, mas predizendo erroneamente sua classe.
- Verdadeiro Negativo (VN): indica a existência de objetos que não foram preditos de maneira correta, ou seja, a resposta do classificador foi que o objeto não pertence a determinada classe e ele não pertence.
- Falso Negativo (FN): corresponde a todo GT que o modelo não conseguiu predizer.

Os VP são normalmente determinados por meio da utilização de um *threshold* de IoU. Os algoritmos que avaliam as predições definirão cada um deles como VP ou FP de acordo com o mínimo de IoU estabelecido na métrica. Por exemplo, se o limite de IoU for definido como 0,5 (50%), as caixas delimitadoras preditas serão marcadas como Verdadeiro Positivo somente se tiverem uma IoU maior ou igual ao limite (0,5). Caso contrário, será definido como FP. Em outras palavras, esse limite define a precisão espacial mínima desejada para definir uma predição como correta. Outro critério importante para definir uma predição como VP, em desafios de detecção de objetos, é que apenas uma predição pode ser dita como um *Ground truth*. Isso significa que, mesmo que três predições tivessem IoU suficiente para um único *Ground truth*, duas delas seriam marcadas como falsos positivos devido à redundância.

Com base nas medidas extraídas pela análise dos resultados de classificação na detecção de objetos, é possível calcular as métricas denominadas Sensibilidade, Precisão e Acurácia, que são comumente utilizadas para avaliar os resultados dos experimentos. A métrica Sensibilidade, definida pela Equação 12, é utilizada para indicar a relação entre as predições definidas como VP realizadas corretamente e todas as predições que realmente são positivas (*Ground truths*). A Precisão é utilizada para indicar a relação entre as predições positivas (VP) realizadas corretamente e todas as predições definidas como VP e FP (Equação 13). A Acurácia (Equação 14) é a métrica mais simples de se calcular, a qual indica a relação entre os VP e todas as outras medidas, inclusive os VP.

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

$$Precisão = \frac{VP}{VP + FP} \quad (13)$$

$$Acurácia = \frac{VP}{VP + FP + VN + FN} \quad (14)$$

Outra métrica importante é a *Average Precision* (AP). Segundo Yilmaz e Aslam (2006), o AP é uma medida estável e altamente informativa para a efetividade da recuperação, sendo uma das métricas mais utilizadas e referenciadas. O AP é calculado para cada classe individualmente no conjunto de dados, mas é comum apresentar a média sobre cada AP de cada classe, o mAP (*mean Average Precision*). Uma curva Precisão / Sensibilidade é calculada, sendo as saídas classificadas de acordo com a confiança da predição. O AP é definido pela precisão média em um conjunto de onze níveis de Sensibilidade igualmente espaçados de zero a um, que define a forma da curva Precisão / Sensibilidade (Equação 15).

$$AP = \frac{1}{11} \sum_{r \in \{0,0.1,\ldots 1\}} p_{interp}\,(r) \quad (15)$$

onde AP é a *Average Precision*; r é a Sensibilidade; p_{interp} é a precisão interpolada em cada nível de Sensibilidade.

Existem diversas variações “mAP”, como mAP-50 e mAP-70. O número no nome da métrica (50 ou 70) está relacionado ao limite mínimo aceitável pela IoU. Isso significa que a métrica mAP-50, que é a mais utilizada, é um pouco mais tolerante em

termos de qualidade de detecção do que o mAP-70. O mAP-50, empregado neste trabalho, requer que as caixas delimitadoras previstas tenham pelo menos 50% (0,5) na pontuação de IoU de acordo com as anotações de *Ground truth*, enquanto o mAP-70 é mais rígido, exigindo 70%.

3.3.2. MÉTRICAS PARA AVALIAÇÃO DO MÉTODO DE PEQUENAS PORÇÕES DE ÁGUA

A avaliação do resultado da aplicação do indicador de água gerado pelo método proposto neste trabalho, baseado em um AG, depende de uma medida que avalia a similaridade entre duas imagens. São inúmeras as medidas existentes na literatura que podem ser usadas para essa finalidade como, por exemplo, correlação, distância euclidiana, média das diferenças absolutas (*Mean Absolute Error* – MAE), ou índice de similaridade estrutural (*Structural Similarity Index* – SSIM), entre outras. Neste trabalho empregamos as últimas duas medidas, descritas a seguir, para avaliar os resultados do indicador de água proposto neste trabalho.

A medida MAE é a soma da diferença absoluta de cada pixel da imagem original (I_ORIG) e da imagem que representa o resultado esperado (I_ESP), dividido pela multiplicação das dimensões da imagem. Esse valor é expresso de acordo com a Equação 16. Valores de MAE próximos de 0,0 indicam que I_ORIG é similar a I_ESP.

$$MAE = \frac{1}{nlnc} \sum_{l=0}^{nl-1} \sum_{c=0}^{nc-1} |I\_ORIG(l,c) - I\_ESP(l,c)| \qquad (16)$$

onde nl e nc representam as dimensões da imagem; l e c representam as coordenadas dos pixels das duas imagens.

A medida $SSIM$ estima a similaridade entre a imagem original (I_ORIG) e a imagem que representa o resultado esperado (I_ESP), comparando três termos: a

luminância $Lm\ (I\_ORIG, I\_ESP)$, o contraste $Co\ (I\_ORIG, I\_ESP)$ e a estrutura $Es\ (I\_ORIG, I\_ESP)$. A medida, que varia de 0,0 a 1,0, é uma combinação multiplicativa dos três termos de acordo com a Equação 17. Valores de $SSIM$ próximos de 1,0 indicam que I_ORIG é similar a I_ESP.

$$SSIM(I\_ORIG, I\_ESP) = [Lm\ (I\_ORIG, I\_ESP)]^{\lambda}[Co\ (I\_ORIG, I\_ESP)]^{\varphi}[Es\ (I\_ORIG, I\_ESP)]^{\xi} \quad (17)$$

onde λ, φ e ξ são parâmetros que definem a importância relativa das componentes de luminância, contraste e estrutura, respectivamente.

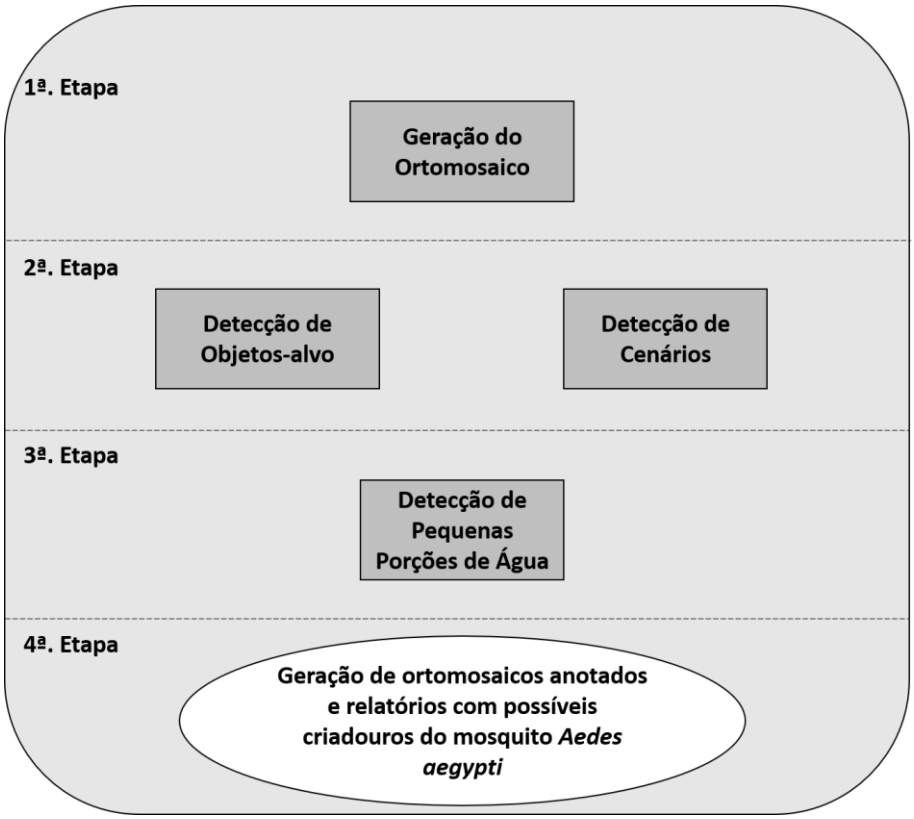
4. ABORDAGEM PROPOSTA E RESULTADOS EXPERIMENTAIS

Neste capítulo é apresentada a abordagem proposta neste trabalho, bem como a descrição dos experimentos realizados com os métodos que a compõem.

4.1. ABORDAGEM PROPOSTA

O problema investigado foi dividido em subproblemas cujas soluções, quando combinadas, levam a uma solução única para problema principal (identificação de possíveis criadouros do mosquito *Aedes aegypti*). Assim, cada uma das 4 etapas que compõe a abordagem proposta, ilustradas na Figura 37 e descritas nas seções 4.2.1 a 4.2.4, é voltada para a resolução de um subproblema.

Figura 37: Diagrama esquemático da abordagem proposta

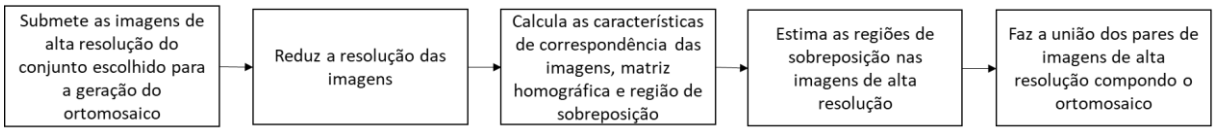


No caso de um conjunto de imagens adquiridas com o uso de um VANT em uma missão, primeiro o ortomosaico é gerado a partir das imagens. Na sequência, as imagens e o ortomosaico são submetidos às tarefas de detecção de objetos-alvo e cenários que caracterizam possíveis criadouros do mosquito. Para a terceira etapa, foi desenvolvido um método para a concepção de um índice para a detecção de pequenas porções de água. A quarta e última etapa consiste na geração de ortomosaicos anotados e relatórios com as indicações dos possíveis criadouros do mosquito. É importante destacar que as etapas 2 e 3 podem ser feitas de forma alternada ou em paralelo.

4.1.1. GERAÇÃO DO ORTOMOSAICO

Nesta etapa, para a geração do ortomosaico, as imagens aéreas adquiridas em uma mesma missão do VANT são processadas pelo método desenvolvido por Tarallo (2013), o qual foi implementado em linguagem C/C++ com o uso da plataforma Microsoft® Visual Studio™ 2015. Na Figura 38 é ilustrado o diagrama de funcionamento do método empregado para a geração dos ortomosaicos.

Figura 38: Diagrama do funcionamento do método para geração de ortomosaicos



Fonte: Adaptado de Tarallo (2013)

Ainda com relação ao método empregado, são realizados os seguintes procedimentos, de maneira automática:

- calibração radiométrica: destinada a corrigir erros esporádicos de transmissão de dados e retificar as distorções fotométricas e espaciais;
- alinhamento das imagens: é realizado o processo de fototriangulação, técnica fotogramétrica que determinar as coordenadas do terreno em relação a um referencial de terreno. O resultado é a geração da nuvem de "*tie points*" ou

pontos fotogramétricos cuja função é materializar o sistema de coordenadas do terreno.

- detecção dos pontos homólogos entre as imagens: é realizado o processamento, nos quais pontos são identificados e extraídos automaticamente das imagens. Em seguida, o algoritmo identifica pontos homólogos em novas imagens comparando pontos candidatos aos correspondentes baseando-se na distância euclidiana dos vetores de posição.
- processo de ortorretificação: as feições das imagens são projetadas ortogonalmente, com escala constante, não apresentando os deslocamentos devidos ao relevo e à inclinação da câmera.

Neste trabalho, para a composição do ortomosaico nas missões programadas para o VANT, as imagens foram adquiridas com uma sobreposição de 50%.

Em alguns casos a geração do mosaico pode não ser bem-sucedida devido às configurações de foco da câmera acoplada ao drone não estarem ajustadas adequadamente, acarretando dificuldades na detecção de pontos homólogos entre as imagens. Outro fator que pode dificultar a criação de mosaicos está relacionado às condições do tempo como, por exemplo, um dia com muito vento que pode desestabilizar o VANT durante a execução do voo para as aquisições das imagens.

4.1.2. DETECÇÃO DE OBJETOS-ALVO E CENÁRIOS

Esta etapa é responsável pela detecção dos objetos-alvo e cenários suspeitos nas imagens. Os reservatórios d’água domésticos (nos seus diversos formatos) e outros recipientes comumente usados para armazenamento de água como os tambores (metálicos ou plásticos), são os objetos que aparecem com mais frequência nas imagens aéreas das áreas urbanas mais periféricas. Além dos reservatórios d’água, outros objetos que podem se tornar criadouros do mosquito também são considerados, tais como pneus velhos, calhas e reservatórios d’água pequenos (containers), como apresentados na seção 1.2. Para os cenários foram feitas simulações de ambientes (em várias situações) contendo lixo inorgânico a céu aberto que pode ainda incluir pneus velhos e pequenos recipientes que podem acumular água, tais como baldes e embalagens plásticas e de papel.

A detecção de objetos de interesse em imagens aéreas é uma tarefa difícil em razão da quantidade de detalhes presentes nelas, principalmente em áreas urbanas. Desde a popularização do uso dos VANTs, a detecção de determinados objetos nas imagens adquiridas por esses equipamentos vem sendo um grande desafio. Nos últimos anos as redes neurais convolucionais (RNCs) têm ganhado destaque na solução de problemas desta natureza (AMMOUR et al., 2017; BEJIGA et al., 2017; XU et al., 2017).

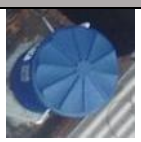
Para a detecção de objetos-alvo e cenários, nesta pesquisa foram utilizadas duas arquiteturas de RNCs pertencentes ao *framework* YOLOv3, sendo uma delas composta por 106 camadas e a outra, denominada tiny-YOLOv3 por 9 camadas.

Neste trabalho explorou-se, além das RNCs, um método que combina *Bag of Visual Words* (BoVW), com o classificador *Support Vector Machine* (SVM), denominado BoVW+SVM, para detecção de objetos e cenários, conforme descrito nas seções 4.1.2.1 a 4.1.2.3 a seguir.

4.1.2.1. DETECÇÃO DE OBJETOS-ALVO UTILIZANDO O FRAMEWORK YOLOV3

Para a detecção de objetos-alvo (reservatórios d'água e outros), foram utilizadas duas arquiteturas do *framework* YOLOv3. É importante dizer que as características dos reservatórios d'água dependem do modelo e do ano em que foram fabricados. Além disso, a matéria-prima utilizada na fabricação de tanques antigos (fibrocimento de amianto) difere muito dos tanques mais atuais (polietileno plástico). A partir da observação das imagens e do conhecimento prévio sobre os diferentes tipos de reservatórios d'água em áreas urbanas, eles foram subdivididos em classes distintas, pois cada tipo de reservatório possui características específicas (cor, tampas, formas circulares e retangulares) que os diferenciam uns dos outros. Assim, foram definidas as classes para o treinamento da RNC, denominada RNC_Detec_Obj_Reserv, de acordo com o Quadro 4.

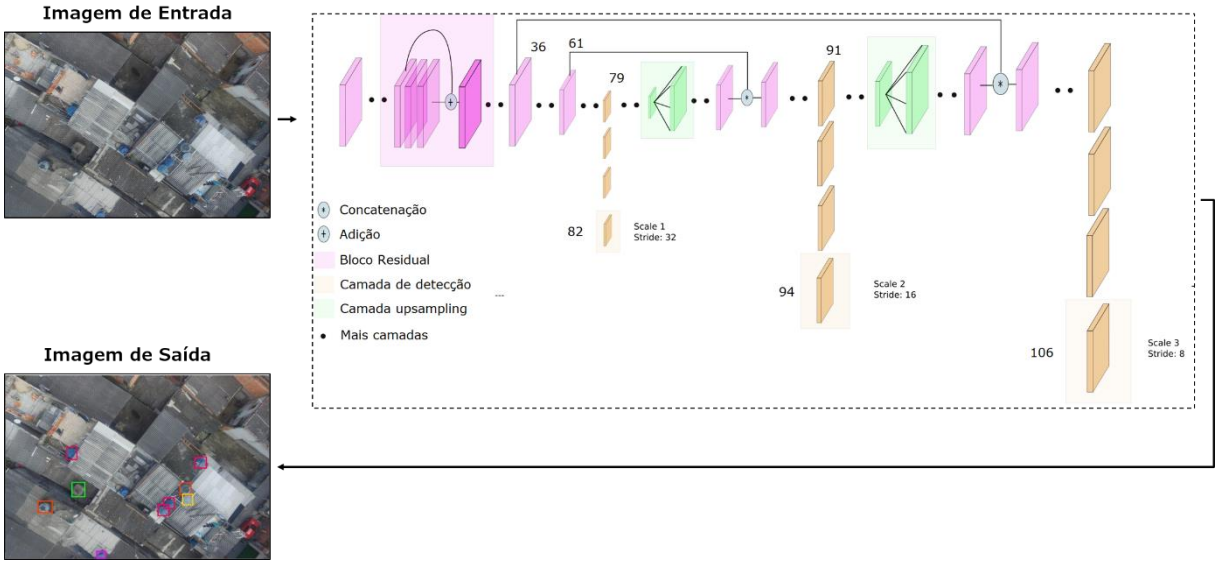
Quadro 4: Classes definidas para o treinamento da RNC_Detec_Obj_Reserv

Subimagem	ID Classe	Classe	Cor da caixa delimitadora
	0	reserv_tipo1	
	1	reserv_tipo2	
	2	reserv_tipo3	
	3	reserv_tipo4	
	4	reserv_tipo5	
	5	reserv_tipo6	
	6	reserv_tipo7	

De 106 imagens pertencentes aos conjuntos DS3 e DS5, foram extraídas, manualmente, 690 subimagens (com diversas dimensões) para compor o conjunto de treinamento. Para os testes foram destinadas 36 imagens do mesmo conjunto. Em outras palavras, 142 imagens selecionadas foram divididas em 2 partes: 75% para o treinamento e 25% para os testes. É importante salientar que o aumento dos dados (quantidade de amostras) para o treinamento é feito automaticamente a cada iteração durante o treinamento do modelo por meio da técnica *data augmentation,* incorporada ao *framework*. Além disso, as imagens de entrada são redimensionadas em determinadas iterações. A arquitetura da RNC_Detec_Obj_Reserv, ilustrada na Figura 39, é composta por 106 camadas, das quais 75 são camadas convolucionais e

as 31 restantes outras camadas (*shortcut*, *route*, *upsample*), sendo as camadas 82, 94 e 106 utilizadas para detecção de objetos em 3 diferentes escalas.

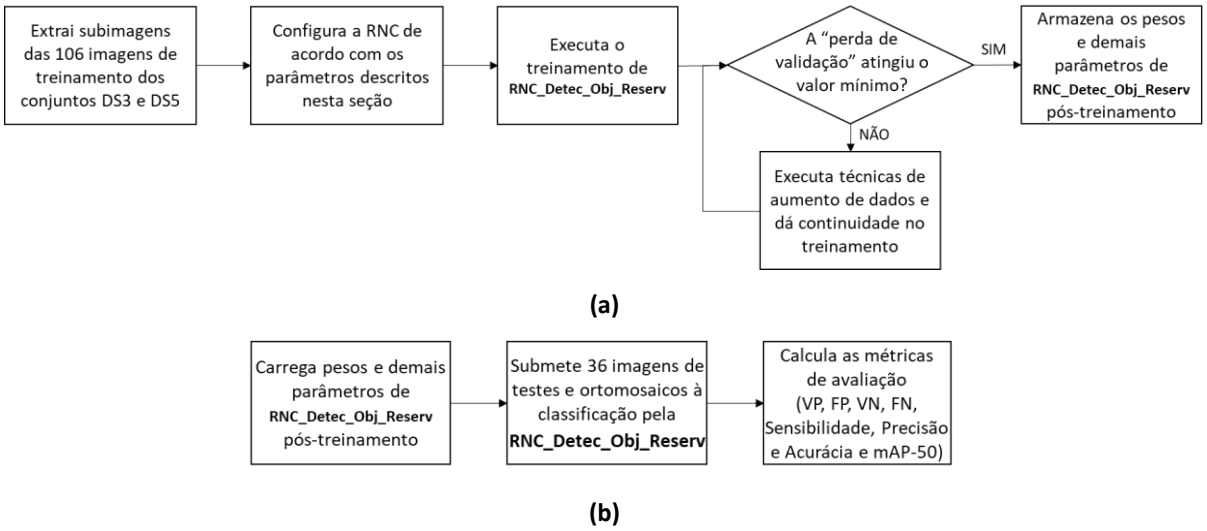
Figura 39: Arquitetura da RNC_Detec_Obj_Reserv



O *framework* YOLOv3 permite a configuração de parâmetros iniciais antes do treinamento tais como o número de batches = 64 e de subdivisões = 32, a quantidade máxima de iterações (2.000 × número de classes = 14.000), os pesos pré-treinados e a taxa de aprendizagem = 0,001. O parâmetro selecionado para parar o treinamento é baseado na falta de melhoria na “perda de validação”, que é ativada se o modelo executar mais de cinco épocas sem melhorar a perda. Após 288 horas de treinamento, totalizando 14.000 iterações, o valor da perda de validação foi de 0,19 na iteração 13.950, com cerca de 1.020.864 subimagens geradas por técnicas de aumento de dados.

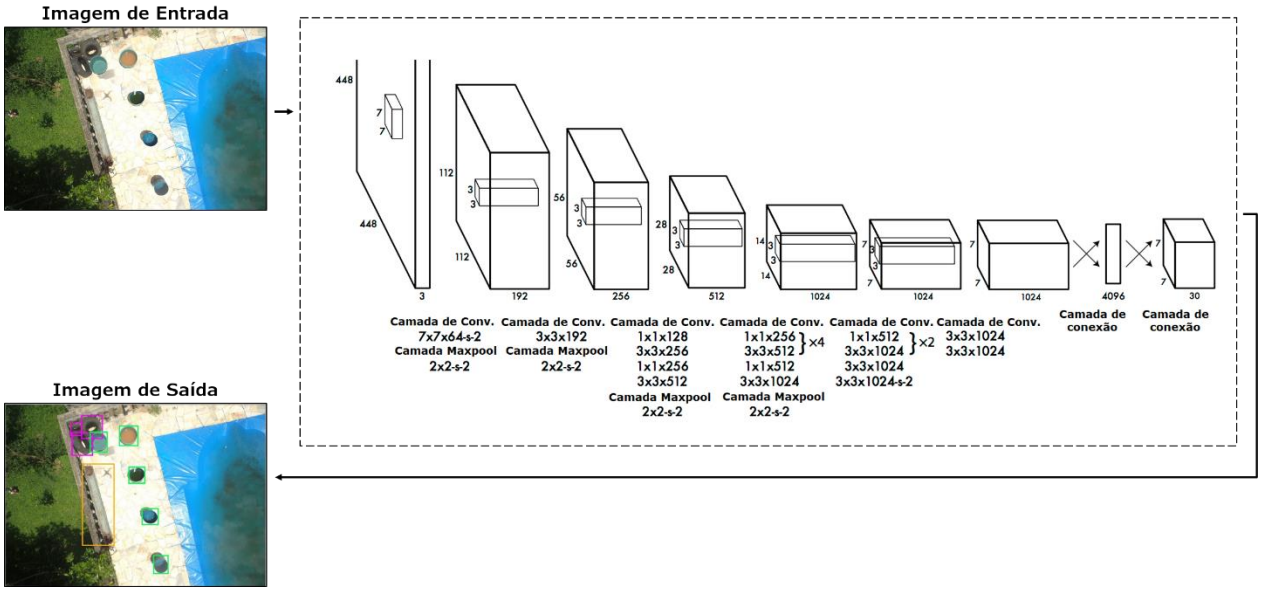
Na Figura 40 é ilustrado o diagrama de funcionamento do método para a detecção dos reservatórios d’água domésticos.

Figura 40: Diagramas do funcionamento do método para a detecção dos reservatórios d'água domésticos: (a) treinamento; (b) testes com a RNC treinada.



Para a detecção de outros objetos que também podem se tornar possíveis criadouros do mosquito, tais como pneus velhos, calhas e reservatórios d'água pequenos (containers), foi empregada a arquitetura tiny-YOLOv3, a qual é ilustrada na Figura 41, e denotada por RNC_Detec_Obj_Outros.

Figura 41: Arquitetura da RNC_Detec_Obj_Outros



Para o treinamento da RNC_Detec_Obj_Outros foram definidas as classes de acordo com o Quadro 5. O conjunto de treinamento foi composto por 500 subimagens

(com diversas dimensões) extraídas manualmente de 42 imagens RGB pertencentes ao conjunto DS4. Para o conjunto de testes foram destinadas 18 imagens. Cabe ressaltar que o aumento dos dados e o redimensionamento das imagens a cada iteração seguem o princípio geral do *framework* YOLOv3.

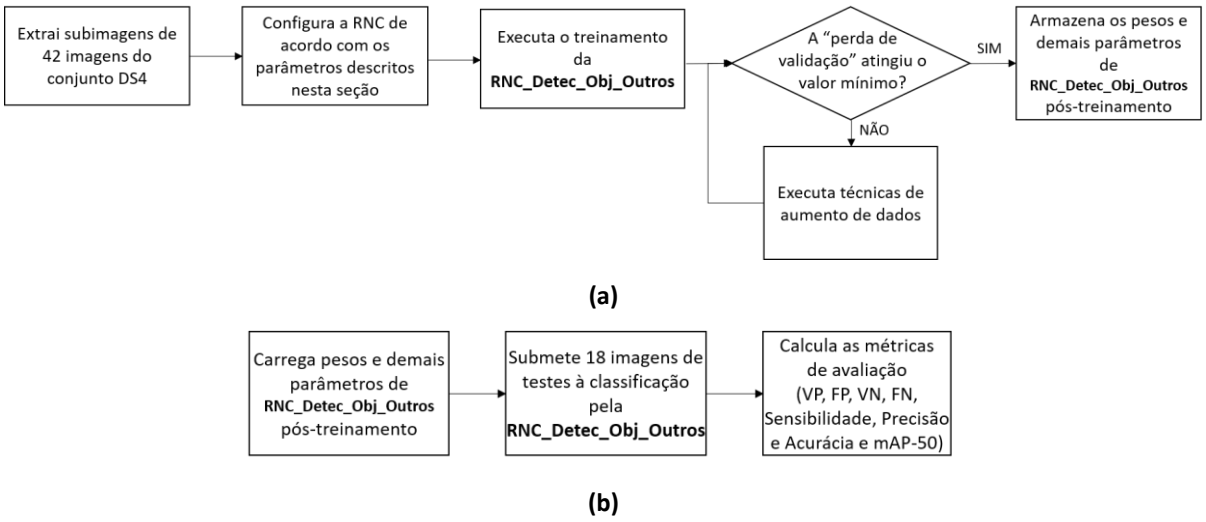
Quadro 5: Classes utilizadas para o treinamento da RNC_Detec_Obj_Outros

Subimagem	ID Classe	Classe	Cor da caixa delimitadora
	0	pneu	
	1	calha	
	2	container	

Foram configurados os seguintes parâmetros para o treinamento da RNC_Detec_Obj_Outros: número de batches = 64; número de subdivisões = 32; quantidade máxima de iterações = 6000; os pesos pré-treinados e a taxa de aprendizagem = 0,001. Após 48 horas de treinamento, totalizando 6.000 iterações, o valor da perda de validação foi de 0,27 na iteração 5.955, com cerca de 381.120 subimagens geradas por técnicas de aumento de dados.

Na Figura 42 são ilustrados os diagramas de funcionamento do método para a detecção de outros objetos-alvo.

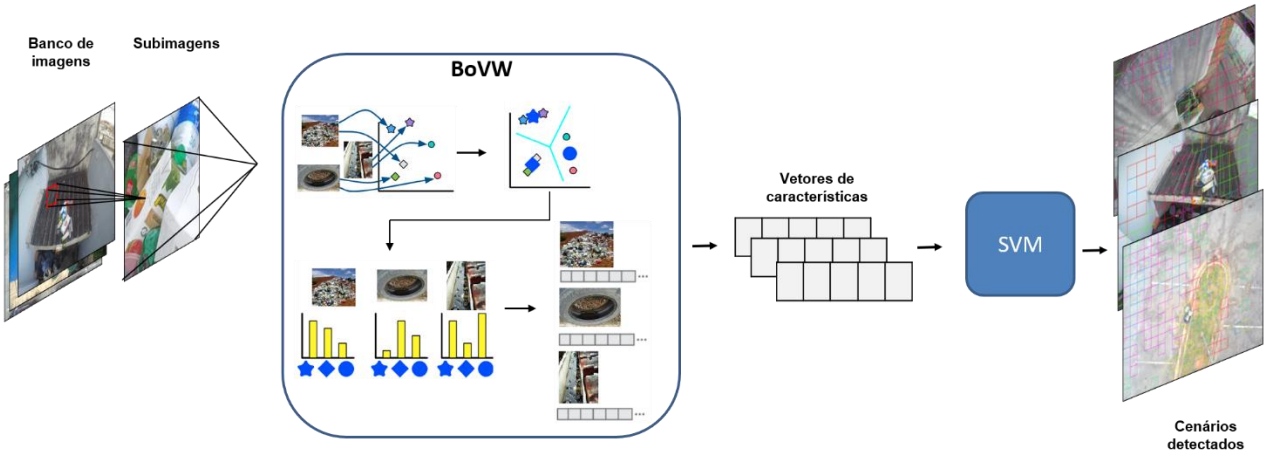
Figura 42: Diagramas do funcionamento do método para a detecção de outros objetos-alvo: (a) treinamento; (b) testes com a RNC treinada.



4.1.2.2. DETECÇÃO DE CENÁRIOS UTILIZANDO BAG OF VISUAL WORDS

Para a detecção de cenários usando BoVW combinada com o classificador SVM multiclasse (BoVW+SVM), adotou-se o esquema ilustrado na Figura 43.

Figura 43: Método BoVW+SVM utilizado para a detecção de cenários



O método BoVW+SVM desenvolvido para a detecção dos cenários consiste das seguintes etapas:

I. Extração de características das subimagens obtidas dos conjuntos DS6 e DS7;
II. Criação do dicionário de palavras visuais a partir das características, utilizando o algoritmo *k-means*;
III. Representação de cada subimagem, a partir do dicionário, por histogramas de palavras visuais (*coding*);
IV. Sintetização dos histogramas de palavras visuais em vetores de características para cada subimagem (*pooling*).
V. Treinamento do classificador SVM utilizando os vetores de características;
VI. Classificação de cada janela deslizante utilizando o classificador SVM treinado.

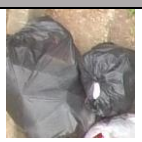
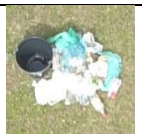
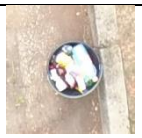
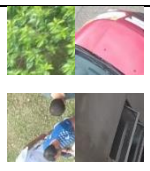
No Quadro 6 é possível observar as classes definidas para o treinamento do SVM. De 160 imagens foram extraídas características de 800 subimagens (100 para cada classe) de 100×100 pixels (subconjunto 1) e de 200×200 pixels (subconjunto 2) para compor os conjuntos de treinamento, sendo que para os testes foram destinadas 70 imagens. Em outras palavras, 230 imagens foram divididas em 2 partes: 70% para o treinamento e 30% para os testes.

Os vetores de características foram obtidos por meio da utilização dos descritores LBP, HOG, histogramas de cores com 128 bins (HIST) e do CLCM. A partir do LBP foram extraídas 2124 características, do HOG 20736 características, dos histogramas de cores 384 características e do CLCM 6 características (descritores de Haralick: segundo momento angular, entropia, contraste, variância, correlação e homogeneidade).

É importante mencionar que, para a extração das características pelo LBP, foram definidos os tamanhos 32 e 16 das células, respectivamente, para as resoluções 200×200 e 100×100, enquanto para o descritor HOG foram definidos os tamanhos 8 e 4 das células, respectivamente.

Os conjuntos de características extraídos das subimagens de resoluções 200×200 e 100×100 para cada combinação de descritores foram submetidos à técnica BoVW, gerando assim vetores de características que foram utilizados como conjuntos de treinamento para o classificador. Vale ressaltar que o tamanho do dicionário de palavras visuais foi definido, empiricamente, como 440.

Quadro 6: Classes utilizadas para o treinamento do SVM

Subimagem	ID Classe	Classe	Descrição	Cor da caixa delimitadora
	0	cenario1	Sacos de lixo fechados	
	1	cenario2	Lixo a céu aberto com pneus velhos	
	2	cenario3	Somente lixo a céu aberto	
	3	cenario4	Lixo a céu aberto com presença de reservatórios d'água pequenos (baldes)	
	4	cenario5	Lixo a céu aberto em caçambas	
	5	cenario6	Lixo a céu aberto em porta-lixos	
	6	cenario7	Sacos de lixo fechados com pneus velhos	
	7	cenario8	Reservatórios d'água contendo lixo	
	8	cenario9	Reservatórios d'água pequenos	
	9	outros	Outros materiais	

Com o objetivo de analisar qual combinação de características seria mais adequada, foi realizado um procedimento de validação cruzada, o qual resultou em

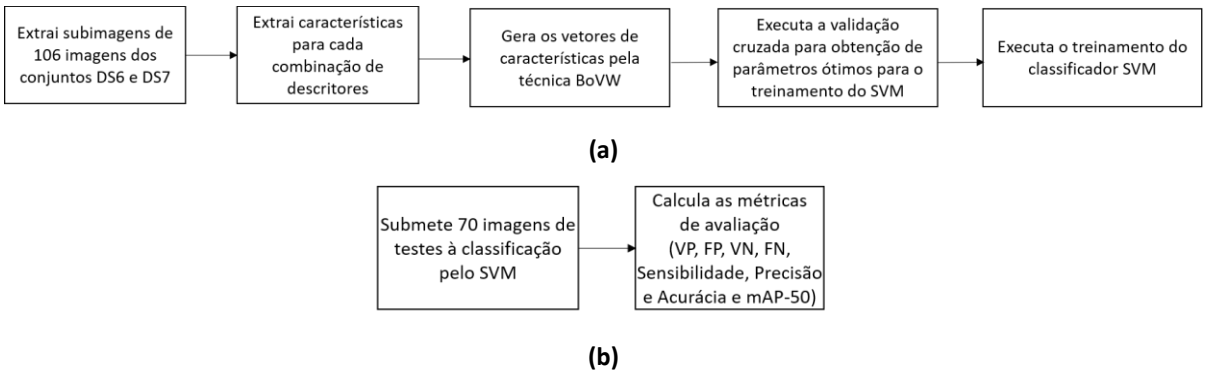
hiperparâmetros ótimos para o treinamento do SVM para as 14 combinações de descritores, de acordo com o Quadro 7. Alguns parâmetros extras (*Box Constraint, Kernel Scale* e *Polynomial Order*) foram definidos para o treinamento do classificador no software Matlab 2018.

Quadro 7: Hiperparâmetros otimizados para o SVM multiclasse

	Descritor (es)	Estratégia	Box Constraint	Kernel Scale	Kernel Function	Polin. Order
Subimagens - 200×200	LBP	Um × Um	507,79	5,5687	gaussiano	-
	HOG	Um × Um	6,6095	-	polinomial	2
	LBP+HOG	Um × Um	0,0010068	-	polinomial	2
	LBP+HOG+HIST	Um × Um	67,956	0	polinomial	2
	LBP+HIST	Um × Um	0,11276	-	polinomial	2
	HOG+HIST	Um × Todos	986,71	18,043	gaussiano	-
	LBP+HOG+HIST+CLCM	Um × Todos	798,09	16,851	gaussiano	-
	LBP+CLCM	Um × Um	28,774	-	polinomial	4
	HOG+CLCM	Um × Todos	84,548	-	polinomial	2
	LBP+HIST+CLCM	Um × Todos	979,61	-	polinomial	2
	LBPR+LBPG+LBPB	Um × Todos	21,105	4,6375	gaussiano	-
	LBPR+LBPG+LBPB+HIST	Um × Todos	951,74	-	polinomial	3
	LBPR+LBPG+LBPB+HIST+CLCM	Um × Todos	832,68	6,8074	gaussiano	-
	CLCM	Um × Um	276,88	0,038998	gaussiano	-
Subimagens - 100×100	LBP	Um × Todos	189,67	-	polinomial	3
	HOG	Um × Um	158,19	-	polinomial	2
	LBP+HOG	Um × Um	990,99	-	polinomial	2
	LBP+HOG+HIST	Um × Um	0,82338	-	linear	-
	LBP+HIST	Um × Todos	999,01	6,4534	gaussiano	-
	HOG+HIST	Um × Um	0,0010506	-	polinomial	2
	LBP+HOG+HIST+CLCM	Um × Todos	0,0010292	-	polinomial	2
	LBP+CLCM	Um × Todos	818,78	3,6502	gaussiano	-
	HOG+CLCM	Um × Todos	0,78794	-	polinomial	2
	LBP+HIST+CLCM	Um × Todos	0,0010069	-	polinomial	2
	LBPR+LBPG+LBPB	Um × Todos	3,4179	-	polinomial	2
	LBPR+LBPG+LBPB+HIST	Um × Todos	55,458	5,8009	gaussiano	-
	LBPR+LBPG+LBPB+HIST+CLCM	Um × Todos	0,0010096	-	polinomial	2
	CLCM	Um × Todos	932,31	0,026316	gaussiano	-

Na Figura 44 são ilustrados os diagramas de funcionamento do método BoVW+SVM para a detecção dos cenários considerando os subconjuntos de treinamento (subimagens 200×200 e 100×100).

Figura 44: Diagramas do funcionamento do método BoVW+SVM para a detecção de cenários: (a) treinamento; (b) testes com o classificador SVM treinado.

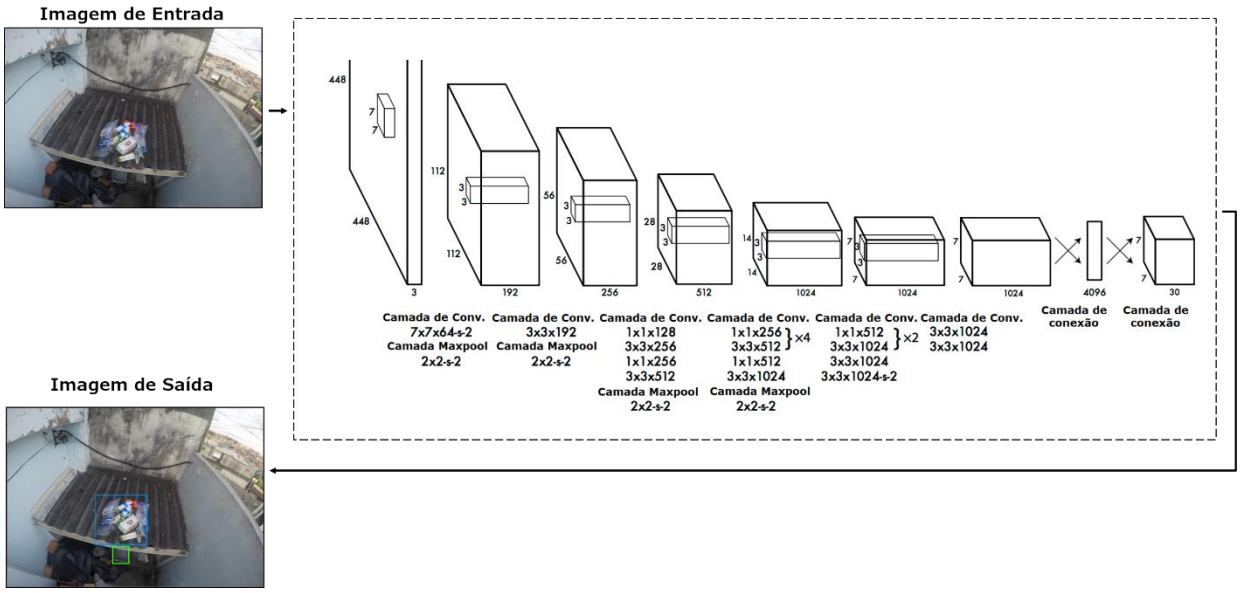


4.1.2.3. DETECÇÃO DE CENÁRIOS UTILIZANDO A ARQUITETURA TINY-YOLOV3

Para a detecção dos cenários também foi empregada uma arquitetura tiny-YOLOv3, denotada por RNC_Detec_Cenarios, e ilustrada na Figura 45. Para o treinamento desta RNC foram definidas as classes apresentadas no Quadro 8.

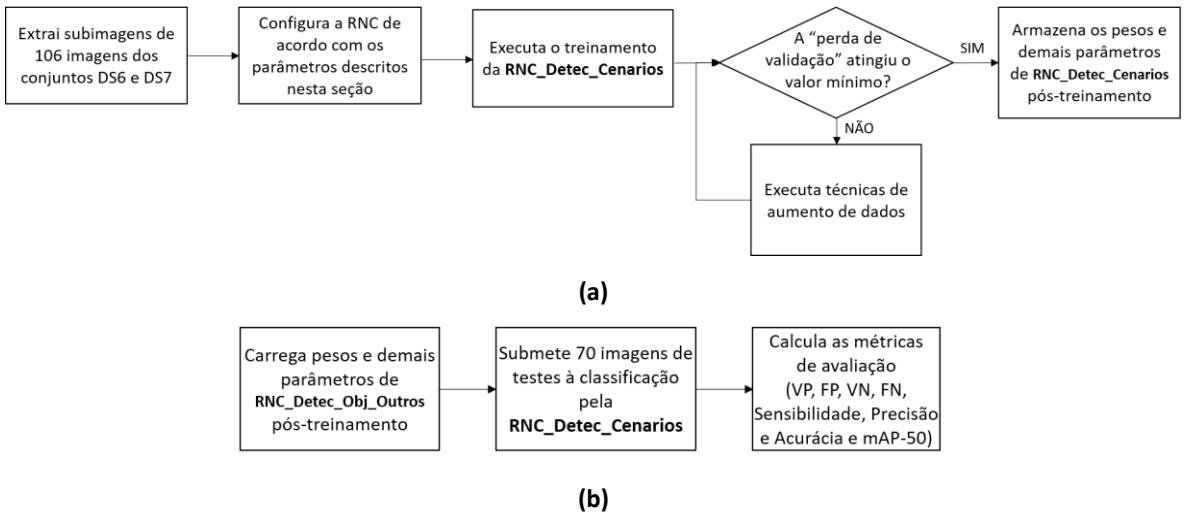
O conjunto de treinamento foi composto por 430 subimagens (com diversas dimensões) extraídas manualmente de 160 imagens RGB pertencentes aos conjuntos DS6 e DS7. Cabe ressaltar que o aumento dos dados e o redimensionamento das imagens a cada iteração seguem o princípio geral do *framework* YOLOv3. Para o conjunto de testes foram destinadas 70 imagens.

Figura 45: Arquitetura da RNC_Detec_Cenarios



Na Figura 46 são ilustrados os diagramas de funcionamento da RNC_Detec_Cenarios para a detecção dos reservatórios d'água domésticos.

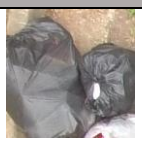
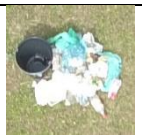
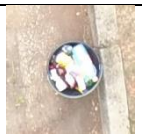
Figura 46: Diagramas do funcionamento do método para a detecção de cenários: (a) treinamento; (b) testes com a RNC treinada.



Foram configurados os seguintes parâmetros para o treinamento da RNC_Detec_Cenarios: número de batches = 64; número de subdivisões = 32;

quantidade máxima de iterações = 18.000; os pesos pré-treinados e a taxa de aprendizagem = 0,001. Após 96 horas de treinamento, totalizando 18.000 iterações, o valor da perda de validação foi de 0,06 na iteração 17.930, com cerca de 1.147.520 subimagens geradas por técnicas de aumento de dados.

Quadro 8: Classes utilizadas para o treinamento do SVM

Subimagem	ID Classe	Classe	Descrição	Cor da caixa delimitadora
	0	cenario1	Sacos de lixo fechados	
	1	cenario2	Lixo a céu aberto com pneus velhos	
	2	cenario3	Somente lixo a céu aberto	
	3	cenario4	Lixo a céu aberto com presença de reservatórios d'água pequenos (baldes)	
	4	cenario5	Lixo a céu aberto em caçambas	
	5	cenario6	Lixo a céu aberto em porta-lixos	
	6	cenario7	Sacos de lixo fechados com pneus velhos	
	7	cenario8	Reservatórios d'água contendo lixo	
	8	cenario9	Reservatórios d'água pequenos	

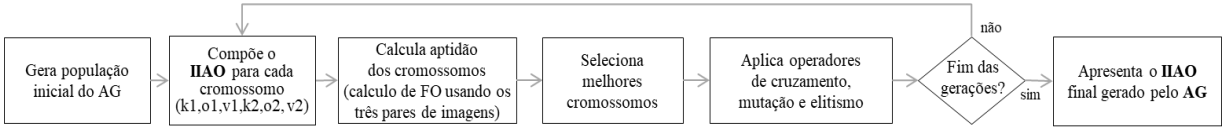
4.1.3. DETECÇÃO DE PEQUENAS PORÇÕES DE ÁGUA

A finalidade desta etapa é detectar a presença de água em reservatórios d'água domésticos, tambores e outros recipientes destampados, os quais representam os principais criadouros, pois podem acumular água parada. Para tanto, foi proposto um método para criação de um índice indicativo de pequenas porções d'água.

Em sensoriamento remoto, são utilizados vários índices indicadores, os quais permitem extrair diversas feições de imagens digitais adquiridas por satélites, inclusive corpos d'água. Murugan et al. (2016) propuseram um algoritmo para seleção de conjuntos ótimos de bandas espectrais visando a criação de índices indicativos de corpos d'água e vegetação em imagens de satélites. Tal algoritmo faz uma busca exaustiva em 242 bandas espectrais disponíveis, o que pode requerer um alto custo computacional.

Contudo, nem sempre os índices indicadores propostos para imagens adquiridas por satélites podem ser utilizados em imagens adquiridas por VANTs, devido à questão da limitação das bandas espectrais, já explicada anteriormente. Analisando tal situação, foi proposto um método para geração de um índice, denominado Índice Indicativo de Água Otimizado ($IIAO$), que pode ser aplicado em imagens aéreas adquiridas por VANTs. Neste método um Algoritmo Genético (AG), cujo Diagrama de funcionamento é apresentado na Figura 47, foi utilizado para fornecer o $IIAO$. É importante destacar que não se trata da otimização de um indicador da literatura, mas da criação de um novo indicador específico para imagens adquiridas por VANTs.

Figura 47: Passos do AG desenvolvido para fornecer o IIAO



Cada cromossomo do AG codifica um conjunto de 6 valores ($v_1, k_1, v_2, k_2, o_1\ e\ o_2$) que compõem um índice a partir de uma combinação aritmética, como descrito na equação 18.

$$\frac{k_1 o_1 v_1 - k_2 o_2 v_2}{k_1 o_1 v_1 + k_2 o_2 v_2} \quad (18)$$

em que $k_1 \, e \, k_2 \in \{1,2,3,4\}$ indicam as bandas selecionadas do conjunto Ω de bandas espectrais disponíveis (bandas do espectro visível e infravermelho), sendo $k_1 \neq k_2$; $v_1 \, e \, v_2 \in [0.1,10.0]$ indicam os valores considerados na operação aritmética e, por fim, $o_1 \, e \, o_2 \in \{+, -, *, /, \hat{}\}$ as operações aritméticas realizadas. Vale dizer que, redimensionando o cromossomo por meio do aumento da quantidade de bandas, dos operadores matemáticos e de constantes, o número de possibilidades para o cômputo do $IIAO$ tende a crescer exponencialmente, fato esse que inviabilizaria a mesma tarefa se fosse realizada por força bruta.

Por exemplo, supondo $\Omega = \{NIR - near\ infrared,\ R - red,\ G - green\ e\ \ B - blue\}$, $k_1 = 1\ (NIR)$, $k_2 = 2\ (R)$, $v_1 = v_2 = 1$ e $o_1 = o_2 = '*'$, tem-se o *Normalized Difference Vegetation Index* – NDVI $(\frac{NIR - R}{NIR + R})$, amplamente utilizado no estudo e avaliação de vegetação.

Cada solução é representada por um conjunto de 6 valores $(k_1, o_1, v_1, k_2, o_2, v_2)$ que compõem o índice e será avaliada por uma função objetivo (FO) (Equação 19), a qual consiste em minimizar a similaridade total SL entre M pares de imagens, cada um deles consistindo de uma imagem gerada com aplicação do $IIAO$ codificado em um solução do AG (I_IIAO) e outra imagem binária anotada manualmente para indicar o resultado esperado (I_ESP).

$$Minimize\ f(v_1, o_1, k_1, v_2, o_2, k_2) = \sum_{j=1}^{M} SL(I\_IIAO_j, I\_ESP_j) \quad (19)$$

Neste trabalho consideramos o *Mean Absolute Error* (MAE), descrito na Equação 20, para medir a similaridade (SL) entre os pares de imagens:

$$SL(I_{IIAO}, I_{ESP}) = \frac{1}{M} \sum_{i=0}^{nl-1} \sum_{j=0}^{nc-1} \left| I\_IIAO_{i,j} - I\_ESP_{i,j} \right| \quad (20)$$

onde nl e nc são as dimensões das imagens comparadas. A medida de similaridade SL pode ser, por exemplo, correlação, soma ou média das diferenças absolutas, distância euclidiana ou índice de similaridade estrutural (*Structural Similarity Index* – $SSIM$), entre outras.

Neste trabalho, a detecção de pequenas porções de água a partir de um par de imagens VIS – NIR é realizada pelo cálculo da Equação 18 ($IIAO$). Este procedimento gera uma imagem em tons de cinza I_GRAY, na qual os pixels são valores reais que variam de 0,0 a 1,0, sendo as porções de água representadas por níveis de cinza próximos de zero (preto).

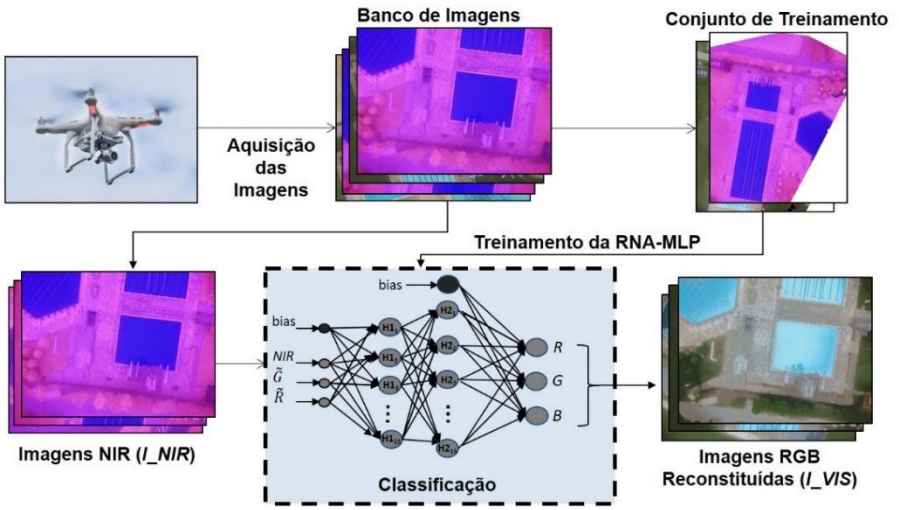
O passo final consiste em converter I_GRAY em uma imagem binária I_BIN por meio do algoritmo de limiarização descrito na Equação 1 usando um limiar $T = 0{,}85$ (obtido empiricamente), sendo que os pixels de I_BIN com valor 0 indicam a presença água.

É importante destacar que uma alternativa mais econômica para a aquisição de imagens contendo a banda infravermelho, que é importante para a detecção de água, é a utilização de uma lente especial acoplada à câmera RGB. Entretanto, o uso da lente para a filtragem da banda espectral infravermelho próximo possui maiores larguras de banda e capta os dados em diferentes comprimentos de onda, causando a contaminação (não linear) das informações de reflectância nas bandas G e R. Para resolver este problema, foi desenvolvido um método para “descontaminação” das bandas espectrais afetadas, o qual é apresentado a seguir.

4.1.3.1. RECONSTITUIÇÃO DAS BANDAS ESPECTRAIS

O método proposto para descontaminar as bandas G e R, denotadas por $\tilde{G}$ e $\tilde{R}$, emprega uma RNA do tipo MLP para reconstituir uma imagem no espectro visível (I_VIS) a partir da imagem infravermelho adquirida com o uso da lente (I_NIR). O funcionamento do método, ilustrado no diagrama da Figura 48, se dá em 2 etapas: extração do conjunto de treinamento e mapeamento das bandas espectrais pela rede neural denominada RNA_MLP_Recons.

Figura 48: Diagrama esquemático do método proposto para reconstituição de bandas espectrais



Extração do Conjunto de Treinamento: para compor o conjunto de dados para treinamento da RNA_MLP_Recons, foram extraídas, manualmente, 60 subimagens de 50×50 pixels de 5 pares de imagens correlacionadas VIS – NIR. Em resumo, para cada pixel escolhido em uma imagem RGB, foi extraída uma subimagem em torno deste pixel e de seu correspondente na imagem NIR, na mesma posição. Assim, cada instância do treinamento é descrita pelos seguintes atributos: NIR, $\tilde{G}$, $\tilde{R}$, R, G, B, sendo os três primeiros extraídos de I_NIR (entradas) e os três últimos de subimagens I_VIS (saídas esperadas).

Mapeamento das bandas espectrais: Formalmente, a função da RNA_MLP_Recons é mapear cada conjunto $\{NIR(l,c), \tilde{G}(l,c), \tilde{R}(l,c)\}$ em outro conjunto $\{R(l,c), G(l,c), B(l,c)\}$. Dessa forma, todos os conjuntos mapeados compõem os pixels da imagem RGB reconstituída. Este procedimento é importante porque evita a realização de duas missões completas para aquisição das imagens RGB e NIR da mesma área, impactando diretamente no tempo gasto na tarefa de aquisição de imagens, bem como economizando a bateria do drone. É válido enfatizar que o $IIAO$ proposto depende das bandas espectrais VIS e NIR.

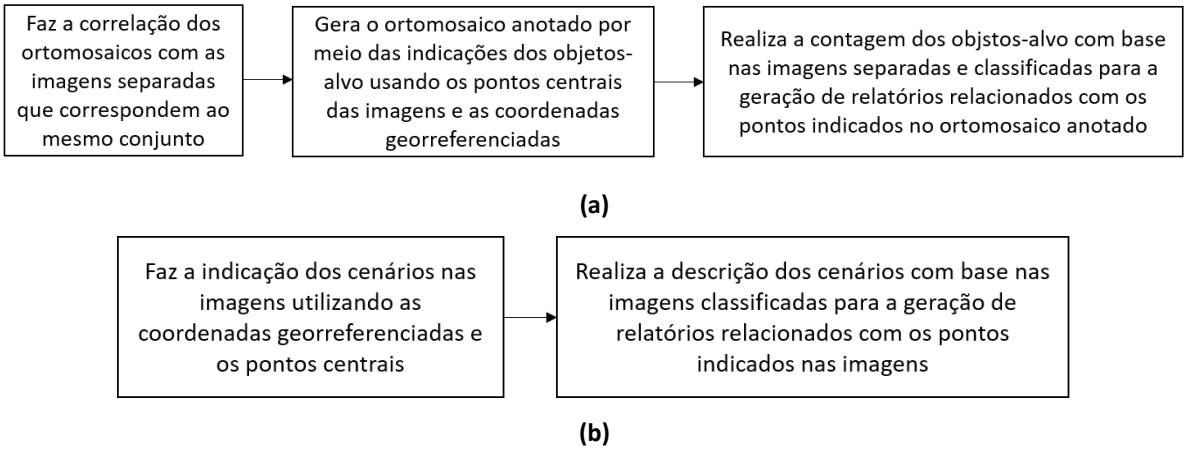
4.1.4. GERAÇÃO DE ORTOMOSAICOS ANOTADOS E RELATÓRIOS COM POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI

O objetivo desta etapa é a geração de ortomosaicos anotados e relatórios com a indicação de locais que representam possíveis criadouros do mosquito *Aedes aegypti*. Com base nas janelas que demarcam os objetos-alvo e cenários nas imagens, nos ortomosaicos e nas coordenadas centrais (latitude e longitude) anotadas nas imagens é possível gerar o relatório com os possíveis criadouros, com suas respectivas coordenadas georreferenciadas que podem ser importantes para nortear os procedimentos de combate aos focos do mosquito *Aedes aegypti*.

Em imagens adquiridas por VANTs, todas as informações a respeito dos parâmetros da câmera utilizada, bem como das coordenadas georreferenciadas, altura absoluta e relativa, dentre outras, ficam registradas nos cabeçalhos dos arquivos. As coordenadas georreferenciadas (latitude e longitude) referem-se aos pontos centrais das imagens adquiridas.

Na Figura 49a é ilustrado o diagrama de funcionamento do método para a geração de ortomosaicos anotados e relatórios para os objetos-alvo e, na Figura 49b, o diagrama de funcionamento do método para as imagens e relatórios para os cenários.

Figura 49: Diagramas de funcionamento do método para a geração de ortomosaicos anotados e relatórios: (a) ortomosaicos anotados e relatórios para objetos-alvo; (b) imagens anotadas e relatórios para cenários



4.2. EXPERIMENTOS CONDUZIDOS COM A ABORDAGEM PROPOSTA

4.2.1. GERAÇÃO DO ORTOMOSAICO

Conforme descrito na seção 4.1.1, para a construção dos ortomosaicos foi empregado o método desenvolvido por Tarallo (2013), que realiza, de forma automática, os procedimentos de calibração radiométrica, alinhamento das imagens, detecção de pontos homólogos e ortorrertificação. Para avaliar o referido método foram gerados os seguintes mosaicos:

a) Ortomosaico_Guaianases_1 (Figura 50b): obtido a partir de 33 imagens pertencentes ao conjunto DS3 – aquisição 1 (Figura 50a);
b) Ortomosaico_Guaianases_2 (Figura 50d): obtido a partir de 107 imagens pertencentes ao conjunto DS3 – aquisição 2 (Figura 50c);
c) Ortomosaico_PortoAreia (Figura 50f): obtido a partir de 101 imagens pertencentes ao conjunto ao conjunto DS5 (Figura 50e).

Dentre as vantagens da utilização de tal método para a geração dos ortomosaicos destacam-se as metodologias empregadas no pré-processamento das

imagens para minimizar possíveis distorções que surgem no processo de aquisição de imagens pelo uso de algoritmos para suavização das emendas das imagens que compõem o ortomosaico, bem como a questão do processamento paralelo que visa a reduzir o tempo de processamento para a construção dos ortomosaicos.

Figura 50: Ortomosaicos gerados com imagens dos conjuntos DS3 (aquisição1), DS3 (aquisição 2) e DS5

(a) (b)

(c) (d)

(e) (f)

É importante mencionar que, posteriormente, os ortomosaicos foram utilizados para a detecção de objetos-alvo.

4.2.2. DETECÇÃO DE OBJETOS-ALVO E CENÁRIOS

4.2.2.1. DETECÇÃO DE OBJETOS-ALVO UTILIZANDO O FRAMEWORK YOLOV3

Para a realização dos experimentos para a detecção dos objetos-alvo foi feita uma adaptação da arquitetura LeNet-5 de Rede Neural Convolucional (RNC). Os resultados da classificação pela solução proposta foram submetidos à métrica mAP-50 e, nos experimentos, o maior valor obtido foi de 0,2789, com janelas deslizantes de 80% de sobreposição para cada imagem, *threshold* do NMS de 0,40 e o escore maior ou igual a 50%. Tendo em vista o desempenho ruim na classificação, optou-se por utilizar o *framework* YOLOv3, que é mais robusto e mais rápido.

Para aferir a precisão da RNC_Detec_Obj_Reserv para a detecção dos reservatórios d'água domésticos, 36 imagens pertencentes aos conjuntos DS3 e DS5 foram submetidas à tarefa de classificação que resultou em 213 detecções após 21 segundos de processamento.

Nas Figura 51 (b), (d) e (f) é possível notar que todas os reservatórios d'água domésticos foram detectadas corretamente, caracterizando os casos de VP. Do total de 152 caixas delimitadoras *Ground truth*, 148 foram classificadas corretamente.

Foram computados 16 casos de FP e 4 de FN. Na Figura 52b é possível identificar um caso de FP, destacado com um círculo vermelho, além de um caso de FN na Figura 52d e outros três na Figura 52f, destacados com círculos pretos.

Vale ressaltar que um fator que contribuiu para a ocorrência de FP foi a faixa de valores de escala adotada para redimensionamento das imagens durante o treinamento da RNC. Mesmo considerando tal redimensionamento, alguns reservatórios d'água domésticos (circulados em preto) presentes na Figura 52f não foram detectados. Vale lembrar que tal imagem pertence à segunda aquisição (conjunto DS3), cuja altura do drone ao solo foi de 50 m.

Figura 51: Resultados das detecções dos reservatórios d'água domésticos pela RNC_Detec_Obj_Reserv

Figura 52: Resultados das detecções dos reservatórios d'água domésticos pela RNC_Detec_Obj_Reserv

Com base nos resultados de classificação da detecção de objetos, foi possível calcular as métricas Sensibilidade e Precisão, as quais são, respectivamente, 0,9700 e 0,9000, que demonstraram que o bom desempenho da arquitetura da RNC_Detec_Obj_Reserv.

Para as 36 imagens classificadas, obteve-se o valor de 0,9651 para o mAP-50, considerando as APs (*Average Precisions*) para cada classe, as quais podem ser observadas na Tabela 1.

Tabela 1: APs calculadas para cada classe da RNC_Detec_Obj_Reserv

Classe	**AP (Average Precision)**
reserv_tipo1	0,9734
reserv_tipo2	0,9933
reserv_tipo3	0,9763
reserv_tipo4	1,0000
reserv_tipo5	0,9040
reserv_tipo6	0,9091
reserv_tipo7	1,0000
mAP-50	**0,9651**

Os resultados demonstrados na Tabela 1 reforçam o bom desempenho da RNC_Detec_Obj_Reserv na detecção dos objetos-alvo. Seu pior desempenho (0,9040) ocorreu na detecção dos reservatórios d'água domésticos do tipo 5, provavelmente devido à sua baixa ocorrência nas imagens analisadas.

Cabe ressaltar que, em situações nas quais existiam reservatórios d'água domésticos muito próximos uns dos outros, a RNC_Detec_Obj_Reserv conseguiu detectá-los individualmente, na maioria dos casos.

Obviamente, para validar a segunda etapa da abordagem proposta, a detecção dos objetos-alvo foi também aplicada nos ortomosaicos. Para tanto, foi necessária a alteração de parâmetros tais como a dimensão da imagem de entrada para a classificação. Nas Figuras 53, 54a e 55a são ilustrados os ortomosaicos, com alguns objetos-alvo detectados, gerados a partir das imagens do conjunto DS3 (aquisição 1), DS3 (aquisição 2) e DS5.

Figura 53: Resultados das detecções dos reservatórios d’água no Ortomosaico_Guaianases_1

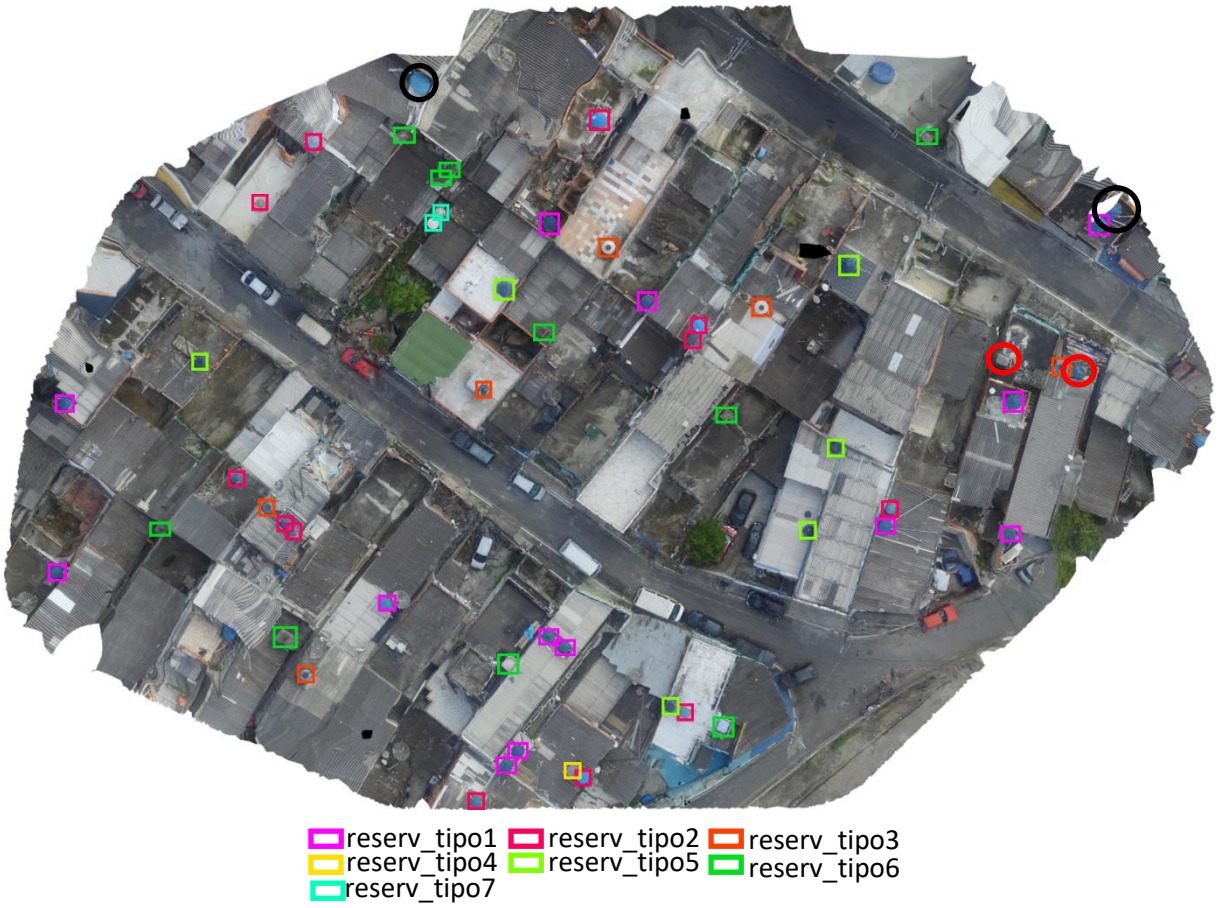


Figura 54: Resultados das detecções dos reservatórios d'água no Ortomosaico_Guaianases_2: (a) ortomosaico completo; (b) subimagens do ortomosaico com os objetos-alvo detectados

(a)

(b)

Foram registrados muitos casos de FN no Ortomosaico_PortoAreia e alguns deles são ilustrados na Figura 55b. Isso deve-se ao fato do redimensionamento das imagens (em 3 escalas) durante o treinamento da RNC do *framework* YOLOv3 não ser suficiente para a detecção de objetos muito pequenos. Diante disso, é importante frisar que a tarefa de detecção de objetos-alvo, em especial os reservatórios d'água domésticos, teve melhor desempenho em imagens adquiridas a uma distância do solo que varia de 20 a 40 m.

É importante destacar também que se um objeto a ser detectado está presente em mais de uma imagem que compõe o ortomosaico, ou seja, está particionado em duas imagens, por exemplo, o procedimento de junção das imagens pode acarretar na deformação de tal objeto, prejudicando a sua detecção. Exemplos desta situação estão destacados com círculos pretos nas Figuras 53 e 54b. Já os círculos vermelhos destacados nas Figuras 53, 54b e 55b caracterizam casos de FN. Sendo assim, justificam-se os experimentos envolvendo a submissão das imagens separadamente para a classificação com o objetivo de minimizar a ocorrência de FN que podem surgir na classificação do ortomosaico.

Figura 55: Resultados das detecções dos reservatórios d'água no Ortomosaico_PortoAreia: (a) ortomosaico completo; (b) subimagem do ortomosaico com os objetos-alvo detectados

(a)

(b)

Para avaliar o desempenho da RNC_Detec_Obj_Outros para classificação de outros objetos que também representam possíveis criadouros do mosquito, tais como pneus velhos, calhas e reservatórios d'água pequenos (containers), 18 imagens oriundas do conjunto DS4 foram classificadas, em 7 segundos, resultando em 150 detecções.

Das 114 caixas delimitadoras definidas como *Ground truth*, 113 caracterizaram casos de VP. Foram computados 1 caso de FN e 13 casos de FP. Na Figura 56b é possível perceber que todos os objetos foram detectados corretamente (VP). Na Figura 56d é possível identificar um caso de FN, que é destacado com um círculo preto e um caso de FP (Figura 56f), destacado com círculo vermelho. Para as métricas Sensibilidade e Precisão foram obtidos os valores de 0,9900 e 0,9000, respectivamente, o que sinaliza um bom desempenho da RNC.

Para as 20 imagens classificadas, obteve-se o valor de 0,9570 para o mAP-50, considerando as APs (*Average Precisions*) para cada classe, as quais podem ser observadas na Tabela 2.

Tabela 2: APs calculadas para cada classe da RNC_Detec_Obj_Outros

Classe	**AP (Average Precision)**
pneu	0,8826
calha	1,0000
container	0,9886
mAP-50	**0,9570**

Figura 56: Resultados das detecções dos cenários pela RNC_Detec_Obj_Outros

Os resultados demonstrados na Tabela 2 reforçam que o desempenho da RNC_Detec_Obj_Outros foi excelente na detecção de objetos diversos. Seu pior desempenho (0,8826) ocorreu na detecção dos pneus, pois nas imagens esses objetos sempre estavam encostados uns nos outros, podendo descaracterizá-los. Um exemplo disso é a ocorrência de um caso de FP assinalado com um círculo vermelho na Figura 56f. Ressalta-se também que o valor máximo obtido na detecção de calhas (AP = 1,0000) foi decorrente do fato de haver somente uma calha presente nas imagens adquiridas.

4.2.2.2. DETECÇÃO DE CENÁRIOS UTILIZANDO BOVW+SVM

A partir dos vetores de características extraídos por meio da técnica BoVW de subimagens de 100×100 e 200×200 pixels, foram realizados os treinamentos utilizando o classificador SVM, considerando as 14 combinações de descritores citadas na seção 4.1.2.2. Em seguida, foram feitas as classificações de 70 imagens utilizando a estratégia de janelas deslizantes com as mesmas dimensões das subimagens, o que demandou 72 horas. Empiricamente, foi definido o valor limiar de 0,60 para a probabilidade posterior, que é calculada para a classe predita. Dessa forma, somente foram consideradas as caixas delimitadoras classificadas com o valor de probabilidade maior do que esse limiar.

Com o intuito de aferir a precisão da detecção dos cenários, foi realizado o cálculo do mAP-50 para diferentes combinações de descritores e tamanhos de janela, sendo os valores ilustrados na Tabela 3, na qual é possível observar que o maior valor de mAP-50 foi de 0,6453 na classificação realizada com janelas deslizantes de dimensão 200×200 para a combinação LBPR+LBPG+LBPB+HIST. Destaca-se que BoVW reduziu o tamanho do vetor de características de 6.756 para 5.404.

De 5.190 detecções, foram computados 1.638 casos de VP e 2.352 casos de FP, caracterizando um número muito alto de FP, levando ao valor mediano obtido pelo mAP-50.

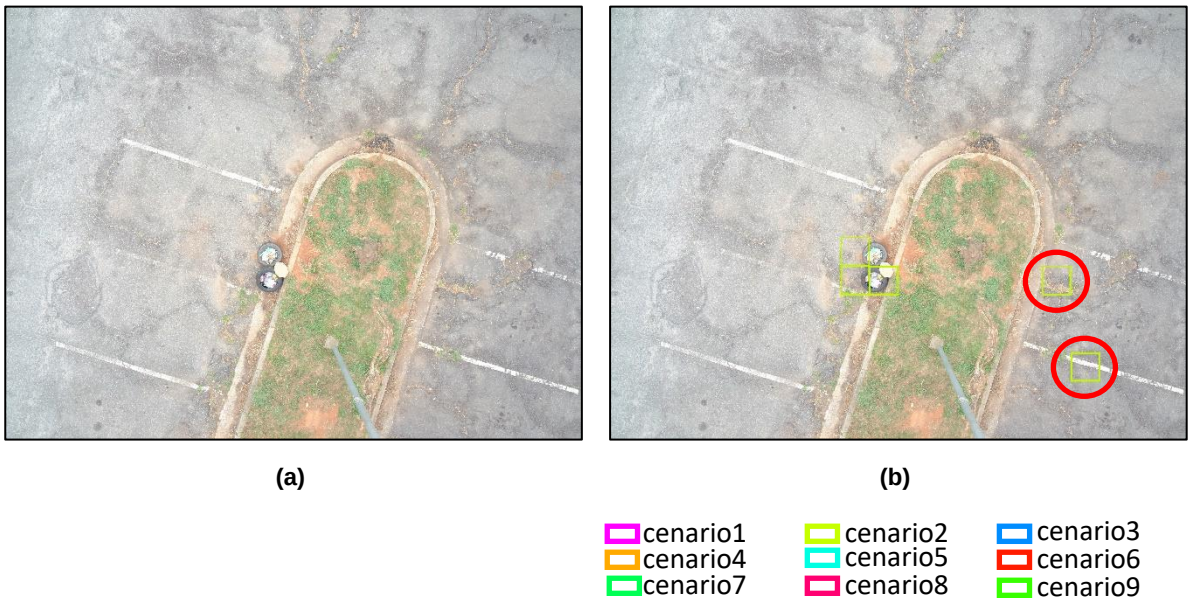
.

Tabela 3: mAPs-50 calculados para cada combinação de descritores

	Descritor (es)	mAP-50
Janela deslizante - 200×200	CLCM	0,4117
	LBP	0,3595
	HOG	0,4173
	LBP+HOG	0,4443
	LBP+HOG+HIST	0,5902
	LBP+HIST	0,4946
	HOG+HIST	0,5787
	LBP+HOG+HIST+CLCM	0,5740
	LBP+CLCM	0,5069
	HOG+CLCM	0,4256
	LBP+HIST+CLCM	0,6131
	LBPR+LBPG+LBPB	0,5019
	LBPR+LBPG+LBPB+HIST	**0,6453**
	LBPR+LBPG+LBPB+HIST+CLCM	0,6353
Janela deslizante - 100×100	CLCM	0,3105
	LBP	0,3733
	HOG	0,3589
	LBP+HOG	0,3705
	LBP+HOG+HIST	0,3904
	LBP+HIST	0,4291
	HOG+HIST	0,3906
	LBP+HOG+HIST+CLCM	0,4088
	LBP+CLCM	0,3820
	HOG+CLCM	0,3843
	LBP+HIST+CLCM	0,4398
	LBPR+LBPG+LBPB	0,3714
	LBPR+LBPG+LBPB+HIST	0,4680
	LBPR+LBPG+LBPB+HIST+CLCM	0,4133

Na Figura 57 é mostrado um exemplo com a imagem de entrada e a imagem de saída com as janelas classificadas, utilizando janelas deslizantes 200×200 e a combinação LBPR+LBPG+LBPB+HIST.

Figura 57: Resultados do processo de classificação da SVM: (a) imagem de entrada; (b) imagem classificada



Para as 70 imagens classificadas, obteve-se o valor de 0,6453 para o mAP-50, considerando as APs (*Average Precisions*) para cada classe, as quais podem ser observadas na Tabela 4.

Tabela 4: APs calculadas para cada classe do método BoVW+SVM

Classe	**AP (Average Precision)**
cenario1	0,6509
cenario2	0,5710
cenario3	0,6544
cenario4	0,6700
cenario5	0,4561
cenario6	0,8511
cenario7	0,7530
cenario8	0,5502
cenario9	0,6510
mAP-50	**0,6453**

Na Figura 57b pode-se observar que a maioria das janelas classificadas para a classe cenario2 caracterizaram casos de VP. Dois casos de FP, destacados com círculos vermelhos, também podem ser observados. Por outro lado, no restante das imagens classificadas, houve muitas ocorrências de FP, indicando que BoVW+SVM, com a configuração adotada neste trabalho, não é o método mais adequado para a solução do problema investigado.

4.2.2.3. DETECÇÃO DE CENÁRIOS UTILIZANDO A ARQUITETURA TINY-YOLOV3

A classificação usando RNC_Detec_Cenarios foi aplicada a 70 imagens oriundas dos conjuntos DS6 e DS7. Esta tarefa foi resolvida em 18 segundos, resultando em 154 detecções. Das 111 caixas delimitadoras definidas como *Ground truth*, 96 foram corretamente classificadas, caracterizando casos de VP, alguns dos quais ilustrados nas Figuras 58b, 58d, 59b e 59d.

Cabe ressaltar que, como ilustrado na Figura 58d, a RNC_Detec_Cenarios obteve um bom resultado mesmo com a maior distância do VANT em relação ao solo. Nesse contexto, é importante dizer que, para um bom desempenho na tarefa de detecção dos cenários, a distância do drone ao solo deve ser de 1 a 15 m.

Foram computados 15 casos de FN e 11 casos de FP. Na Figura 60b é possível identificar dois casos de FP, que são destacados com círculos vermelhos e quatro casos de FN (Figura 60d e Figura 60f), destacados com círculos pretos. As métricas Sensibilidade e Precisão apresentaram os valores de 0,8600 e 0,9000, respectivamente, corroborando o bom desempenho da RNC_Detec_Cenarios.

Figura 58: Resultados das detecções dos cenários pela RNC_Detec_Cenarios

Figura 59: Resultados das detecções dos cenários pela RNC_Detec_Cenarios

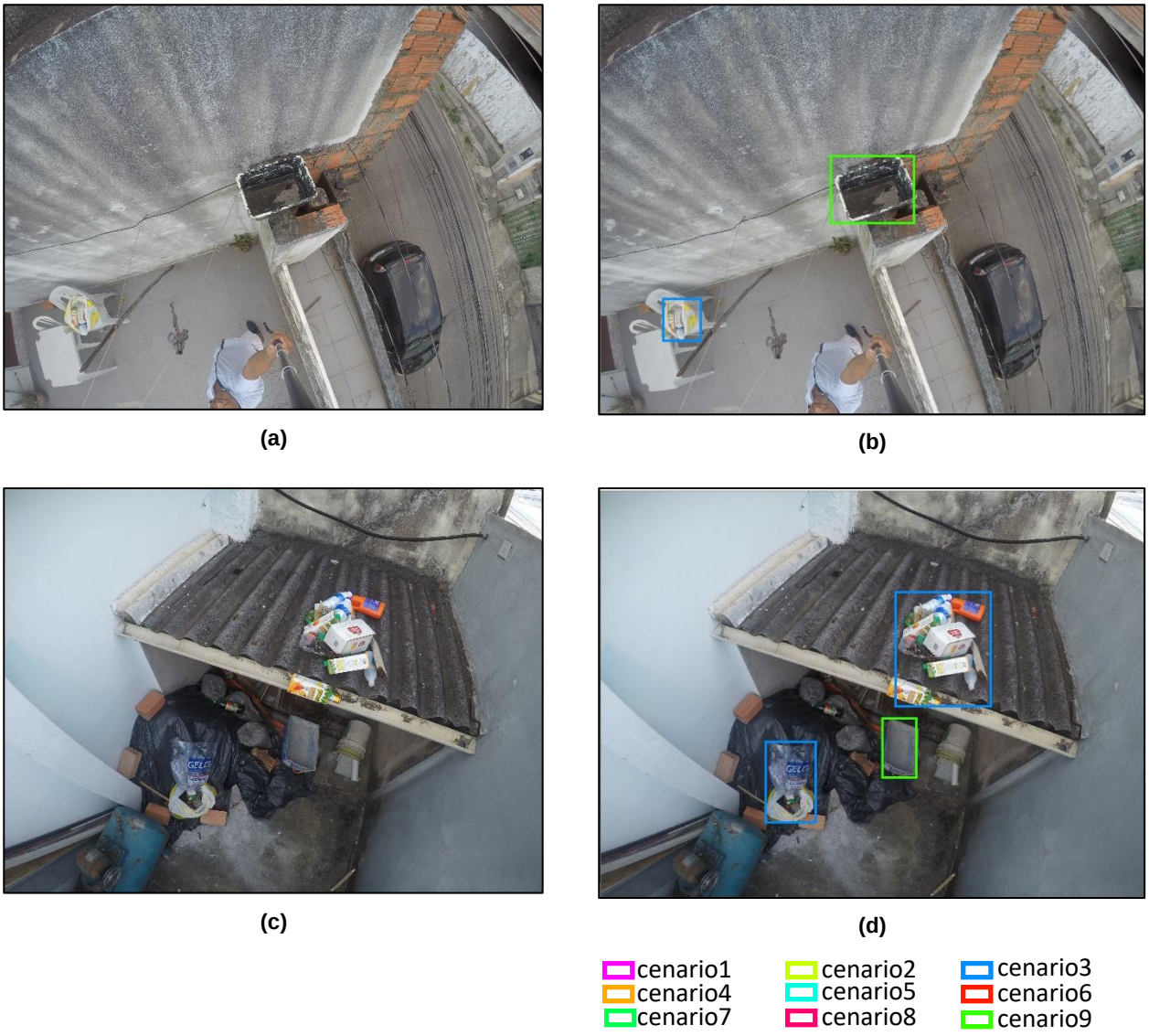
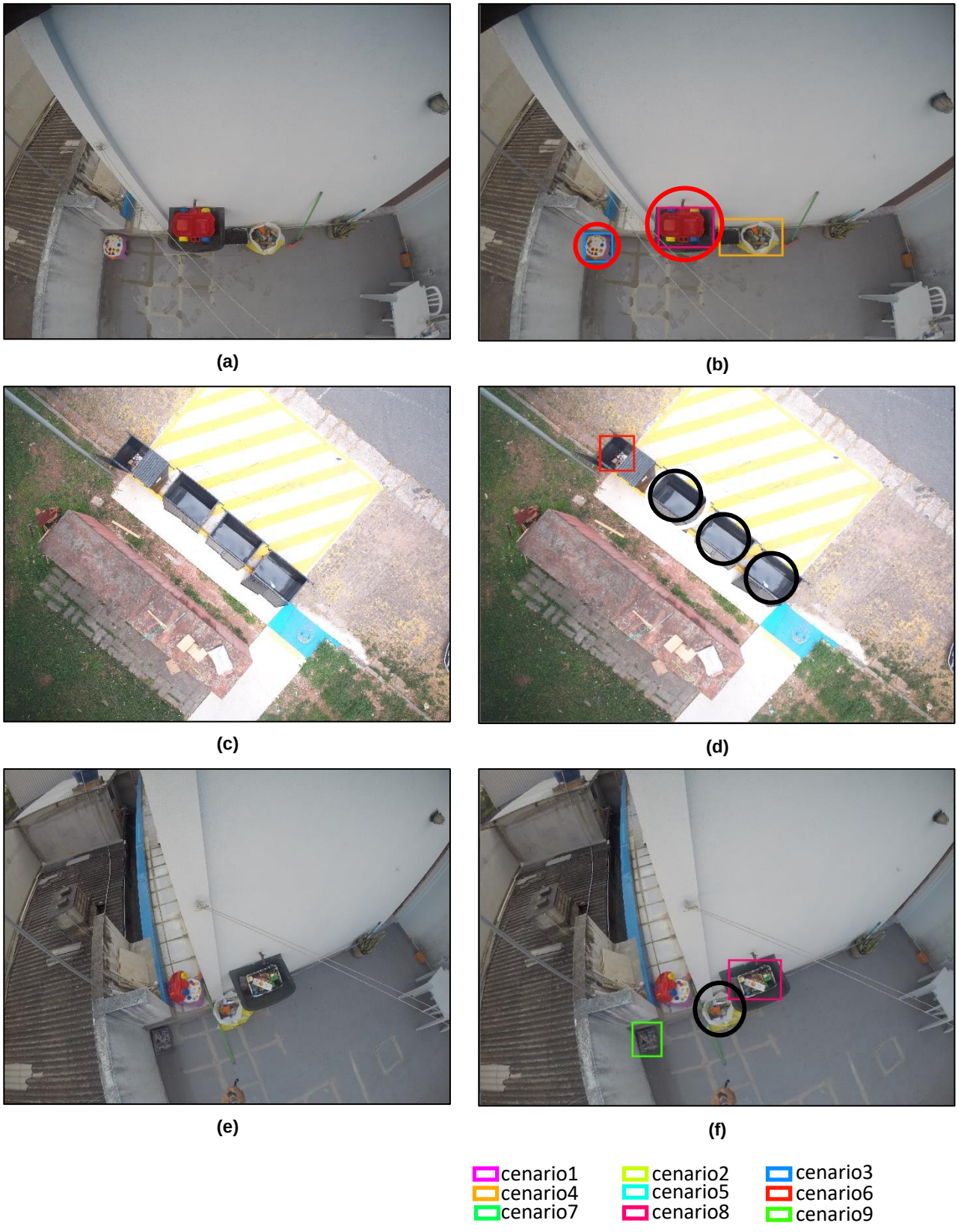


Figura 60: Resultados das detecções dos cenários pela RNC_Detec_Cenarios



Para as imagens classificadas, obteve-se o valor de 0,9028 para o mAP-50, considerando as APs (*Average Precisions*) para cada classe, as quais podem ser observadas na Tabela 5.

Tabela 5: APs calculadas para cada classe da RNC_Detec_Cenarios

Classe	**AP (Average Precision)**
cenario1	0,8182
cenario2	0,9860
cenario3	0,8098
cenario4	0,9924
cenario5	1,0000
cenario6	0,8182
cenario7	1,0000
cenario8	0,7915
cenario9	0,9091
mAP-50	**0,9028**

Os resultados demonstrados na Tabela 5 reforçam o bom desempenho RNC_Detec_Cenarios na detecção dos cenários. Seu pior desempenho (0,7915) ocorreu na detecção do cenário 8 (reservatório d'água com lixo), provavelmente pela baixa ocorrência destes tipos de cenário nas imagens de treinamento. Um exemplo disso é a ocorrência de dois casos de FP assinalados com círculos vermelhos na Figura 60b.

4.2.3. DETECÇÃO DE PEQUENAS PORÇÕES DE ÁGUA

Nos experimentos para a detecção de pequenas porções de água, foram utilizadas as imagens pertencentes ao conjunto DS2. Para validar a função objetivo do Algoritmo Genético, 3 pares de imagens (RGB e NIR) do conjunto DS2 foram correlacionadas e, além disso, foram criadas as respectivas imagens anotadas (imagens binárias) com o resultado esperado após a aplicação do índice.

Para configurar o AG, os seguintes parâmetros foram definidos: tamanho da população = 250; número de gerações = 400 (usado como critério de parada); taxa de população de substituição = 0,8; cruzamento = 0,85; taxa de mutação = 0,10.

Nos experimentos conduzidos, após a convergência do AG, obteve-se o $IIAO$ dado na Equação 21.

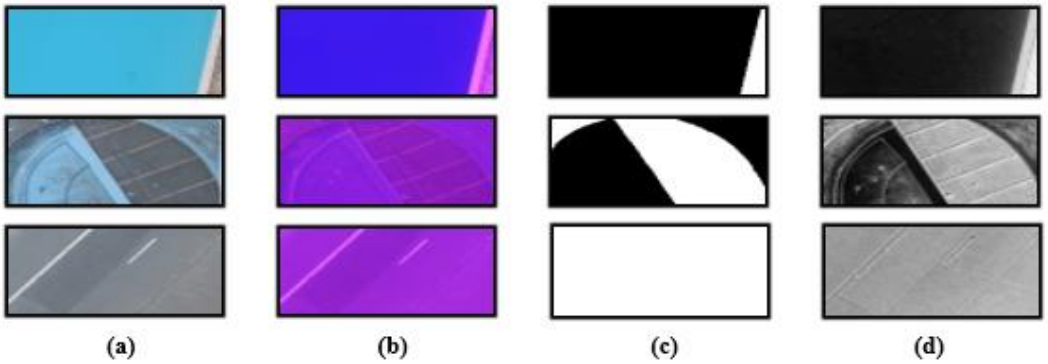
$$IIAO = \frac{6.855NIR - 2.3475B}{6.855NIR + 2.3475B} \quad (21)$$

As bandas sugeridas pelo AG para compor o $IIAO$ estão em consonância com a literatura, já que muitas pesquisas na área de sensoriamento remoto indicam uma combinação das bandas NIR e VIS para detecção de corpos d'água. Isso se deve ao fato de que a água apresenta alta absorção na faixa NIR e as bandas do espectro visível, quando combinadas com NIR, permitem caracterizar a qualidade da água. Finalmente, a importância de determinar os pesos para cada banda, feita pelo AG, deve ser destacada. Tais pesos são responsáveis pelo ajuste do índice e, como pode ser observado na literatura, são geralmente obtidos de forma empírica ou exaustiva.

É importante destacar que, embora o método proposto para fornecer o $IIAO$ considere apenas 4 bandas espectrais, ele poderia ser facilmente adaptado para um maior número de bandas ajustando o cromossomo do AG.

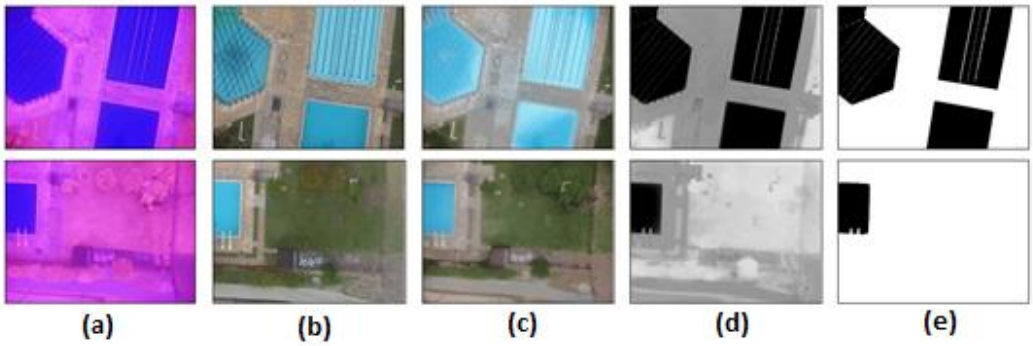
Na Figura 61 pode-se observar que as imagens geradas usando o $IIAO$ proposto (coluna d) são muito semelhantes às imagens anotadas (coluna c), evidenciando os bons resultados obtidos. Em ambas as colunas, as regiões de imagens contendo água são indicadas por níveis de cinza próximos ao preto.

Figura 61: Alguns resultados obtidos com o IIAO criado. (a) imagens RGB (I_VIS); (b) imagens NIR (I_NIR); (c) Imagens com os resultados esperados (imagens anotadas); (d) Imagens geradas pelo IIAO considerando as bandas NIR e B extraídas das imagens mostradas nas colunas a e b



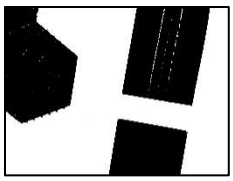
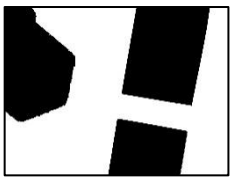
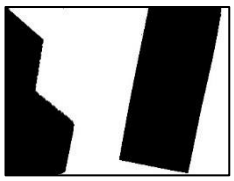
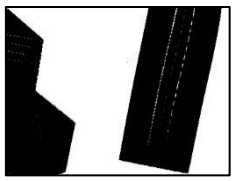
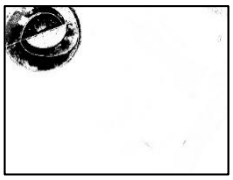
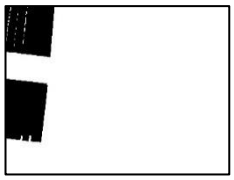
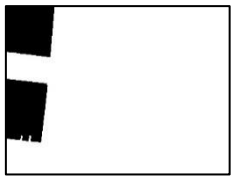
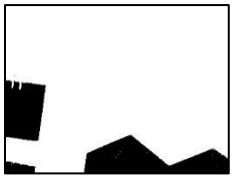
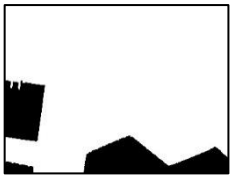
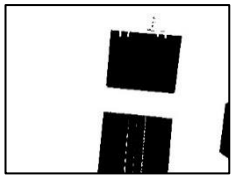
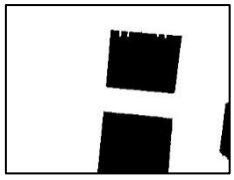
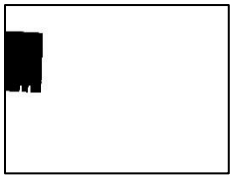
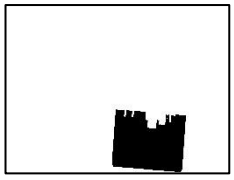
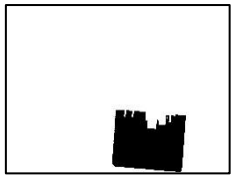
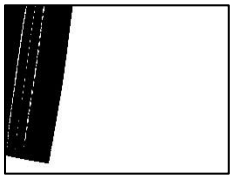
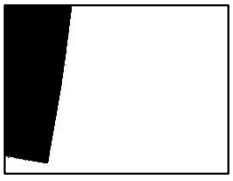
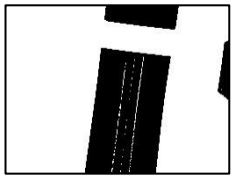
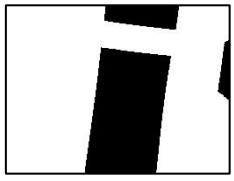
Com o propósito de realizar experimentos quantitativos e qualitativos para validar o $IIAO$, a RNA_MLP_Recons foi aplicada nas 50 imagens NIR do conjunto DS2 de imagens, para reconstituir suas correspondentes versões RGB visando a descontaminação das bandas espectrais $\tilde{G}$ e $\tilde{R}$. Duas imagens reconstituídas são mostradas na coluna c da Figura 62. Ao comparar essas imagens com as imagens RGB originais, mostradas na coluna b, pode-se observar que os resultados são muito satisfatórios. Também com base em uma análise qualitativa, é possível observar pelas imagens representadas nas colunas d e e que o método proposto para detecção de pequenas porções de água apresentou bons resultados.

Figura 62: Resultados da detecção de pequenas porções de água. (a) imagens NIR (I_NIR); (b) Imagens RGB originais (I_VIS); c) Imagens RGB reconstituídas; (d) Imagens resultantes da aplicação do IIAO (I_GRAY); (e) imagens binárias (I_BIN)



Na Tabela 6 são apresentados alguns resultados quantitativos obtidos com a aplicação do $IIAO$ em um subconjunto de 10 imagens (este subconjunto inclui imagens contendo pequenas porções de água como, por exemplo, piscinas e uma pequena fonte), para as quais foram criadas manualmente as versões anotadas com regiões contendo água destacadas em preto. A qualidade dos resultados é indicada pelas medidas MAE e $SSIM$, que são amplamente utilizadas para expressar a similaridade entre as imagens.

Tabela 6: Resultados qualitativos obtidos de experimentos considerando imagens anotadas

Imagem anotada	Imagem resultante da aplicação do IIAO	Imagem anotada	Imagem resultante da aplicação do IIAO
			
MAE: 0,0100 / SSIM: 0,9732		MAE: 0,0068 / SSIM: 0,9736	
			
MAE: 0,0350 / SSIM: 0,9345		MAE: 0,0016 / SSIM: 0,9949	
			
MAE: 0,0030 / SSIM: 0,9904		MAE: 0,0036 / SSIM: 0,9879	
			
MAE: 0,0003 / SSIM: 0,9978		MAE: 0,0013 / SSIM: 0,9946	
			
MAE: 0,0072 / SSIM: 0,9830		MAE: 0,0067 / SSIM: 0,9811	

Com o intuito de validar o $IIAO$ sugerido pela Equação 21 na detecção de porções de água ainda menores que as mostradas na Tabela 6, contidas por exemplo em caixas d'água para uso doméstico, realizou-se um experimento com a aplicação do $IIAO$ em algumas imagens contendo apenas um balde com diferentes quantidades de água, que foram adquiridas com uma câmera multiespectral (bandas R, G, B e NIR) diferente daquelas utilizadas na aquisição das imagens que compõem a base descrita na seção 3.2.1.

Nesse experimento constatou-se, pela observação da imagem resultante da aplicação do $IIAO$, que não foi possível reproduzir o mesmo resultado das imagens do conjunto DS2. Isso pode ser explicado pelo fato das imagens usadas em tais experimentos serem diferentes (em termos das larguras e das faixas das bandas espectrais) das imagens que compõem o conjunto DS2, em virtude das diferenças entre os sensores empregados nas aquisições das imagens. Isso indica que o método deve ser utilizado para computar o $IIAO$ levando em conta o conjunto de imagens em que ele será aplicado. Assim, para diferentes conjuntos de imagens pode-se ter cômputos de indicadores diferentes.

Uma outra alternativa para solução do problema seria a utilização de imagens termográficas. No entanto, no verão (estação mais crítica para a proliferação do mosquito) a água dos reservatórios d'água de polietileno plástico (que são as mais comuns) pode esquentar e isso pode comprometer a detecção da feição de água, já que a diferença térmica da água em relação a outros materiais pode não ser perceptível. Por exemplo, pode acontecer uma situação em que a água esteja tão quente quanto o próprio material do reservatório usado para armazená-la. Assim, uma solução mais efetiva para a tarefa ainda continua sendo um desafio de pesquisa.

4.2.4. GERAÇÃO DE ORTOMOSAICOS ANOTADOS E RELATÓRIOS COM POSSÍVEIS CRIADOUROS DO MOSQUITO AEDES AEGYPTI

Com base nas detecções dos objetos feitas nas imagens do conjunto DS3 separadamente e, realizando correlações entre elas e o ortomosaico, foi possível fazer as indicações conforme pode ser observado no ortomosaico anotado (ortomosaico

anotado) ilustrado na Figura 63. Além disso, foram utilizadas as coordenadas georreferenciadas centrais de cada imagem para as demarcações no ortomosaico anotado.

Figura 63: Ortomosaico_Guaianases_1 com as detecções dos objetos-alvo e as demarcações de acordo com as coordenadas georreferenciadas

O relatório com informações a respeito dos objetos-alvo, ilustrado no Quadro 9, foi gerado com base nas imagens e no Ortomosaico_Guaianases_1.

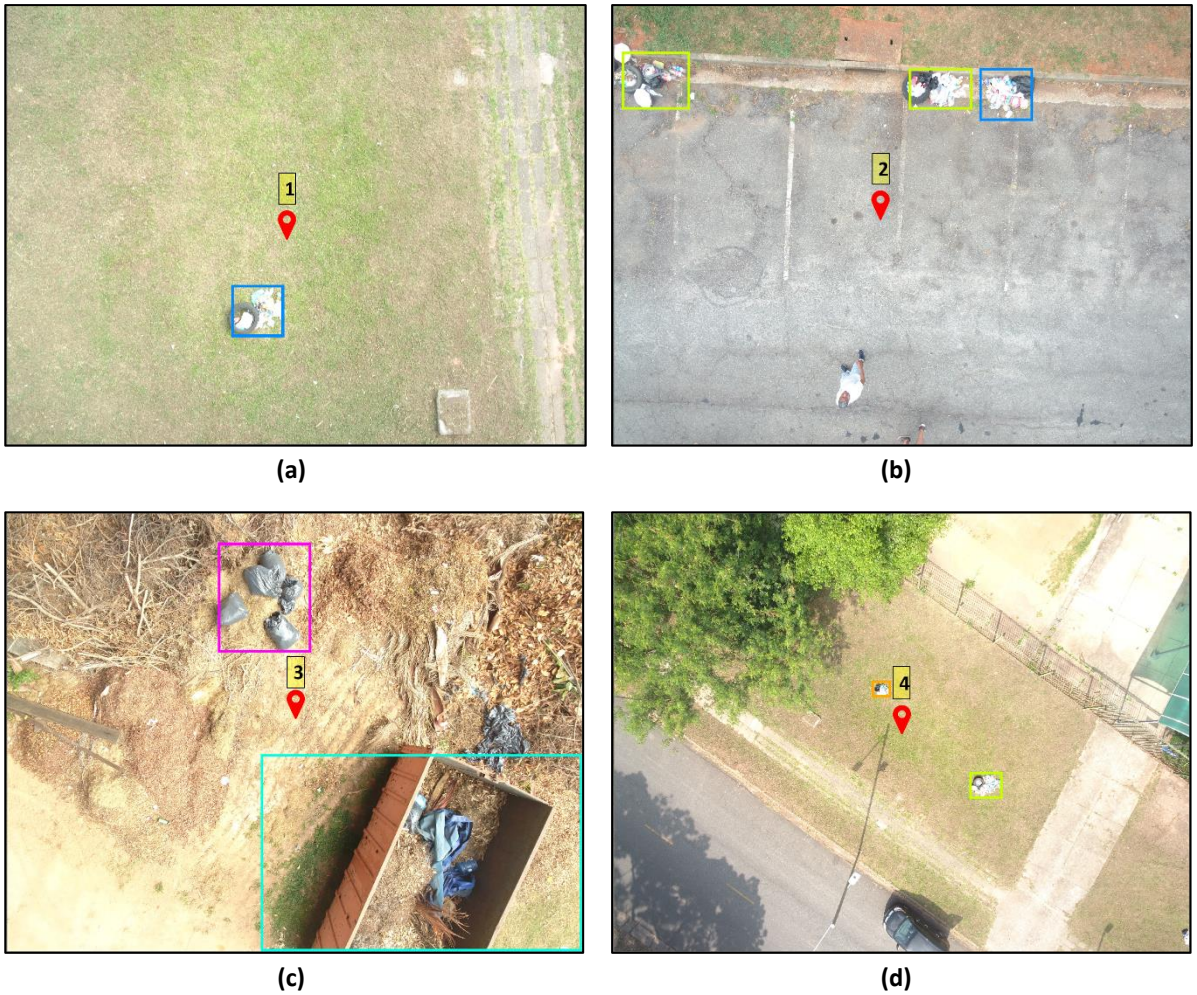
Quadro 9: Relatório dos objetos-alvo baseado no conjunto DS3

Ponto	Latitude	Longitude	Possíveis criadouros (reservatórios d'água)
1	-23,530189	-46,396979	6
2	-23,53009	-46,39698	8
3	-23,529984	-46,39698	10
4	-23,529882	-46,396979	5
5	-23,529725	-46,397132	6
6	-23,529841	-46,397138	8
7	-23,530205	-46,397124	3
8	-23,530365	-46,397117	7
9	-23,530466	-46,397117	5
10	-23,530488	-46,397275	6
11	-23,530388	-46,397276	6
12	-23,530184	-46,397276	4
13	-23,530083	-46,397277	3
14	-23,529981	-46,397277	5
15	-23,529829	-46,397276	7
16	-23,529678	-46,397277	7
17	-23,52958	-46,39728	6
18	-23,529587	-46,397428	7
19	-23,529687	-46,97429	6
20	-23,529838	-46,397427	4
21	-23,529939	-46,397426	5
22	-23,530041	-46,397429	5
23	-23,530141	-46,397426	10
24	-23,530192	-46,397427	7
25	-23,530293	-46,397428	2
26	-23.530345	-46,397429	2
27	-23,530181	-46,397578	7
28	-23,530029	-46,397577	5
29	-23,529978	-46,397577	1
30	-23,529825	-46,397578	2

Para cada imagem foi feita a contagem dos objetos-alvo detectados que, neste caso, referem-se a reservatórios d'água de uso doméstico. Decidiu-se utilizar as imagens separadamente, pois nelas a detecção dos objetos é mais precisa do que no ortomosaico. No entanto, na contagem não foram considerados objetos que, porventura, foram detectados em mais de uma imagem, ou seja, há casos onde o mesmo objeto foi detectado em duas imagens. A latitude e a longitude foram obtidas a partir do metadados de cada imagem e são baseadas na sua coordenada central.

No Quadro 10 pode-se observar o relatório que contém informações sobre cenários detectados em quatro imagens, ilustradas na Figura 64, pertencentes ao conjunto DS6. Nesse caso, os pontos foram indicados, baseando-se nas coordenadas georreferenciadas e nas coordenadas centrais das imagens.

Figura 64: Imagens classificadas pela RNC_Detec_Cenarios com detecções dos objetos-alvo e as demarcações de acordo com as coordenadas georreferenciadas



Quadro 10: Relatório dos cenários baseado no conjunto DS6

Ponto	Latitude	Longitude	Possíveis criadouros
1	-23,55859	-46,719308	Lixo a céu aberto com pneu
2	-23.560614	-46.730188	Lixo a céu aberto com pneus
3	-23.559224	-46.71937	Sacos de lixo fechados e lixo em caçamba
4	-23.558591	-46.719316	Lixo a céu aberto com pneu e reservatório d'água (balde)

Um fator importante que deve ser levado em conta é que, dependendo das condições climáticas, do relevo, o GPS do VANT pode perder o sinal e falhar ou demorar para atualizar as informações. Nesse caso, a margem de erro de posicionamento geográfico de GPS como os utilizados costuma ficar entre 3 a 5 m.

Cabe esclarecer que neste trabalho ainda não foi possível indicar com precisão as coordenadas georreferenciadas de cada possível criadouro, pois as imagens adquiridas não foram adquiridas com o uso de marcadores (pontos de referência).

Apesar de não ter sido plenamente explorada, a abordagem proposta neste trabalho serve para nortear o desenvolvimento de um sistema computacional para a finalidade a que ele se propõe.

5. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi proposta uma abordagem para a identificação automática de possíveis criadouros do mosquito *Aedes aegypti* em imagens aéreas (de edificações urbanas e cenários simulados) adquiridas por VANTS.

O emprego do *framework* YOLOv3 para a identificação dos possíveis criadouros do mosquito (objetos-alvo e cenários) mostrou ser uma ótima alternativa por apresentar poucas ocorrências de FP e, principalmente, pela precisão na detecção, resultando em o valor médio de 0,9319 para a métrica mAP-50.

Na tarefa de detecção de objetos em imagens separadas, na qual obteve-se mAP-50 = 0,9651, e em ortomosaicos, os resultados foram considerados satisfatórios. No entanto, as detecções ficaram prejudicadas nos casos de deformação dos objetos quando da junção das imagens para a composição dos ortomosaicos, tornando importante a detecção dos objetos, também, nas imagens separadas. Além disso, deve ser levado em conta também a rapidez com que as RNCs realizaram as detecções nas imagens.

No que tange à detecção de cenários suspeitos, foram investigadas duas soluções: uso da técnica BoVW combinada com o classificador SVM multiclasse (BoVW+ SVM) e a RNC tiny-YOLOv3. A primeira solução não apresentou resultados muito satisfatórios (mAP-50 = 0,6453) em virtude do elevado número de FP nas imagens classificadas. Isso ocorreu devido à grande quantidade de detalhes presentes nas imagens, o que influenciou negativamente na extração de características locais utilizadas para o treinamento do SVM. Já a tiny-YOLOv3, pelo fato trabalhar com várias escalas e com aumento de dados durante o treinamento da RNC_Detec_Cenarios, seu desempenho foi bem superior (mAP-50=0,9028).

Na etapa de detecção de pequenas porções de água, foram realizados experimentos usando uma câmera RGB com uma lente especial (equipamentos de baixo custo) para a filtragem da banda infravermelho próximo. Tendo em vista a contaminação de algumas bandas pelo uso da lente, foi desenvolvido um método para a reconstituição dessas bandas e, mesmo assim, foi possível a detecção de porções bem rasas de água com a aplicação do indicador proposto, denominado $IIAO$.

Sobre o método baseado em Algoritmos Genéticos para geração do $IIAO$, é importante enfatizar que, embora apenas 4 bandas espectrais (R, G, B e NIR)

tenham sido consideradas nos experimentos, o método pode ser aplicado em imagens com um número maior de bandas, bastando para isso apenas o redimensionamento do cromossomo. Contudo, ressalta-se que para obtenção de bons resultados, o cômputo do índice indicador deve ser baseado no conjunto de imagens em que o método será aplicado, pois as larguras e as faixas das bandas espectrais podem ser diferentes em virtude às características de cada sensor empregado na aquisição de imagens. Provavelmente, esse foi o motivo dos resultados não terem sido satisfatórios em experimentos considerando a aplicação do $IIAO$ em algumas imagens de pequenos recipientes contendo água.

No trabalho também foi demonstrada a geração de ortomosaicos anotados e relatórios com indicações de coordenadas georreferenciadas dos possíveis criadouros (objetos-alvo e cenários) do mosquito *Aedes aegypti*. Os mosaicos anotados e relatórios têm como objetivo proporcionar uma visão geral de localizações suspeitas de serem criadouros e podem ser empregadas para agilizar os processos de inspeção das áreas mapeadas pelo VANT.

A aplicação da etapa 3 (detecção de pequenas porções de água) nas imagens processadas pelas etapas 1 e 2 da abordagem proposta, permanece como um desafio de pesquisa que deverá ser contornado em trabalhos futuros. Um dos motivos é que não é nada simples solicitar a um morador de alguma residência que retire a tampa de seu reservatório d'água para que se possa fazer a aquisição de imagens de cenários reais. Também não é fácil reproduzir cenários simulados envolvendo o uso de reservatórios de água para uso doméstico em virtude dos tamanhos desses reservatórios.

Apesar das lacunas que ainda restaram, a abordagem proposta neste trabalho pode trazer contribuições significativas para a implementação de sistemas computacionais que visem auxiliar os agentes de saúde no planejamento e execução de atividades voltadas para o combate ao mosquito *Aedes aegypti* com o uso de VANTs. Neste sentido, a detecção de objetos-alvo, principalmente os reservatórios d'água para uso doméstico nos mais diversos tipos e formatos, e que são apontados como um dos principais criadouros *Aedes aegypti*, constitui uma contribuição relevante, pois não foi encontrada na literatura nenhuma abordagem para a detecção desses objetos em imagens aéreas.

Outra contribuição importante e que deve ser levada em consideração é método desenvolvido para a geração de índices indicadores de pequenas porções de água, visto que os indicadores utilizados em sensoriamento remoto não reproduzem os mesmos resultados em imagens adquiridas por VANTs.

No decorrer da pesquisa, constatou-se que uma das principais dificuldades encontradas para solução do problema investigado neste trabalho, inclusive relatada em outros trabalhos encontrados na literatura, é a aquisição de imagens contemplando as mais diversas situações de locais com possíveis criadouros do mosquito. Por conta disso, a base de imagens concebida neste trabalho também pode ser considerada como uma importante contribuição científica. Ressalta-se que parte dela (conjunto de imagens contendo os cenários simulados) será disponibilizada em repositórios públicos para que outros pesquisadores possam testar seus métodos.

Por fim, o método desenvolvido para a reconstituição de bandas espectrais também é destacado com uma contribuição deste trabalho, pois permite a utilização de equipamentos de baixo custo (câmeras economicamente mais acessíveis com o uso de lentes especiais) para a aquisição de imagens com o uso de VANTs.

Em trabalhos futuros pretende-se: i) conduzir novos experimentos envolvendo a aplicação da etapa 3 (detecção de pequenas porções de água) nas imagens processadas pelas etapas 1 e 2; ii) melhorar o método para detecção de pequenas porções de água permitindo que sejam consideradas mais de duas bandas espectrais no indicador, e que uma gama maior operações matemáticas possam ser realizadas com essas bandas. Além disso, pretende-se conduzir experimentos com esta nova forma de gerar o indicador; iii) implementar um SVC que incorpore as 4 etapas da abordagem proposta neste trabalho e que seja capaz de gerar ortomosaicos anotados e relatórios com indicações georreferenciadas mais precisas dos potenciais criadouros do mosquito, com o uso da tecnologia RTK (*Real Time Kinematic*) e, por fim, iv) conduzir novos experimentos para uma melhor investigação de BoVW+SVM na detecção de cenários, considerando também outros descritores de texturas como, por exemplo, os histogramas baseados em *attribute profiles*.

REFERÊNCIAS

AGRAWAL, A.; CHAUDHURI, U.; CHAUDHURI, S.; SEETHARAMAN, G. Detection of potential mosquito breeding sites based on community sourced geotagged images. In: SPIE 9089, Geospatial InfoFusion and Video Analytics IV and Motion Imagery for ISR and Situational Awareness II, 2014.

AGUIRRE-GÓMEZ, R; SALMERÓN-GARCÍA, O; GÓMEZ-RODRÍGUEZ, G; PERALTA-HIGUERA, A. Use of unmanned aerial vehicles and remote sensors in urban lakes studies in Mexico. In: International Journal of Remote Sensing, 2016.

ALBUQUERQUE, R. W; COSTA, M. O; FERREIRA, M. E; JORGE, L. A. C; SARRACINI, R. H; ROSA, E. O. Uso do índice MPRI na avaliação de processos de Restauração Florestal (RF) utilizando sensor RGB a bordo de VANT quadricóptero. Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto-SBSR, INPE Santos-SP, Brasil. p. 4795-4802, 2017.

ALVES, M. O; FERREIRA, R. V; CUSTÓDIO, V. B. Interpretação de imagens de drone e do sensor OLI/ Landsat 8 para identificação de pragas e doenças na cana-de-açúcar. Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto-SBSR, INPE Santos-SP, Brasil. p. 1432-1438, 2017.

AMMOUR, N.; ALHICHRI, H.; BAZI, Y.; BENJDIRA, B.; ALAJLAN, N. e ZUAIR, M. Deep Learning Approach for Car Detection in UAV Imagery. Journal Remote Sensing, v.9, n.4, p. 1-15, 2017.

APPOLINÁRIO, F. Metodologia da Ciência – filosofia e prática da pesquisa. São Paulo: Editora Pioneira Thomson Learning, 2006.

BEJIGA, M. B.; ZEGGADA, A.; NOUFFIDJ, A. e MELGANI, F. A Convolutional Neural Network Approach for Assisting Avalanche Search and Rescue Operations with UAV Imagery. Journal Remote Sensing, v. 9, n. 2, 100, 2017.

BENJDIRA, B.; KHURSHEED, T.; KOUBAA, A.; AMMAR, A.; OUNI, K. Car Detection using Un-manned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3. In: Proceed-ings of the 1st International Conference on Unmanned Vehicle Systems (UVS), Muscat, Oman, 2019.

BEZERRA, E. Introdução a Aprendizagem Profunda. In Ogasawara, V., editor, Topicos em Gerenciamento de Dados e Informações - Simpósio Brasileiro de Banco de Dados, p. 57-86. Sociedade Brasileira de Computação, 2016.

BISOGNIN, G. Utilização de Máquinas de Vetores de Suporte Para Predição de Estruturas Terciárias de Proteínas. São Leopoldo, Universidade do Vale do Rio dos Sinos, Ciências Exatas e Tecnológicas, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada. Dissertação de Mestrado, 2007.

BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. Redes neurais artificiais:

Teoria e aplicação. Rio de Janeiro: LTC, 262 p., 2000.

CANDIDO, A. K. A. A.; SILVA, N. M.; FILHO, A. C. P. Imagens de alta resolução de veículos aéreos não tripulados (VANT) no planejamento do uso e ocupação do solo. Anuário do Instituto de Geociências - UFRJ. 38, p. 147-156, 2015.

CAPOLUPO, A.; PINDOZZI, S.; OKELLO, C.; BOCCIA, L. Indirect field technology for detecting areas object of illegal spills harmful to human health: applications of drones, photogrammetry and hydrological models. Geospatial Health 8, p. S699-S707, 2014.

CASSEMIRO, G. H. M; PINTO, H. B. Composição e processamento de imagens aéreas de alta-resolução obtidas com Drone. Monografia submetida ao curso de graduação em Engenharia Eletrônica da Universidade de Brasília, como requisito parcial para obtenção do Título de Engenheiro Eletrônico. Brasília, DF, 2014.

COLET, M. E; BRAUN, A; MANSSOUR, I. H. A new approach to turbid water surface identification for autonomous navigation. 24th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2016), Plzen, Czech Republic. p. 317-326, 2016.

CONCI, A. .; AZEVEDO, E.; LETA, F. R. Computação Gráfica: teoria e prática., v. 2. Rio de Janeiro: Elsevier, 2008.

CORTES, C.; VAPNIK, V. Support vector networks. Machine Learning, 20:273-297, 1995.

COSMO, D. L. Detecção de Pedestres Utilizando Descritores de Orientação do Gradiente e Auto Similaridade de Cor. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Espírito Santo, 2014.

COSTA, L. F.; CESAR, JR, R. M. Shape analysis and classification: theory and practice. CRC Press, 659 pgs., 2000.

DALAL, N., TRIGGS, B., Histograms of Oriented Gradients for Human Detection. In: In CVPR, p. 886-893, 2005.

DEMUTH, H.; BEALE, M. Neural network toolbox for use with Matlab. User's guide, The MathWorks, Inc., 2003.

DINIZ, M. T. M.; MEDEIROS, J. B. Mapeamento de criadouros de reprodução de aedes aegypti na cidade de Caicó/RN com o auxílio de veículo aéreo não tripulado. Revista GeoNordeste, n. 2, p. 196-207, 2018.

EGMONT-PETERSEN, M.; RIDDER, D.; HANDELS, H. Image processing with neural networks - a review. Pattern Recognition, v. 35, p. 2279-2301, 2002.

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

FORNACE, K.M.; DRAKELEY, C.J.; WILLIAM, T.; F. ESPINO, J. Cox. Mapping infectious disease landscapes: Unmanned aerial vehicles and epidemiology, Trends in Parasitology, 30(11):514-519, 2014.

FU, C.; LIU, W.; RANGA, A.; TYAGI, A.; BERG, A. C. DSSD : Deconvolutional Single Shot Detector. CoRR, abs/1701.06659, 2017.

G1-DF – TV Globo (Portal G1)/Distrito Federal. Força-tarefa quer usar drones para procurar Aedes aegypti em lotes fechados no DF. Disponível em: https://g1.globo.com/df/distrito-federal/noticia/2019/01/14/forca-tarefa-quer-usar-drones-para-procurar-aedes-aegypti-em-lotes-fechados-no-df-entenda.ghtml. Acesso em 14 de fev de 2019.

GAO, B. NDWI–A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. Remote Sensing of Environment. 58, p. 257-266, 1996.

GIL, A. C. Como Elaborar Projetos de Pesquisa. 4 Edição ed. São Paulo, 2002.

GONZALEZ, R. C.; WINTZ, P. Digital image processing. Massachussetts: Addison-Wesley, 503 p., 1987.

GONZALEZ, R. C.; WOODS, R. E. Digital Image Processing. 1. ed. Massachusetts: Addison-Wesley, 2002.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep learning: The MIT Press, 800 p., 2016.

HAGAN, M. T.; MENHAJ, M. Training feedforward networks with the marquardt algorithm. IEEE Transactions on Neural Networks, v. 5, n. 6, p. 989-993, 1994.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. Textural Features for Image Classification. IEEE. Transactions on Systems, Man, and Cybernetics, v. SMC-3, n. 6, November, 1973.

HARDY, A.; MAKAME, M.; CROSS, D.; MAJAMBERE, S.; MSELLEM, M. Using low-cost drones to map malaria vector habitats. Parasit. Vectors, 10, 29, 2017.

HAYKIN, S. Redes Neurais – Princípios e práticas. 2. ed. Pearson, 2003.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, p. 770-778, 2016.

JAIN, A. K.; MAO, J.; MOHIUDDIN, K. M. Artificial neural networks: A tutorial. IEEE

Computer, v. 29, n. 3, p. 31-44, 1996.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 22, n. 1, p. 4-37, 2000.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. ACM Comput. Surv., v. 31, n. 3, p. 264-323, 1999.

JEGOU, H.; HARZALLAH, H.; SCHMID, C. A contextual dissimilarity measure for accurate and eficiente image search. In: CVPR, 2007.

JARDIM, A. Agricultura de precisão: uma nova fronteira agrícola. AgroANALYSIS, v. 37, n. 10, p. 48, 2018.

KHANDELWAL, A.; KARPATNE, A., MARLIER, ME., KIM, J.; LETTENMAIER, DP; KUMAR, V. An approach for global monitoring of surface water extent variations in reservoirs using MODIS data. Remote Sensing Environment, 202:113-128, 2017.

LECUN, Y., BOTTOU, L., BENGIO, Y. & HAFFNER, P. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278-2324, 1998.

LIN, T.-Y.; GOYAL, P.; GIRSHICK, R.; HE, K.; DOLLAR, P. Focal Loss for Dense Object Detection. 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.-Y.; BERG, A. C. SSD: Single Shot Multibox Detector. In: European conference on computer vision, p. 21-37, 2016.

LOWE, D. G., Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vision, v. 60, n. 2, p. 91-110, Nov. 2004.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, v. 5, p. 115-133, 1943.

MEHRA, M.; BAGRI, A.; JIANG, X.; ORTIZ, J. Image analysis for identifying mosquito breeding grounds. In: 2016 IEEE International Conference on Communication and Networking (SECON Workshops), p. 1-6, 2016.

MITCHELL, T. M. Machine learning. Published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997.

MS – MINISTÉRIO DA SAÚDE. Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue. Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf. Acesso em 29 de fev de 2019.

MURUGAN, P.; SIVAKUMAR, R.; PANDIYAN, R.; ANNADURAI, M. Algorithm to select optimal spectral bands for hyperspectral index of feature extraction. Indian Journal of Science and Technology 9 (37), p. 1-13, 2016.

OJALA, T., PIETIKÄINEN, M., HARWOOD, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. 29(1), 51-59, 1996.

ONUBR – Nações Unidas no Brasil. ONU usa drones para combater Aedes aegypti no Brasil. Disponível em: https://nacoesunidas.org/onu-usa-drones-para-combater-aedes-aegypti-no-brasil/. Acesso em 14 de fev de 2019.

OSP – O SÃO PAULO. Surto de Aedes aegypti pode afetar mais de 500 cidades neste verão. Disponível em: http://www.osaopaulo.org.br/tags/dengue. Acesso em 28 de fev de 2019.

PASSOS, W. L.; DIAS, T. M.; ALVES JUNIOR, H. M.; BARROS, B. D.; ARAUJO, G. M.; LIMA, A. A.; SILVA, E. A. B.; LIMA NETTO, S. Acerca da Detecção Automática de Criadouros do Mosquito Aedes aegypti. In: Anais do XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, p. 1-5, 2018.

PAVLIDIS, T. Algorithms for graphics and image processing. Computer Science Press, 416 p., 1982.

PEDRINI, H.; SCHWARTZ, W. R. Análise de Imagens Digitais - Principios, Algoritmos e Aplicações. São Paulo: Thomson, 2008.

PMS – Prefeitura Municipal de Santos. Drone volta a ser utilizado no combate ao Aedes aegypti. Disponível em: http://www.santos.sp.gov.br/?q=noticia/drone-volta-a-ser-utilizado-no-combate-ao-aedes-aegypti. Acesso em 14 de fev de 2019.

POLIDORIO, A. M; IMAI, N. N; TOMMASELLI, A. M. G. Índice indicador de corpos d'água para imagens multiespectrais. In: I Simpósio de Ciências Geodésicas e Tecnologias da Geoinformação. 9, Recife-PE, 2014.

PONTI, M.; COSTA, G, P. Como funciona o Deep Learning. Tópicos em Gerenciamento de Dados e Informações, Book Chapter, 2017.

PRASAD, M. G.; CHAKRABORTY, A.; CHALASANI, R.; CHANDRAN, S. Quadcopter-based stagnant water identification. In: Proceedings of Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (IEEE-NCVPRIPG), p. 1-4, 2015.

RAHMAN, M. A; WANG, Y. Optimizing intersection-over-union in deep neural networks for image segmentation. In ISVC, 2016.

REDMON, J; FARHADI, A. Yolov3: An incremental improvement. Computing Research Repository (CoRR), 2018.

REDMON, J.; DIVVALA, S. K.; GIRSHICK, R. B.; FARHADI, A. You only look once: Unified, real-time object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), abs/1506.02640, 2016.

REEVES, R. G. Manual of Remote Sensing. American Society of Photogrammetry. Falls Church, Virginia. 2144 p., 1975.

REINECKE, M; PRINSLOO, T; CUSTÓDIO, V. B. The influence of drone monitoring on crop health and harvest size. In: 2017 1st International Conference on Next Generation Computing Applications (NextComp). Mauritius, p. 5-10, 2017.

ROUSE, J.W.; HAAS, R.H.; SCHELL, J.A.; DEERING, D.W. Monitoring vegetation systems in the Great Plains with ERTS, In: S.C. Freden, E.P. Mercanti, and M. Becker (eds) Third Earth Resources Technology Satellite–1 Syposium. Volume I: Technical Presentations, NASA SP-351, NASA, Washington, D.C., p. 309-317, 1974.

RUMELHART, D. E.; MCCLELLAND, J. L. Parallel distributed processing, v. 1. The MIT Press, 576 p., 1986.

SAHOO, P. K.; SOLTANI, S.; WONG, A. K. C.; CHEN, Y. C. A survey of thresholding techniques. Comput. Vision Graph. Image Process., v. 41, n. 2, p. 233-260, 1988.

SILVA, R.; AIRES, K.; SANTOS, T.; ABDALLA, K.; VERAS, R. Segmentação classificação e detecção de motociclistas sem capacete. In: XI Simpósio Brasileiro de Automação Inteligente (SBAI), Fortaleza, Ceará - Brasil, 2013.

SONKA, M.; HLAVAC, V.; BOYLE, R. Image processing, analysis, and machine vision. 2 ed.. Pacific Grove: Brooks Cole, 770 p., 1999.

SZELISKI, R. Computer Vision: Algorithms and Applications. New York: Springer Science & Business Media, 2011.

TARALLO, A. S. Construção Automática de Mosaicos de Imagens Digitais Aéreas Agrícolas Utilizando Transformada SIFT e Processamento Paralelo. Dissertação (Mestrado) — Universidade de São Paulo - Escola de Engenharia de São Carlos, 2013.

TIAN, Y.; YANG, G.; WANG, Z.; WANG, H.; LI, E.; LIANG, Z. Apple detection during different growth stages in orchards using theimproved YOLO-V3 model. Computers and Eletronics in Agriculture, v. 157, p. 417-426, 2019.

TOU, J. T.; GONZALEZ, R. C. Pattern recognition principles. Massachusetts: Addison-Wesley, 377 p., 1974.

WEEKS, JR, A. R. Fundamentals of electronic image processing. SPIE/IEEE Press, 570 p., 1996.

XU, Y.; YU, G.; WANG, Y.; WU; X. E MA, Y. (2017). Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation, p. 1-10, 2017.

YANG, X.; CHEN, L. Evaluation of automated urban surface water extraction from Sentinel-2A imagery using different water indices. Journal of Applied Remote Sensing, 11, 026016, 2017.

YI, Z.; YONGLIANG, S.; JUN, Z. An improved tiny-yolov3 pedestrian detection algorithm. Optik - International Journal for Light and Electron Optics, v. 183, 2019.

YILMAZ, E.; ASLAM, J. A. Estimating average precision with incomplete and imperfect judgments. In: Proceedings of the 15th ACM international conference on Information and knowledge management, p. 102-111, 2006.

ZHOU, Y.; LUO, J.; SHEN, Z.; HU, X.; YANG, H. Multiscale water body extraction in urban environments from satellite images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens, 7, p. 4301-4312, 2014