



Universidade Federal
do Rio de Janeiro

Escola Politécnica

AUMENTO DE DADOS EM PROCESSAMENTO DIGITAL DE
IMAGENS NO AUXÍLIO À DETECÇÃO DE FOCOS DE *AEDES*
AEGYPTI

Pedro Gil Oliveira de Magalhães Couto

Projeto de Graduação apresentado ao Curso de Engenharia Eletrônica e de Computação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientadores: Eduardo A. B. da Silva
Wesley L. Passos

Rio de Janeiro
Fevereiro de 2021

AUMENTO DE DADOS EM PROCESSAMENTO DIGITAL DE
IMAGENS NO AUXÍLIO À DETECÇÃO DE FOCOS DE *Aedes*
Aegypti

Pedro Gil Oliveira de Magalhães Couto

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO
DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO DA ESCOLA PO-
LITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO
PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU
DE ENGENHEIRO ELETRÔNICO E DE COMPUTAÇÃO

Autor:



Pedro Gil O. de M. Couto

Orientador:



Prof. Eduardo A. B. da Silva, Ph.D.

Co-orientador:



Wesley L. Passos, M.Sc.

Examinador:



Prof. Sergio L. Netto, Ph.D

Examinador:



Prof. José Gabriel R. C. Gomes, Ph.D

Rio de Janeiro

Fevereiro de 2021

Declaração de Autoria e de Direitos

Eu, Pedro Gil Oliveira de Magalhães Couto, CPF 148.363.817-03, autor da monografia *Aumento de Dados em Processamento Digital de Imagens no Auxílio à Detecção de Focos de Aedes Aegypti*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na disciplina de Projeto de Graduação da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.
2. Excetua-se do item 1. eventuais transcrições de texto, figuras, tabelas, conceitos e idéias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.
3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.
4. O autor pode, excepcionalmente, encaminhar à Comissão de Projeto de Graduação, a não divulgação do material, por um prazo máximo de 01 (um) ano, improrrogável, a contar da data de defesa, desde que o pedido seja justificado, e solicitado antecipadamente, por escrito, à Congregação da Escola Politécnica.
5. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.
6. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.
7. Por ser verdade, firmo a presente declaração.



Pedro Gil Oliveira de Magalhães Couto

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es).

DEDICATÓRIA

Para meus avós, e para minha mãe. Obrigado por tudo.

AGRADECIMENTO

Os agradecimentos que seguem são quase tão longos quanto os anos que passei nesse curso. Dentre idas e vindas, altos e baixos, certeza de desistir e convicção de conseguir chegar ao fim, o processo de me tornar engenheiro foi árduo, e se imiscui indissociavelmente ao processo que fez de mim a pessoa que sou hoje. As pessoas que fizeram parte dessa trajetória, e que citarei nas próximas linhas, tornaram-se, portanto, parte de mim; e as carregarei com carinho e saudade até que essa bricolagem de sensações e experiências que posso definir como *eu*, chegue ao fim.

Antes de tudo, gostaria de agradecer a meus pais, Heloisa e Gilmar, que me acompanharam ao longo de todo o processo. Apesar de todas as dificuldades que vivemos, vocês sempre foram pedra fundamental em minha vida. O amor que vocês têm por mim - e que eu tenho por vocês - é o que me deu, dia após dia, confiança de que eu conseguiria chegar ao outro lado desse enorme desafio. A vocês, ficam toda a minha gratidão, e a certeza de que estaremos juntos sempre.

Voltando ao curso, posso dizer que existem três períodos principais pelos quais passei. Do primeiro, até 2014, guardo principalmente a companhia dos meus amigos para vida toda, Felipe De Dios e Luke Gonzalez, egressos da mesma turma do CAp-UERJ, e igualmente incomodados com as incongruências da UFRJ. Guardo, também, o apoio e a compreensão que me foram oferecidas pelo professor Carlos Teodósio - que continua a oferecer todo apoio possível ainda hoje, ao longo de diversos processos e trâmites burocráticos. Sinceramente, se não fosse por vocês, teria desistido do curso anos atrás.

O segundo foi o período mais paradoxal dos três. Foi um momento de muita dificuldade, mas também o momento de construção de alguns dos maiores vínculos afetivos que carrego: o período das Semanas de Eletrônica. Se o apoio oferecido pelo Teodósio foi o que me impediu de desistir, foram as pessoas que conheci no processo de organização das SEs, Hackathons e Copas Minerva que me deram forças para seguir em frente. O que construímos juntos virou o CT inteiro de cabeça pra baixo, e nos alçou a lugares que jamais imaginaríamos. Victor, Marina, Luiz, Gabi, Tamine, Antonio, Bobô, Cossetti, Renan: eu não consigo imaginar minha vida sem vocês; e, apesar de todas as coisas difíceis que vivemos, sinto saudades constantes

de estarmos juntos. Confabulando no GECOM, nos estressando no Auditório do Bloco A, vendendo camisetas e organizando inscrições nas mesinhas do H... Vocês são minhas melhores lembranças da faculdade, e do que eu mais vou sentir falta nos anos que virão.

Aqui, também, não posso deixar de citar algumas pessoas sem as quais nada do que construímos seria possível. Aos professores Casé e Joarez, fica o agradecimento pelo apoio incondicional aos eventos, e pela parceria de sempre. Ao Agnaldo, à Rita e ao Daniel, pelo apoio crucial para a realização do Hackathon. Ao Wallace, à Drielly e ao Rodrigo, pela ajuda com a tramitação dos processos. Aos nossos muitos voluntários, colaboradores, palestrantes e jogadores, por serem o motivo de existir de cada um de nossos eventos.

O terceiro e último ato compreende um arco de reconstrução. Foi nele que, enfim, juntei os cacos do destroço que havia se configurado e, pouco a pouco, fui construindo a pessoa que sou hoje. E não há ninguém que tenha colaborado mais com esse processo do que minha companheira, Carol. De fato, é difícil conceber esses últimos três anos sem sua presença. A tua parceria faz de mim uma pessoa muito melhor; muito mais capaz de enfrentar os desafios e dificuldades da vida e de correr atrás dos meus sonhos e da minha felicidade. Conhecer você foi como voltar a experienciar o que há de bom; e, por isso, serei eternamente grato.

Há mais duas outras pessoas sem as quais meu processo de reconstrução dentro do curso muito provavelmente não teria acontecido. Luiz e Moreno, vocês não têm ideia do quanto eu agradeço por ter contado com vocês na minha vida. Não consigo nem conceber como retribuir todas as vezes que vocês se dispuseram a me ajudar com as matérias, não importando o quão difíceis fossem as condições. Vocês foram, sem dúvidas, os professores mais importantes que tive na faculdade.

Aqui, também, cabe citar outros dois professores, que me ofereceram um voto de confiança gratuito, sem os quais tudo teria sido muito mais difícil. Luiz Wagner e Eduardo, obrigado por confiarem em mim e em minha capacidade, mesmo quando nem eu mesmo acreditei. Vocês são minhas maiores referências, e eu só posso torcer para, um dia, ser uma fração do que vocês são hoje. A vocês ficam minha gratidão e minha admiração.

Finalizando esse terceiro ato, gostaria de agradecer a meus amigos Cezar,

Mabel e Maria, e reforçar meu agradecimento ao Luiz e à Marina. Sem a ajuda e o afeto de vocês, especialmente naquela segunda metade de 2017, eu não sei como teria conseguido seguir em frente.

Gostaria de agradecer também a todas as pessoas que contribuíram com minha formação, mas que não pertencem especificamente a um dos momentos descritos acima. A meus avós, Neyde e Manoel, fica o agradecimento por todo o afeto, e pela integridade infalível; espero, um dia, conseguir ser uma pessoa que nem vocês. À minha irmã, Chris, um enorme obrigado por ter me incentivado a seguir na área, e pelo companheirismo de sempre. Aos professores Rodrigo, Pino, Moreirão, Natanael e AC, um agradecimento pela confiança e pelos exemplos dados. Ao professor Sergio, pela oportunidade oferecida, pela orientação ao longo desse período de pesquisa e pela franqueza que admiro. Ao Wesley, meu co-orientador, pela ajuda ao longo desses últimos anos, inclusive para a construção desse trabalho. Aos amigos Nascimento, Ian, Fabi, Carol Menezes e Luísa, pelo apoio e companhia providenciais em algumas das matérias mais difíceis que cursei. Ao Alan, Adriano, Capella, Igor e à Caroline Uchôa, meu agradecimento pela amizade incondicional, que pretendo levar para o resto da vida.

Por fim, gostaria de agradecer ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e à Petrobras pelo apoio oferecido para a construção desse Trabalho de Conclusão de Curso, e à banca de avaliação deste trabalho pelas contribuições oferecidas.

RESUMO

Este trabalho tem como objetivo desenvolver metodologias capazes de refinar um fluxo de aumento de dados, construído anteriormente sobre uma base de vídeos gravados por VANTs (veículos aéreos não tripulados) e utilizada para o treino de uma rede neural que almeja detectar focos de mosquito *Aedes aegypti*. Para isso, são empregados conceitos tradicionais de processamento de imagens e desenvolvidas metodologias locais e globais de variação de luminância e de casamento de histogramas de cor. Para a execução das metodologias globais, são construídas uma base de objetos reais e algoritmos de seleção responsáveis por definir, caso a caso, quais destes objetos devem ser utilizados como referência. Por fim, uma pesquisa subjetiva nos mostra que as sucessivas metodologias resultam na criação de objetos de aumento progressivamente mais verossimilhantes.

Palavras-Chave: Aumento de Dados, Visão Computacional, *Aedes Aegypti*, Dengue, Mosquito, Detecção de Objetos.

ABSTRACT

This work's objective is the development of methodologies capable of improving a data augmentation pipeline, built previously upon a database of videos recorded by UAVs (unmanned aerial vehicles) and used for training a neural network whose focus is detecting possible *Aedes aegypti* breeding grounds. In order to achieve this, we employ traditional image processing concepts and devise local and global luminance change and color histogram matching methodologies. For the global methods, we create a real object database and selection algorithms responsible for defining, for each case, which of these objects should be used as reference. At last, a subjective research shows us that the successive methodologies result in the creation of progressively credible augmentation objects.

Key-words: Data Augmentation, Computer Vision, Aedes Aegypti, Dengue, Mosquito, Object Detection.

SIGLAS

ANAC - Agência Nacional de Aviação Civil

CDF - Cumulative Distribution Function

CNN - Convolutional Neural Networks

CPU - Central Processing Unit

CRT - Cathode Ray Tube

GAN - Generative Adversarial Network

GLO - Generative Latent Optimization

GPU - Graphics Processing Unit

HSV - Hue, Saturation, Value

LBP - Local Binary Patterns

LoG - Laplaciano da Gaussiana

MBG - Mosquito Breeding Grounds

PDF - Probability Distribution Function

RGB - Red, Green, Blue

RPN - Region Proposal Network

R-CNN - Region Convolutional Neural Network

SIFT - Scale Invariant Feature Transform

SVM - Support Vector Machine

UFRJ - Universidade Federal do Rio de Janeiro

VANT - *Veículo Aéreo Não Tripulado*

YCbCr - Luma, Blue Chrominance, Red Chrominance

YOLO - You Only Look Once

Sumário

1	Introdução	1
1.1	Objetivos	2
1.2	Organização do Trabalho	3
2	Base de Dados <i>Mosquito Breeding Grounds</i>	5
3	Aumento de Dados: Primeiros Passos e Desafios	10
3.1	Aumento de Dados	11
3.2	Os Dois Caminhos	11
3.3	Metodologia Original	13
3.3.1	A Base de Pneus de Aumento	13
3.3.2	Os Métodos de Inserção	14
3.3.3	Lidando com Vídeos	15
3.3.4	A Base Aumentada	16
3.3.5	A Detecção	17
3.4	Conclusões	18
4	Adaptação de Luminância via Background	19
4.1	Luminância e suas nuances	20
4.2	Fundamentos	21
4.2.1	Luminância como base	22
4.2.2	Luma como base	22
4.3	Os Descritores	25
4.4	Transformação dos pneus de aumento	25
4.5	Resultados e problemas	28

5	Repensando o Processo: Criação da Base de Pneus Reais	31
5.1	A Base de Pneus Reais	31
5.1.1	Montagem da Base	32
5.1.2	Extração de Descritores	33
5.2	Casamento de Luminância via Base de Dados	35
5.2.1	Comparação de <i>Patches</i>	35
5.2.2	Casamento de Luminâncias	37
5.2.3	Resultados e problemas	37
5.3	Casamento de Histogramas de Cor via Base de Dados	40
5.3.1	Extração de novas <i>features</i>	40
5.3.2	Mudanças na comparação de <i>patches</i>	42
5.3.3	Casamento de Histogramas	43
5.3.4	Resultados	46
6	Resultados Subjetivos	49
6.1	Elaboração do Teste Subjetivo	50
6.2	Resultados	52
7	Conclusão e Trabalhos Futuros	54
	Bibliografia	56

Lista de Figuras

2.1	Vídeo 0002	7
2.2	Vídeo 0031	7
2.3	Vídeo 0043	7
2.4	Vídeo 0019	7
2.5	Tela do Zframer – extraído de [5]	8
2.6	Áreas Cobertas	9
3.1	Exemplo de colagem sem contexto – extraído de [20]	12
3.2	Exemplos das imagens iniciais	13
3.3	Exemplo de segmentação	13
3.4	Comparação dos métodos de inserção – Colagem, Mistura e Desfoque	15
3.5	Exemplo de consistência entre <i>quadros</i> – extraído de [5]	16
3.6	Exemplo de quadro gerado com o método de desfoque – extraído de [5]	17
4.1	Pneu real em dia claro	19
4.2	Pneu real em dia escuro	19
4.3	Imagem no espaço $YCbCr$ decomposta em seus canais	24
4.4	Variação de brilho em um objeto de aumento – extraído de [5]	25
4.5	Arquivo de parâmetros antes das alterações	26
4.6	Exemplo de extração	27
4.7	Arquivo de parâmetros após alterações	27
4.8	Quadro do vídeo 0033 antes das alterações	28
4.9	Quadro do vídeo 0033 após a adaptação de luminância	29
4.10	Comparação de objetos inseridos	29
4.11	Objeto sobre parede – método original	30
4.12	Objeto sobre parede – método atual	30

5.1	Imagem ilustrativa do modelo de mesa digitalizadora utilizado	32
5.2	Exemplos de <i>crop</i> dos objetos de interesse	33
5.3	Exemplos de máscaras para os mesmos pneus	33
5.4	Exemplos de máscaras para os fundos	34
5.5	Início do arquivo de uma das versões da base de dados de pneus reais	34
5.6	Exemplo do fluxo concebido	36
5.7	Exemplo de <i>quadro</i> do vídeo 0019 contendo pneus inseridos com o método de casamento de luminâncias e um pneu real	38
5.8	Exemplo de <i>quadro</i> do vídeo 0038 contendo pneus inseridos com o método de casamento de luminâncias e um pneu real	38
5.9	Exemplo de <i>quadro</i> do vídeo 0033 contendo pneus inseridos com o método de casamento de luminâncias	39
5.10	Extração dos vetores descritores dos histogramas	45
5.11	Visão geral do processo de transformação	46
5.12	Exemplos do método isolado	47
5.13	Comparação em vídeo escuro – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor	47
5.14	Comparação de objetos sobre gramado – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor	48
5.15	Comparação de objetos sobre concreto – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor	48
6.1	Exemplos de Pneus Reais Seleccionados	50
6.2	Exemplo de par para casamento de histogramas	51
6.3	Outro exemplo de par para casamento de histogramas	51
6.4	Exemplo de par para adequação de luminância	52
6.5	Outro exemplo de par para adequação de luminância	52

Lista de Tabelas

3.1	Resultados de detecção – extraído de [5]	18
6.1	Resultados do teste para o método de casamento de luminância . . .	53
6.2	Resultados do teste para o método de casamento de histogramas de cor	53

Capítulo 1

Introdução

O mosquito *Aedes aegypti* é responsável por transmitir algumas das doenças com maior taxa de contágio no Brasil, como dengue, febre amarela, chicungunha e zika. Até o início de 2020, que contou com a chegada da maior pandemia das últimas décadas ao país, sob a forma da SARS-COVID-19, ninguém seria capaz de disputar com ele a hegemonia dos noticiários de saúde.

Embora o contexto geral tenha mudado, o *Aedes aegypti* continua a proliferar doenças. Tomemos como exemplo uma das enfermidades citadas: a dengue. Até o dia 21 de março, segundo dados do Ministério da Saúde [1], eram 414.224 casos prováveis, um aumento de 61,5% em relação ao mesmo período de 2019. Ano este que, por sua vez, já havia contado com um aumento de casos de 264% [2] em relação a 2018.

A dengue possui altas taxas de contágio e ainda não há vacina ou cura. Sendo assim, a única política de saúde pública capaz de evitar futuras mortes é impedir que ocorra contaminação. Para isso, é necessário neutralizar o vetor de transmissão. No caso da dengue, o vetor é o mosquito; que, em sua incapacidade de obedecer a orientações e leis, precisa ser eliminado.

Atualmente, devido à altíssima taxa de fertilidade do *Aedes aegypti*, a principal ferramenta de contenção dos mosquitos se dá na tentativa de evitar sua reprodução. Para isso, é necessário localizar – e destruir – seus potenciais criadouros. Criadouros dessa espécie de mosquito consistem de água parada, sobre a qual as fêmeas depositam seus ovos. Em ambientes predominantemente urbanos, especialmente em épocas do ano em que há alto grau de precipitação, os criadouros podem

assumir inúmeras formas, tanto naturais quanto a partir de construção humana.

Na cidade do Rio de Janeiro – lugar escolhido para nosso estudo – as principais estratégias de combate são conduzidas por agentes de saúde, que percorrem os aglomerados populacionais em busca de possíveis focos de reprodução. Seguindo as orientações encontradas em [3], procuram por objetos capazes de acumular água.

Em alguns casos, utilizam-se de veículos aéreos não tripulados (VANT), com câmeras acopladas, para realizar a varredura da região onde atuarão. Essas varreduras têm como objetivo permitir a identificação prévia de objetos de interesse, permitindo que os agentes não percam tempo vistoriando propriedades sem focos em potencial.

Embora haja considerável ganho de tempo quando da aplicação desta metodologia, ainda é necessário que os agentes realizem o sobrevoo com os VANTs de forma manual, e que identifiquem visualmente os potenciais criadouros de mosquito em tempo real, enquanto pilotam o equipamento. Facilitar o processo de voo permitiria aos agentes economizar tempo; e auxiliar na detecção dos objetos poderia aumentar a acurácia – e, portanto, a efetividade – de seu trabalho, garantindo que as visitas seriam feitas apenas a propriedades com focos em potencial. Essas pequenas melhorias permitiriam aos agentes cobrir uma área efetiva muito maior do que são atualmente capazes.

A busca por essas soluções é o elemento de motivação inicial que fundamenta toda essa pesquisa.

1.1 Objetivos

O objetivo mais amplo do projeto no qual este trabalho está inserido, então, é desenvolver metodologia eficaz para colaborar com os agentes de saúde, através da automatização tanto do processo de voo dos VANTs quanto da detecção dos possíveis focos de mosquito.

O processo de voo é automatizado via aplicativo Litchi [4] e destrinchado em [5]. A detecção é realizada utilizando-se rede neural convolucional (CNN) [6], treinada com base de dados em vídeo construída com gravações extraídas de sobrevoos de drones, controlados pelo mesmo processo de automatização, e que pode ser

conferida em [7].

A detecção é restrita a objetos específicos. Em primeiro estágio, a rede é treinada para identificar pneus. A decisão se deve a evidências, que podem ser conferidas em [8], que indicam que eliminar focos apenas dos objetos com maior probabilidade de incidência de criadouros de mosquitos produz resultados similares àqueles encontrados quando do tratamento de todos os objetos possíveis.

Realizar sobrevoos com drones sobre regiões urbanas mostra-se particularmente desafiador. A necessidade de liberação do plano de voo por entidades oficiais e não oficiais – especialmente se considerando que os pontos de maior interesse para detecção de criadouros se localizam em áreas socialmente vulneráveis – faz com que cada gravação tome tempo considerável. A necessidade de deslocamento e de manejo anterior e posterior do VANT demanda pessoal. Além disso, a necessidade de marcação dos objetos de interesse presente nas gravações, em contexto de vídeo, demanda tempo e dedicação, mesmo quando da utilização de softwares de apoio.

A necessidade de obtenção de dados em uma velocidade maior do que é possível gerar se configura, assim, em um desafio que precisa ser superado para a obtenção de resultados progressivamente melhores em nossa detecção.

O objetivo específico desse trabalho, portanto, consiste em gerar, artificialmente, dados que possam ser utilizados para o treinamento da rede neural escolhida para a detecção dos possíveis criadouros de *Aedes aegypti*. Em nosso caso, através da inserção de pneus artificiais em imagens retiradas de vídeos feitos por drones, quando de seu sobrevoo por regiões urbanas da cidade do Rio de Janeiro.

1.2 Organização do Trabalho

O Capítulo 2 aborda os movimentos iniciais da pesquisa, ao redor da criação da base de dados e dos primeiros passos da detecção automática de objetos. O Capítulo 3 discorre sobre os problemas encontrados; o principal destes sendo a impossibilidade de obtenção de novos dados em quantidade suficiente – o que nos leva ao paradigma do aumento de dados. Situa então nossa escolha metodológica inicial pela utilização de métodos clássicos de processamento de imagens para a realização do aumento de dados.

Os Capítulos 4 e 5 discorrem sobre as diferentes metodologias desenvolvidas para o aumento de dados ao longo desta pesquisa, situando cada uma em suas particularidades e apresentando suas justificativas e embasamento teórico, assim como seus resultados. O Capítulo 6 faz um breve apanhado teórico sobre detecção de objetos utilizando redes neurais convolucionais e apresenta os resultados de uma pesquisa subjetiva que compara os métodos desenvolvidos.

Por fim, o Capítulo 7 encerra este trabalho com suas conclusões e possíveis desenvolvimentos futuros de pesquisa.

Capítulo 2

Base de Dados *Mosquito Breeding Grounds*

A detecção automatizada de criadouros de mosquito requer a definição de duas variáveis importantes, abordadas em sequência neste capítulo. A primeira se refere à metodologia utilizada para realizar a detecção; a segunda concerne à geração dos insumos necessários, por parte de nosso detector de escolha, para que a detecção seja possível.

A identificação de objetos em imagens é uma das principais linhas de pesquisa em Visão Computacional. Embora seja um desafio sempre presente, ganhou tração em anos mais recentes, tornando-se comum inclusive em situações do cotidiano, como detecção de faces para autenticação de identidades e vigilância [9] e aplicação de filtros em fotos [10].

O mesmo ganho de tração se apresenta para a detecção de objetos em imagens. Embora similar em propósito à identificação, a detecção de objetos requer alguns passos a mais. É necessário não apenas identificar que um objeto conste na imagem, mas identificar onde na imagem ele se localiza. Também busca-se detectar não apenas o objeto alvo em específico, mas todos os objetos que sejam do mesmo tipo que o alvo.

Usemos nosso caso como exemplo: é preciso saber exatamente onde o possível criadouro se situa na imagem aérea, pois precisar sua localização é essencial para o agente de saúde. Também é necessário detectar todas as possíveis variações de cada objeto de interesse. Existem pneus e caixas d'água de diversas tonalidades,

tamanhos e condições de preservação, e é necessário que detectemos, idealmente, todos.

Um dos principais motivos para o ganho de tração citado advém da consolidação do uso de técnicas de *Deep Learning* [11] – especificamente, Redes Neurais Convolucionais (CNN) [6] – para a identificação de objetos. A apresentação da “AlexNet” no desafio da ImageNet [12], em 2012, serve como ponto de referência a partir do qual, ano após ano, são desenvolvidas novas arquiteturas de CNN que alcançam resultados progressivamente melhores – e vastamente superiores àqueles obtidos por metodologias prévias.

Nos métodos clássicos, é comum que se realize a extração de descritores (*features*) – com uso de extratores como o *scale invariant feature transform* (SIFT) [13] – de diferentes seções da imagem. Esses descritores são, então, comparados com descritores extraídos de dados de referência, via algum classificador tradicional – como *support vector machine* (SVM) [14].

É impossível prever o tamanho que o objeto de interesse terá em uma imagem alvo. Isso faz com que seja necessário extrair seções de diversas dimensões, produzindo grande número de possibilidades. Com isso, surgem dois grandes problemas: falhas de detecção pela impossibilidade de realizar corretamente essa previsão de forma generalizada; e alto custo computacional, pela necessidade de extração de descritores para um grande número de seções.

As CNN, em comparação, são capazes de extrair descritores das imagens de forma dinâmica, superando as limitações encontradas nas metodologias clássicas. Contudo, sua consolidação recente não se dá sem justificativa. Para que operem com sucesso, CNNs necessitam de ampla capacidade computacional – consolidada recentemente, em especial por meio da utilização de placas gráficas – e de grande quantidade de dados [12].

Em termos de infraestrutura, a condição da capacidade computacional da CNN considera-se cumprida; a segunda condição justifica o desenvolvimento do presente trabalho. É pela demanda de um conjunto de dados mais completo que nos debruçamos sobre a base de dados *Mosquito Breeding Grounds* (MBG).

Sua construção vem ocorrendo de forma contínua ao longo dos últimos dois anos e meio – processo que é descrito extensivamente em [5] e [7]. A MBG é composta

por vídeos aéreos, gravados por VANTs, de regiões urbanizadas – em maior ou menor grau – contidas na Região Metropolitana do Rio de Janeiro, como pode ser observado nas Figuras 2.1, 2.2, 2.3 e 2.4. Surge da necessidade de criação de uma base de dados que cubra, de forma ampla e generalizada, objetos que possam conter criadouros de mosquito. Embora existam bases que se adequem parcialmente a essa necessidade – em especial [15] –, nenhuma delas possui quantidade suficiente de dados para que se treine nosso classificador com as premissas desejadas.



Figura 2.1: Vídeo 0002



Figura 2.2: Vídeo 0031



Figura 2.3: Vídeo 0043



Figura 2.4: Vídeo 0019

A escolha pela utilização de vídeos gravados por VANTs ocorre para que se aproxime a base de treino ao máximo da situação real encontrada pelos agentes de saúde do Estado. Além disso, uma série de diretrizes é concebida para que a base possua regularidade entre os diferentes vídeos.

Primeiro, é desenvolvido sistema automatizado de controle de voo, através do aplicativo *Litchi* [4]. Ele permite que o VANT execute sua trajetória de forma estável, com altura e velocidade fixas, e seguindo padrão de zigue-zague – formado sempre por seções de trajetória retilínea.

Então, é desligado qualquer ajuste automático da câmera. Os parâmetros são configurados à mão, para garantir padronização. Realiza-se também a calibração da

câmera do VANT e a retificação dos vídeos, através de metodologia apresentada em [16], para redução dos efeitos de distorção radial existentes.

São definidas como diretrizes, também: (i) a cobertura de múltiplos tipos diferentes de *backgrounds* para os sobrevoos; (ii) a presença de várias instâncias de cada um dos objetos de interesse; (iii) a presença destes objetos em diferentes vídeos da base; (iv) a existência de marcações confiáveis e precisas.

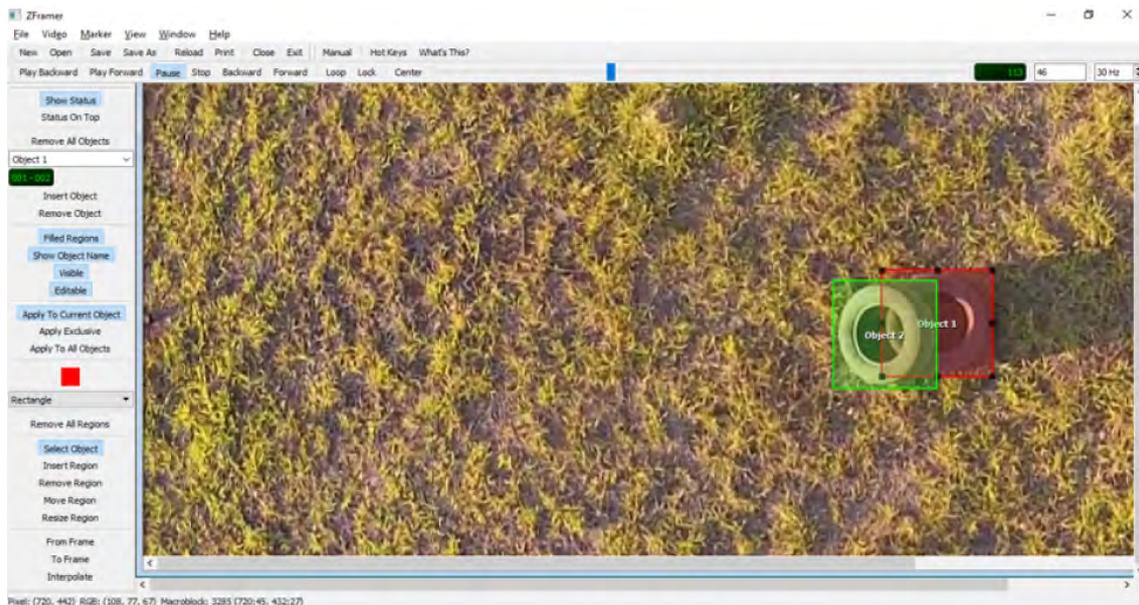


Figura 2.5: Tela do Zframer – extraído de [5]

Todos os vídeos passam pelo processo de anotação. Os objetos de interesse são marcados manualmente, a nível de quadro, através do software Zframer¹. O programa suporta a marcação de diversos objetos por quadro, e é capaz de realizar interpolação entre marcações de um mesmo objeto em diferentes quadros, eliminando a necessidade de se marcar cada objeto em todos os quadros em que aparece. O programa tem como saída um arquivo de texto contendo as anotações correspondentes à posição que cada objeto ocupa em cada *frame*.

Até o presente momento, a base de dados é composta por 18 vídeos já anotados, com alguns ainda em processo de anotação. Compreendem regiões diversas da Cidade Universitária – na Ilha do Fundão – e de Tubiacanga – na Ilha do Governador.

O processo de gravação e marcação desses 18 vídeos, porém, não aconteceu

¹<http://www.smt.ufrj.br/tvdigital/Software/zframer>



Figura 2.6: Áreas Cobertas

sem empecilhos. Os desafios que enfrentamos, assim como as soluções que concebemos para contornar as dificuldades, são fundamentais para a construção deste trabalho, e são o ponto focal do próximo capítulo.

Capítulo 3

Aumento de Dados: Primeiros Passos e Desafios

Produzir os vídeos que compõem nossa base de dados consiste em tarefa não trivial. Restrições dos mais variados tipos se impõem sobre o processo de construção, tornando-o custoso e demorado.

Relacionadas à gravação dos vídeos em si, podemos citar, entre outras: a necessidade de autorizações, concedidas pela ANAC¹, para realização de sobrevoos em regiões específicas da cidade; a dificuldade logística imposta pela necessidade de deslocamento, tanto de pessoal quanto de equipamentos, para os locais escolhidos para as gravações; a impossibilidade de sobrevoar áreas da cidade sob controle de organizações paralelas ao Estado – áreas estas que, por concentrarem possíveis criadouros de mosquito, se configuram como pontos de interesse de nosso projeto.

Relacionado à pós-produção, o processo de marcação dos vídeos, mesmo com o auxílio dos softwares já citados, consome tempo e recursos computacionais e humanos e necessita revisões e alterações constantes até que se possa considerar um vídeo como pronto.

Adicionando-se a isso as limitações orçamentárias e de pessoal, comuns a ambientes de pesquisa que necessitam desenvolver trabalho de campo não usual, encontra-se uma dificuldade importante na geração de dados para o treinamento da rede.

Com isso em mente, buscar uma solução técnica que possa contornar essas

¹<https://www.anac.gov.br/assuntos/paginas-tematicas/drones>

limitações se torna um dos imperativos de nossa pesquisa.

3.1 Aumento de Dados

Explorando-se a literatura especializada, observa-se que o aumento artificial de dados é amplamente utilizado [17] em situações onde existe escassez de dados. A termos gerais, esse tipo de processo busca produzir dados artificialmente, partindo de um conjunto de dados reais. Esses dados artificiais são, então, mesclados com os reais, gerando maior volume e variabilidade para o treinamento de algoritmos de detecção/inferência.

Para construir esses dados artificiais, é possível utilizar diferentes abordagens e metodologias. Tomando o campo de classificação de imagens e detecção de objetos como base, pode-se citar como exemplo mais simples a realização de transformações – rotações, escalamentos, inversões, deformações elásticas – nas imagens de entrada de uma base de treinamento. As imagens resultantes, distintas entre si, são então adicionadas às imagens originais.

Considerando-se que, para uma única imagem original, é possível realizar múltiplas transformações distintas, o potencial de criação de dados se mostra bastante relevante. Contudo, essa abordagem é limitada a produzir apenas pequenas variações sobre a base de dados. Para casos em que a necessidade de aumento de variabilidade é muito maior – como nesse trabalho – é preciso utilizar métodos mais complexos.

3.2 Os Dois Caminhos

A procura por referências que indicassem quais métodos utilizar nos direcionou para uma escolha entre dois caminhos: de um lado, a utilização de métodos tradicionais de processamento de imagens; de outro, a construção de Redes Neurais capazes de gerar/modificar imagens.

O uso de *Deep Learning*, de histórico mais recente e de importância crescente, tem se baseado principalmente na utilização de *Generative Adversarial Networks* (GANs) [18] para transferir características entre duas imagens de entrada, mesclando-as.

Em resumo, a ideia por trás das GANs foi construída em cima de conceitos de teoria dos jogos. Elas utilizam duas redes adversariais, que entram em disputa: uma é responsável por produzir a imagem mais convincente possível, tendo como base um conjunto de imagens reais; a outra, por tentar discernir com a maior precisão possível quais imagens que recebe são reais, e quais são geradas pela primeira rede.

O embate entre as duas gera resultados progressivamente melhores, até que, idealmente, a rede geradora seja capaz de gerar imagens tão verossímeis que a rede discriminadora não seja mais capaz de discernir entre imagens reais e artificiais. A partir daí, a rede geradora pode ser utilizada para a criação de dados artificiais.

Inúmeras abordagens para GANs seguem sendo criadas e desenvolvidas, como, por exemplo, a exposta em [19]. Nesta, uma variação do conceito de GAN, denominada GLO (*Generative Latent Optimization*), é utilizada para gerar imagens de arraias, submersas, em vídeos gravados por VANTs em voos realizados sobre regiões de oceano.

O aumento por métodos clássicos de processamento de imagens, por sua vez, insere dados artificiais por meios tradicionais: em artigos como o em [20], são criadas e segmentadas imagens variadas de objetos de interesse; essas imagens são coladas sobre *backgrounds* relacionáveis à base de dados reais. As novas imagens, resultantes desse processo, são então adicionadas à base de dados original.

Uma importante discussão trazida por esse artigo diz respeito à necessidade de verossimilhança entre dados gerados artificialmente e originais. Neste caso, a simples colagem das imagens dos objetos, sem maiores preocupações com posicionamento, escala ou contextualização se mostrou suficiente para a resolução do problema apresentado.



Figura 3.1: Exemplo de colagem sem contexto – extraído de [20]

Porém, outros artigos, como [21], demonstram que, para casos mais comple-

xos, a inserção de dados utilizando-se de correta conformação contextual é relevante para a obtenção de melhores resultados de detecção.

3.3 Metodologia Original

Toma-se a decisão de, em primeiro momento, adotar o Aumento de Dados por métodos clássicos de processamento de imagens. A partir deste ponto, desenvolve-se algoritmo extenso, descrito detalhadamente em [5], que apresentamos brevemente nesta seção.

3.3.1 A Base de Pneus de Aumento

O processo começa com a criação de uma pequena base de imagens de pneus, para utilização no aumento de dados. Para isso, são tiradas fotografias de seis modelos de pneus, em diferentes posições e combinações. As fotos são feitas com os objetos postos em cima de uma lona azul para facilitar a subsequente segmentação pela qual passam as imagens – a Figura 3.2 mostra alguns exemplos. Após a segmentação, explicitada na Figura 3.3, são obtidas as máscaras e as *bounding boxes* correspondentes a cada pneu da base. No total, a base conta com 62 imagens.



Figura 3.2: Exemplos das imagens iniciais



Figura 3.3: Exemplo de segmentação

Para aumentar a variabilidade da base, quatro operações de manipulação de imagens são incorporadas: inverter a imagem horizontalmente (com probabilidade de 50%); rotacioná-la por um ângulo arbitrário; reescalá-la por uma pequena margem e alterar seu valor médio de luminância, também por uma pequena margem.

Quando de sua posterior inserção na base de dados original, as imagens passam por processo de escalamento, para que suas dimensões se mantenham semelhantes a suas dimensões reais. Para isso, é levada em consideração a distância do VANT ao solo em cada um dos sobrevoos.

3.3.2 Os Métodos de Inserção

Partindo das referências bibliográficas já abordadas, o primeiro método desenvolvido apenas cola os pneus de aumento sobre a imagem de destino, utilizando-se de suas máscaras e seguindo as manipulações de escala e variabilidade descritas.

Um segundo método promove, após a colagem do pneu, um processo de *blending* das regiões de fronteira. Esse processo aplica um filtro gaussiano sobre a máscara do objeto de aumento, suavizando suas bordas. Após experimentações com diferentes valores, define-se o tamanho do *kernel* do filtro em $k = 3$.

O terceiro e último método promove, além da mistura das regiões de fronteira, o borramento completo de cada pneu inserido, para que seu aspecto de borramento pareça com aquele do *background* sobre o qual foi colado. Como os vídeos são gravados por VANTs em movimento, é possível perceber, nos quadros, desfoque de movimento (*motion blur*). Esse desfoque contrasta com as fotografias estáticas e controladas a partir das quais os pneus de aumento foram extraídos, gerando incongruências.

Para tanto, o terceiro método define um filtro Laplaciano da Gaussiana (LoG) bidimensional como forma de mensurar o nível de borramento de uma imagem. Esse filtro é construído no espaço contínuo e é posteriormente discretizado, sendo transformado em um *kernel* aplicável sobre as imagens. Sua construção leva em consideração a altura a partir da qual cada um dos vídeos foi gravado, visando respeitar as variações de escala existentes na MBG.

O algoritmo de inserção é iterativo: caso determine-se, através do LoG, que o nível de borramento do objeto de aumento é inferior ao do fundo sobre o qual foi inserido, um filtro gaussiano bidimensional é aplicado sobre o pneu. Após a aplicação, comparam-se novamente os níveis de borramento: se o nível de borramento do objeto de aumento ainda estiver menor, amplia-se a largura do filtro gaussiano, e este é novamente aplicado. Repete-se esse processo até que o objeto de aumento apresente

borramento similar ao do fundo.



Figura 3.4: Comparação dos métodos de inserção – Colagem, Mistura e Desfoque

A Figura 3.4 mostra um comparativo visual entre os três métodos apresentados. Nela, é fácil observar que o segundo método é capaz de suavizar com sucesso os artefatos de fronteira presentes no primeiro método, e que o terceiro método aproxima o nível de borramento do objeto daquele do fundo sobre o qual foi inserido.

3.3.3 Lidando com Vídeos

Todas as metodologias descritas acima cobrem a inserção de objetos em imagens estáticas. Para que o processo funcione em vídeos, é necessário que as imagens sejam inseridas nos quadros vizinhos levando em conta alguma forma de consistência temporal. Para isso, decide-se por utilizar correlação de fase [22] para determinar os deslocamentos relativos entre cada um dos quadros extraídos de cada vídeo.

A metodologia de correlação de fases utilizada calcula a função de correlação cruzada entre o quadro atual e o subsequente. Para isso, modela o quadro subsequente como um vetor de deslocamento aplicado sobre o quadro original

$$f_1(x, y) = f_0(x - t_x, y - t_y)$$

e calcula a transformada de Fourier discreta de ambos, considerando m e n como as frequências espaciais:

$$F_1(m, n) = F_0(m, n) \cdot e^{-2\pi i(mt_x + nt_y)}$$

Então, calcula a correlação cruzada normalizada entre os quadros (com * indicando o complexo conjugado)

$$C(m, n) = \frac{F_0 \cdot F_1^*}{|F_0 \cdot F_1^*|} = \frac{F_0 \cdot F_0^*}{|F_0 \cdot F_0^*|} \cdot e^{-2\pi i(mt_x + nt_y)},$$

toma a transformada inversa e mostra que a correlação cruzada normalizada equivale à função delta de Dirac,

$$c(x, y) = \delta(x - t_x, y - t_y),$$

que indica o deslocamento translacional entre os quadros em questão.

Como os vídeos são gravados em alturas constantes e não há rotação da câmera, podemos modelar o deslocamento entre seus quadros como translações. Assim, faz sentido assumir que o pico da função de correlação cruzada descrita acima representa uma boa aproximação para esse deslocamento.

Munido das informações de correlação de fase, o método é capaz de inserir os objetos em *quadros* sequenciais levando em conta o deslocamento relativo da câmera em relação ao *background* e, assim, garante a sensação de consistência temporal. Um exemplo de resultado pode ser observado na Figura 3.5.



Figura 3.5: Exemplo de consistência entre *quadros* – extraído de [5]

3.3.4 A Base Aumentada

Com o *pipeline* descrito acima construído, as três metodologias de inserção são utilizadas para inserir objetos na base de dados original, gerando três versões

da Base de Dados MBG Aumentada. Essas versões são alimentadas a um detector, para auferir sua colaboração para o processo de detecção. A Figura 3.6 retrata um dos quadros gerados.



Figura 3.6: Exemplo de quadro gerado com o método de desfoque – extraído de [5]

3.3.5 A Detecção

O processo de detecção é efetuado através de uma CNN do tipo YOLO [23]. Ela é alimentada quadro a quadro, e não leva em conta correlações temporais ou espaciais entre diferentes quadros. São estabelecidos quatro grupos de comparação: dois sem aumento de dados (Caso 1 e Caso 2), um para as três metodologias de inserção desenvolvidas (Caso 3) e um com o aumento de dados provido tanto pelo terceiro método quanto pela própria rede de detecção (Caso 4) – nesse caso, a YOLO aplica, automaticamente, pequenas variações nas imagens da base de treino.

Ao final, como pode ser observado na tabela abaixo, conclui-se que o terceiro método reduz significativamente os falsos positivos, e mantém resultados similares – ou um pouco melhores – de detecção em relação ao grupo sem aumento de dados. Aqui, as colunas TP_b representam as detecções corretas e as colunas FP_b representam os falsos positivos.

Tabela 3.1: Resultados de detecção – extraído de [5]

	CCMN		FAU		BLI3		Total	
	TP _b	FP _b						
Case 1	0	27	8	27	7	71	15	125
Case 2	0	11	5	13	3	60	8	84
Case 3 (Paste)	0	0	3	4	4	6	7	10
Case 3 (Blend)	0	0	2	2	5	3	7	5
Case 3 (Blur+Blend)	0	1	10	2	7	7	17	10
Case 4	0	0	7	3	5	4	12	8

3.4 Conclusões

A metodologia desenvolvida sana muitas dificuldades relacionadas à realização de aumento de dados, via métodos tradicionais de processamento de imagens, em uma base de dados de vídeos. Contudo, os resultados obtidos mostram que seu principal objetivo – suprir a ausência de dados para permitir melhora na capacidade de detecção – é apenas parcialmente alcançado. A diminuição significativa de falsos positivos aparenta ser um bom indicativo de que o aumento no número de objetos na base ajudou o detector a entender o que não se configura como objeto de interesse; contudo, os níveis de detecção continuam abaixo do que se almeja.

Este trabalho, então, parte desta primeira fase de experimentações, e da base técnica e de conhecimento construída até aqui. Em cima dela, tenta construir novas metodologias, capazes de obter resultados de detecção que permitam seu uso de forma ampla e consistente em situações reais.

Essas novas tentativas, que compõem o coração deste texto, são descritas nos próximos capítulos.

Capítulo 4

Adaptação de Luminância via Background

Para realizar novas tentativas, é preciso definir primeiro qual problema atacar. Os resultados obtidos pelos métodos já descritos parecem apontar que uma representação mais fidedigna dos objetos inseridos resulta em melhores resultados de detecção. As questões que se põem são: como fazer com que os pneus inseridos pareçam mais com os reais? O que, nos resultados obtidos, distoa de forma mais imediata do que esperamos observar?

Dentre as possibilidades, aquela que mais se destaca é a discrepância de luminância. Nossa base de vídeos possui conteúdos gerados em dias com diferentes condições climáticas. A iluminação observada em um vídeo como o 0010 (Figura 4.1), gravado em um dia ensolarado, difere drasticamente daquela presente no vídeo 0033 (Figura 4.2), capturado sob a penumbra de um dia nublado.



Figura 4.1: Pneu real em dia claro



Figura 4.2: Pneu real em dia escuro

Contudo, os objetos de aumento sempre aparecem com as mesmas carac-

terísticas, salvo a pequena variação de luminância empregada para aumentar sua variabilidade. Então, enquanto os resultados obtidos em vídeos mais claros parecem fazer sentido, o problema se evidencia quando de sua aplicação nos vídeos mais escuros.

Com o caminho definido, busca-se encontrar uma forma de adaptar a luminância dos objetos de aumento àquela observada no fundo sobre o qual eles são inseridos. A primeira destas formas realiza esse processo de forma local, transformando os pixels inseridos de acordo com a região imediatamente a seu redor. Para isso, define brilho e contraste como descritores básicos para modelar a luminância percebida, e os utiliza para comparar os objetos e o fundo sobre os quais são inseridos.

4.1 Luminância e suas nuances

A construção de imagens artificiais depende da capacidade de reprodução dos padrões de percepção visual humanos. Conceitos como brilho, luminância, *luma*, *gama* e luminosidade foram criados para tentar mapear as complexas relações decorrentes da tentativa de modelar nossa visão, e de criar sistemas capazes de gerar artefatos que ela seja capaz de compreender.

A medida de luminância – ou seja, da intensidade luminosa em si – segue comportamento linear, e é baseada nas relações físicas que regem a luz visível. Contudo, a percepção humana de luminosidade apresenta comportamento não linear, demonstrando maior sensibilidade para tons mais escuros, e menor sensibilidade para tons mais claros. Logo, mesmo que consideremos o cenário mais simples possível, de uma imagem monocromática, já se apresentam dificuldades para que consigamos codificar informações visuais.

Quando falamos de imagens coloridas, a dificuldade se torna ainda maior. Nas imagens monocromáticas, a relação entre a informação codificada e a luminosidade é direta: mesmo com as dificuldades postas acima, uma imagem monocromática captura, por essência, a intensidade da luz em cada ponto que representa. Imagens coloridas, porém, codificam a intensidade da luz de forma indireta, uma vez que as relações entre as componentes de cor e a percepção aparente de luminosidade não

são imediatas.

Isso porque o sistema visual humano não determina a intensidade luminosa que percebe de forma similar entre as diferentes cores. Se decomposermos a luz para suas três componentes básicas (vermelho, verde e azul) percebemos que a luz verde apresenta influência muito maior para a percepção de claridade do que as outras duas – com a luz azul sendo a que, de longe, menos influencia nossa visão.

Portanto, quando criamos imagens coloridas, é necessário respeitar estas relações; e, quando desejamos retirar informações de imagens coloridas já construídas, precisamos estar conscientes das nuances que as acompanham.

4.2 Fundamentos

As imagens com que trabalhamos são inicialmente codificadas seguindo um padrão global para imagens digitais: o *RGB*. Este padrão utiliza três componentes distintas, uma para cada componente básica da luz – R para vermelho, G para verde e B para azul – e a informação é transformada por uma função não-linear. Essa transformação, comumente chamada de correção de gama, tem como intuito seguir a relação não-linear da visão humana, descrita previamente, alocando maior quantidade de informação para os tons mais escuros, e menor para os tons mais claros (nos quais a quantidade extra de informação codificada não seria percebida de qualquer forma).

Partindo de imagens codificadas neste padrão, é possível atuar por dois caminhos para extrair as informações que desejamos mapear em nossos descritores: o primeiro envolve desfazer a transformação não-linear e trabalhar em cima de valores absolutos de luminância; o segundo, converter a informação para outro padrão, que represente a luminosidade de forma independente das componentes de cor, mantendo a não-linearidade que aproxima nossos dados à percepção humana de fato. Para nos referirmos à informação de luminosidade tratada de forma linear, adotamos o termo luminância; para aquela modelada em conjunto com a não-linearidade típica da correção de gama, adotamos luma.

4.2.1 Luminância como base

Nesse caso, começamos com a normalização dos valores de cada componente de cor para cada pixel da imagem para um intervalo entre 0 e 1, dividindo-se os valores por 255. Então, realiza-se a transformação inversa da não-linearidade originalmente aplicada, seguindo a forma definida em [24]:

$$V_{linear} = \begin{cases} \frac{V'}{12.92} & V' \leq 0,04045 \\ \left(\frac{V'+0,055}{1,055}\right)^{2,4} & V' > 0,04045 \end{cases}$$

onde V' representa o valor de intensidade de um pixel para um dado canal em uma imagem, ainda sob efeito da não-linearidade, e V representa o valor do pixel após o processo de linearização.

Finalmente, para encontrar a luminância Y de cada pixel, os canais são ponderados de acordo com o peso com que cada componente contribui:

$$Y = R_{linear} \times 0,2126 + G_{linear} \times 0,7152 + B_{linear} \times 0,0722$$

onde R_{linear} , G_{linear} e B_{linear} representam as intensidades de um mesmo pixel, já linearizadas pelo processo descrito acima, em cada um dos três canais da imagem (R , G e B).

Para realizar a transformação dos objetos de aumento, este processo precisa ser aplicado tanto nesse sentido (para calcular os descritores da imagem original do pneu de aumento) quanto no sentido contrário (para, após a transformação do objeto em Y , reconstruirmos seus valores em RGB).

4.2.2 Luma como base

Nesse caso, utiliza-se da mesma premissa adotada em [5] para realizar a variação aleatória de luminância: efetuar transformação no espaço de cores de RGB para $YCbCr$, e tomar a informação de luminosidade não-linearizada (canal Y) como base para as futuras transformações. Os cálculos referentes à extração dos descritores se dão diretamente sobre este canal, e não necessitam de outras operações – como as descritas na seção anterior. A transformação entre os espaços de cores é abordada nas subseções abaixo.

4.2.2.1 Espaços de cores

Espaços de cores são bases vetoriais sobre as quais construímos sistemas de coordenadas para mapear cores. Tratando-se de imagens digitais, são discretizados de acordo com a resolução de *bits* disponível para sua representação.

Cada espaço de cores resulta de uma transformação de base a partir de outro espaço de cores, de forma que é sempre possível realizar transformações, em ambas as direções, entre diferentes espaços – assumindo-se que a base da qual se parte é completa e bem definida, e que não há inserção de não-linearidades, como correção de gama, no sistema.

A existência de diferentes espaços de cores surge de diferentes necessidades de representação. Aplicações tecnológicas possuem enfoques e pontos de atenção distintos, e realizar transformações deste tipo pode resultar em ganhos expressivos. Seja no caso do próprio $YCbCr$, construído originalmente para diminuir o gasto de banda na transmissão de sinais de televisão, seja no caso do HSV , tão utilizado para criação de paletas de cores em softwares de *design*, a construção e utilização de diferentes espaços de cores tem sido basilar nas áreas de síntese e processamento de imagens.

4.2.2.2 O espaço de cores $YCbCr$

O espaço de cores $YCbCr$ foi originalmente concebido levando-se em conta sua utilização em televisores do tipo CRT (tubos de raios catódicos), e tentava combater a grande redundância de informações existente em representações do tipo RGB . Levando em conta que a intensidade luminosa absoluta é mais relevante para a percepção visual humana do que as componentes de cores, o $YCbCr$ transforma o espaço de representação construindo um canal responsável exclusivamente por mapear a intensidade percebida da luz (luma, representada no canal Y); e deixa aos outros dois canais (diferença-azul – Cb – e diferença-vermelho – Cr) a responsabilidade de representar as cores (crominância). Uma decomposição visual ilustrativa nesse espaço de cores pode ser observada na imagem 4.3.

Desta forma, era possível propositalmente aumentar a quantidade de informação (resolução) disponível para o canal Y , em detrimento dos canais C_b e C_r , sem que fossem perceptíveis falhas nas imagens, aumentando a eficiência na

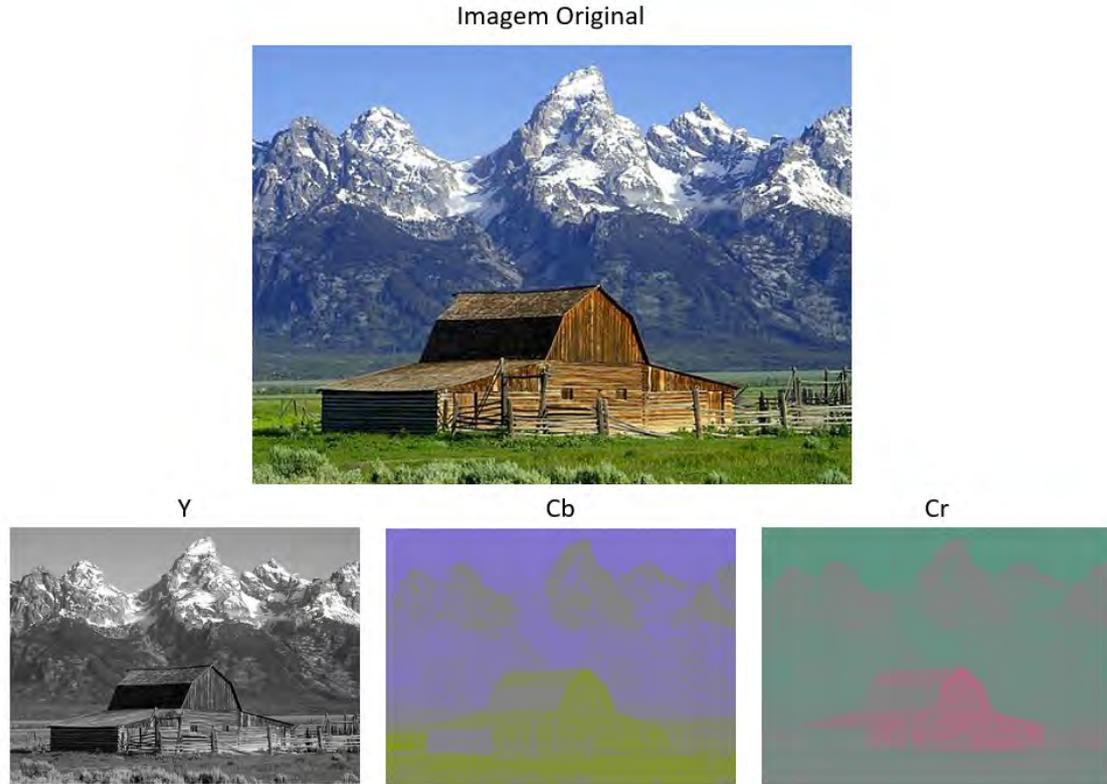


Figura 4.3: Imagem no espaço $YCbCr$ decomposta em seus canais

transmissão de informação imagética.

Partindo-se de uma base RGB não-linear, como a RGB , a transformação de base para o espaço $YCbCr$ pode ser descrita pela matriz de transformação

$$\begin{bmatrix} Y' \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} K_r & K_g & K_b \\ -\frac{1}{2} \frac{K_r}{1-K_b} & -\frac{1}{2} \frac{K_g}{1-K_b} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \frac{K_g}{1-K_r} & -\frac{1}{2} \frac{K_b}{1-K_r} \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

onde os coeficientes K_r , K_g e K_b , em nosso caso, são referentes ao espaço RGB e, assim como no primeiro caminho, valem, respectivamente, 0.2126, 0.7152 e 0.0722.

É possível observar que a transformação que resulta no canal Y' segue o mesmo procedimento daquela realizada no primeiro caminho, salvo a normalização e a remoção da não-linearidade, que não acontecem aqui.

4.3 Os Descritores

Brilho e contraste são descritores comumente utilizados para representar luminosidade, e possuem construção matemática simples: o brilho representa o nível médio e o contraste, o desvio padrão da luminosidade de dada imagem.

A opção por suas utilizações tem base no significado perceptivo de cada uma dessas grandezas. O brilho representa o quão “clara” ou “escura” é uma imagem, como exemplificado na Figura 4.4. O contraste, quando levada em conta apenas a luminosidade, caracteriza as relações que o material representado na imagem possui com a luz – como reflexividade, opacidade, etc.



Figura 4.4: Variação de brilho em um objeto de aumento – extraído de [5]

Partindo-se dessas definições, pode-se inferir que a discrepância observada nos resultados originais, nos vídeos mais escuros, entre os pneus de aumento e os fundos sobre os quais foram inseridos se configura, principalmente, como uma discrepância de brilho. O ajuste mais simples que pode amenizar a sensação de inverossimilhança percebida é, portanto, ajustar esta grandeza.

O ganho de desempenho de detecção observado quando da adição do passo de desfoque no *pipeline* original parece indicar, também, que características de textura são importantes para a obtenção de bons resultados. Vem daí a opção por levar em conta o contraste.

4.4 Transformação dos pneus de aumento

Neste método, utilizamos como base o *pipeline* descrito no capítulo anterior. Para cada pneu de aumento a ser inserido, são escolhidos um quadro aleatório do vídeo alvo (que serve como referência) e uma coordenada aleatória dentro deste quadro. A partir dessas informações, é aplicada correlação de fase para definir a posição do objeto nos *quadros* prévios e posteriores. Quando, em qualquer uma das

direções, é inferido que as coordenadas do objeto no próximo *quadro* superam as dimensões máximas do quadro, o processo é interrompido.

Os dados posicionais são mapeados em um arquivo de parâmetros, que contém as informações necessárias para que o método realize a inserção de fato de cada objeto: *quadro* onde será inserido, coordenadas (x e y), tipo de objeto, caminho para a imagem do pneu de aumento e parâmetros de variabilidade aleatórios (rotação, mudança de escala e mudança de luminância aleatórias). Uma seção de um desses arquivos pode ser observado na Figura 4.5.

frame	x	y	objectid	objPath	rotAngle	flipHor	lumShift	scaleFactor
3240	613	944	0	tire-0042.png	85	0	70.75732129880234	0.29356695155120033
3241	612	957	0	tire-0042.png	85	0	70.75732129880234	0.29356695155120033
3242	612	970	0	tire-0042.png	85	0	70.75732129880234	0.29356695155120033
3243	611	983	0	tire-0042.png	85	0	70.75732129880234	0.29356695155120033
3244	610	995	0	tire-0042.png	85	0	70.75732129880234	0.29356695155120033

Figura 4.5: Arquivo de parâmetros antes das alterações

A partir das coordenadas definidas no arquivo, é estabelecida a *bounding box* completa do objeto, de acordo com as regras de escala derivadas da altura de voo do drone, e sua imagem base é colada. O objeto passa pelos processos de mistura nas bordas e desfoque e isso conclui o processo original.

Contudo, a nova metodologia muda o *pipeline* descrito em dois pontos. Primeiro, ao adicionar, ainda na criação do arquivo de parâmetros, um estágio de extração de descritores da região onde o objeto será inserido. Depois, inserindo um novo passo após a conclusão do processo original, para realizar a transformação da luminância dos objetos inseridos.

O primeiro adendo começa com a avaliação do *background* sobre o qual cada objeto será inserido. Após a definição do quadro de referência e das coordenadas iniciais, infere-se a *bounding box* correspondente. A área inscrita na *bounding box* é destacada do resto do quadro. Sobre ela, é realizada conversão do espaço de cores para o espaço *YCbCr*, seguindo a decisão de projeto apresentada em [5], e abordada previamente. A biblioteca *Numpy*, construída sobre linguagem *Python*, é utilizada para calcular vetorialmente a média (brilho) e o desvio padrão (contraste) da componente luma da imagem (canal Y). Esses valores são salvos no arquivo de

parâmetros, e são vinculados ao objeto correspondente. O fluxo é exemplificado na Figura 4.6, e a Figura 4.7 retrata a nova configuração do arquivo de parâmetros.



Figura 4.6: Exemplo de extração

frame	x	y	objectid	objPath	rotAngle	flipHor	lumShift	scaleFactor	brilliance	contrast
4680	2004	117	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4681	2004	144	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4682	2003	171	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4683	2003	197	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4684	2002	224	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4685	2002	251	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235
4686	2001	277	0	tire-0037.png	85	1	81.92647483869152	0.02123268449584231	70.6732796032238	29.067346159235

Figura 4.7: Arquivo de parâmetros após alterações

O segundo adendo resgata os valores salvos no arquivo de parâmetros para, após as operações de mistura e desfoque, realizar a transformação de fato da luminância dos objetos inseridos.

A transformação ocorre em todos os pixels correspondentes ao objeto. Para isso, é aplicada operação de normalização seguida de desnormalização. O processo é concebido também vetorialmente, e segue a forma

$$P_{out} = \left(\frac{P_{in} - \bar{P}}{\sigma_P} \right) \times \sigma_T + \bar{T}$$

onde o pixel pré-transformação é representado por P_{in} e o pós-transformação por

P_{out} ; \bar{P} e σ_P representam o valor médio e o desvio padrão originais do pneu de aumento e \bar{T} e σ_T os valores de brilho e contraste referentes aos descritores extraídos do *background*, e que consideramos como alvo de nossa transformação.

4.5 Resultados e problemas

Os resultados obtidos mostram que o objetivo inicial de diminuir a discrepância percebida nos vídeos gravados em dias mais escuros foi alcançado. Com a aplicação da nova metodologia, observa-se que os objetos de aumento mudam de aspecto de acordo com as condições do fundo sobre o qual foram colados. A imagem 4.8 mostra o resultado do método original sobre um dos quadros iniciais do vídeo 0033 – um dos vídeos gravados em um dia nublado. O mesmo quadro, quando processado pelo novo *pipeline*, pode ser observado na Figura 4.9.



Figura 4.8: Quadro do vídeo 0033 antes das alterações

É fácil observar, neste exemplo, que a vasta maioria dos objetos de aumento inseridos pelo novo método apresenta maior verossimilhança com as condições de iluminação presentes. A Figura 4.10 apresenta, de forma mais detalhada, a transformação sofrida por alguns dos objetos.

Curiosamente, a adaptabilidade que o método busca construir é exatamente o que nos faz questionar sua eficácia. Ao atrelar o aspecto final de cada objeto de



Figura 4.9: Quadro do vídeo 0033 após a adaptação de luminância



Figura 4.10: Comparação de objetos inseridos

aumento a sua própria seção de *background*, o método gera mais variabilidade do que se é possível observar no mundo real. Pequenas variações no material do fundo – especialmente de suas características reflexivas – fazem com que pneus muito próximos, em condições de iluminação semelhantes, destoem razoavelmente entre si. O pneu inserido parcialmente sobre a parede da pequena construção, detalhado nas Figuras 4.11 e 4.12, exemplifica o fenômeno, ao ficar mais claro do que originalmente



Figura 4.11: Objeto sobre parede – método original



Figura 4.12: Objeto sobre parede – método atual

estava. A coloração intensamente branca da parede justifica esse comportamento, e acende o sinal de alerta.

Dado o contexto de nossas gravações, tomadas sempre em regiões externas, à luz do sol, é necessário buscarmos algum tipo de referência global para os objetos de aumento que são inseridos nos mesmos quadros. Para isso, é desenvolvida uma base de dados de pneus reais, e se muda por completo a definição dos valores alvo.

Essas mudanças são abordadas no próximo capítulo deste trabalho.

Capítulo 5

Repensando o Processo: Criação da Base de Pneus Reais

O novo desafio compreende tentar reduzir, de alguma forma, a sensação de artificialidade descrita previamente. No “mundo real”, em ambientes externos, é normal que o nível de luminosidade de objetos em diferentes áreas de uma mesma localidade sejam similares (dado que sejam compostos por materiais não muito diferentes em sua relação com a luz). Apenas em regiões de sombra – ou sob efeitos peculiares de reflexão – percebe-se variação significativa desta característica.

Contudo, descrever o comportamento que os objetos de aumento devem assumir em cada situação não é trivial. É necessário que estabeleçamos algum padrão de análise que nos ajude a determinar como realizar esse mapeamento. Daí, surge a ideia de utilizar os pneus reais capturados nos vídeos como referencial para os objetos de aumento.

5.1 A Base de Pneus Reais

A construção desse referencial se alicerça sobre uma base de dados composta por informações extraídas de pneus reais. Os vídeos da MBG, como descrito no Capítulo 2, são acompanhados de marcações que descrevem em que quadros – e em quais coordenadas dentro desses quadros – é possível localizar os objetos que se deseja detectar. Desta forma, torna-se simples a tarefa de extraí-los individualmente.

Essa facilidade nos permite construir uma base de imagens composta por

cerca de 100 pneus reais – diferentes versões da base foram construídas ao longo do tempo, dadas pequenas alterações de concepção, o que resulta em pequenas variações no número total de objetos contidos na base. Essas imagens sempre contêm o pneu, centralizado, e uma região de *background* a seu redor. O quanto de *background* será incluído ao redor do pneu se configura como um hiperparâmetro, que é variado arbitrariamente para obtenção de comparativos experimentais.

A partir dessas imagens, é realizado processo de extração de descritores. A base que é, de fato, utilizada em nosso fluxo é composta apenas pelos dados numéricos obtidos – o que reduz seu tamanho e a complexidade de sua utilização.

5.1.1 Montagem da Base

A montagem da base começa com a extração dos objetos de interesse a partir dos quadros dos vídeos da MBG. Para isso, utiliza-se um algoritmo que percorre as anotações dos vídeos: para cada anotação, ele realiza o *crop* do quadro correspondente, com base nas coordenadas marcadas para cada objeto, e salva os resultados em novos arquivos de imagem. Esse processo resulta em uma grande quantidade de imagens, dado que as marcações são feitas a nível de quadro, e não a nível de objeto.

No próximo passo, selecionam-se as imagens extraídas, retirando-se as redundâncias. Também são descartadas as imagens nas quais os objetos de interesse estão ocluídos, ou em que a qualidade da captura se encontra mais baixa, sendo difícil distinguir com precisão a fronteira entre os objetos e o fundo.



Figura 5.1: Imagem ilustrativa do modelo de mesa digitalizadora utilizado

Com a seleção final de imagens pronta, começa o processo de marcação dos objetos e do fundo. Com o auxílio de uma mesa digitalizadora *Wacom Bamboo*, são desenhadas duas máscaras por imagem: a primeira, contendo a região da imagem que

pode ser definida, com certeza, como fazendo parte do pneu; a segunda, seguindo os mesmos preceitos, contendo o fundo. As regiões de incerteza fronteira, as marcadas por sombras ou reflexões muito destoantes e aquelas que contém outros objetos que não o de interesse são excluídas de ambas as máscaras. Exemplos do processo – e de seus resultados – podem ser observados nas Figuras 5.2, 5.3 e 5.4.



Figura 5.2: Exemplos de *crop* dos objetos de interesse



Figura 5.3: Exemplos de máscaras para os mesmos pneus

Neste ponto, nossa base conta com três imagens por objeto, totalizando cerca de 300 itens. Esses servem de base para o próximo passo: a extração de descritores.

5.1.2 Extração de Descritores

A extração segue processo similar àquele realizado sobre os pneus de aumento e seus respectivos fundos, descrito no Capítulo 4. Para cada objeto da base, realiza-se

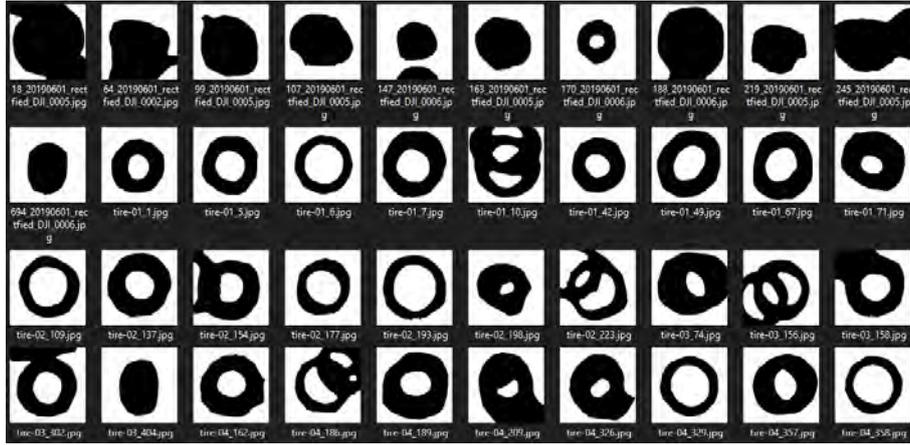


Figura 5.4: Exemplos de máscaras para os fundos

a transformação do espaço de cores para $YCbCr$ e utiliza-se o o canal Y , responsável pelo mapeamento de Luma, para a extração dos descritores de brilho e contraste.

Esse processo é realizado duas vezes: uma sobre a região delimitada pela máscara que representa o objeto de interesse, e uma sobre a região delimitada pela máscara representando o fundo. Ao final deste processo, temos, para cada pneu, quatro valores mapeados: brilho do pneu, constraste do pneu, brilho do fundo e contraste do fundo.

Esses valores são embutidos em quadrivetores, que são salvos, um por linha, em um arquivo `csv`. Esse arquivo compacto, de cerca de 100 linhas, é a forma final da primeira versão de nossa base de dados de pneus reais, e é de suma importância para o método descrito a seguir. Um trecho desse arquivo pode ser observado na Figura 5.5.

mean_tire	std_tire	mean_back	std_back
76,1163575	29,7973675	97,8063819	42,031058
137,765203	26,1594446	165,5328825	45,5337569
140,3044205	36,3441014	130,7432556	51,0501092
136,2508787	34,2648006	117,0211361	52,4943121
82,2172235	10,39158	98,9027825	20,198142
89,0854341	37,4920867	109,0167769	54,6422344

Figura 5.5: Início do arquivo de uma das versões da base de dados de pneus reais

5.2 Casamento de Luminância via Base de Dados

A base de dados de pneus reais nos permite tentar inferir o comportamento que os objetos de aumento devem seguir quando inseridos nos quadros dos vídeos da base de dados MBG. A intuição por detrás das informações concentradas nessa nova base é a de que existe relação direta entre as características de luminância do pneu e aquelas do fundo sobre o qual ele está. Em outras palavras, conhecendo-se as características do *background* sobre o qual o objeto de aumento será inserido, é possível definir que comportamento o objeto deverá assumir se observarmos as relações mapeadas na base.

O método aqui apresentado busca encontrar, dentre as possibilidades mapeadas na base de dados de pneus reais, o fundo que mais se adequa a cada situação de inserção. Com o fundo escolhido, é fácil extrair do quadrivetor correspondente os valores de brilho e contraste relativos ao pneu real associado a ele. Esses valores são, então, utilizados como guia para a transformação dos objetos de aumento.

5.2.1 Comparação de *Patches*

Mesmo após definir-se um modelo para o comportamento dos objetos de aumento, resta ainda um desafio: como encontrar um único conjunto de valores de referência de luminância para todo um quadro? Em outras palavras, a base de dados de pneus reais resolve o problema de, definindo-se o fundo, conseguirmos realizar as transformações necessárias sobre o objeto de aumento; porém, ainda não resolve qual fundo deve ser o escolhido.

Para conseguirmos sanar as incongruências observadas na conclusão do capítulo anterior, essa escolha precisa culminar em resultados consistentes para toda a extensão de um mesmo *quadro*. Isso faz com que nossa nova metodologia precise contar com boa capacidade de generalização, sendo capaz de observar características de todas as regiões da imagem – e não só da região imediatamente ao redor do objeto, como no método anterior. À solução que apresentamos para esse desafio demos o nome de comparação de *patches*.

A comparação de *patches* busca generalidade enquanto tenta manter a atenção às características locais de diferentes pontos do quadro. Para isso, divide a imagem

em diferentes subseções, de igual tamanho. Para cada uma delas, realiza a extração dos descritores de brilho e contraste, seguindo a mesma metodologia já descrita. Então, os compara com todas as entradas da base de dados de pneus reais: para cada quadrivetor da base, são comparados com os descritores recém-extraídos os dois valores que correspondem ao mapeamento do *background* do pneu real. Dentre todas as comparações, aquela com maior similaridade é salva.

Ao final do processo, os casamentos de todas as subseções são comparados, e aquele com o melhor resultado é escolhido como referência para o fluxo. Além deste, todos os casamentos que estiverem dentro de uma pequena margem arbitrária – se houver algum – são também marcados como referência, para que exista a possibilidade de pequenas variações. Um esquema simplificado do processo pode ser observado na Figura 5.6.

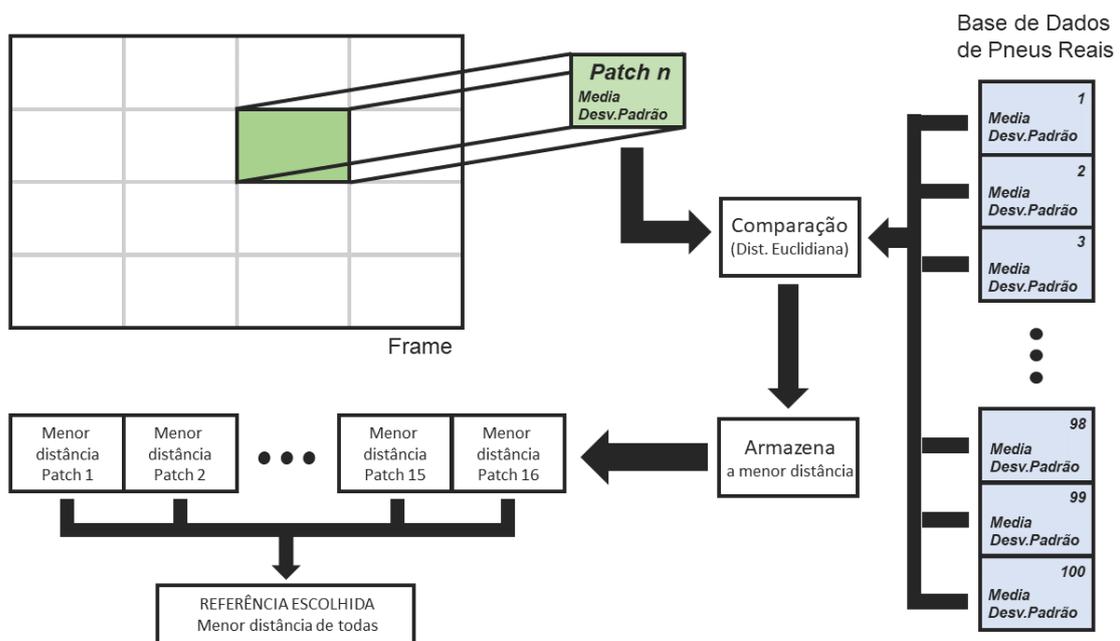


Figura 5.6: Exemplo do fluxo concebido

A comparação utiliza como métrica de similaridade a distância euclidiana entre os vetores, de acordo com a expressão

$$Similaridade_{v_1, v_2} = \sqrt{(v_{1b} - v_{2b})^2 + (v_{1c} - v_{2c})^2}$$

Aqui v_1 e v_2 são bivectores, da forma $v_1 = [v_{1b} \ v_{1c}]$ e $v_2 = [v_{2b} \ v_{2c}]$, onde o primeiro elemento representa o brilho e o segundo, o contraste.

Para utilização dessa métrica, os vetores passam por processo de normalização

$$v_{normalizado} = \frac{v_{base} - v_{medio}}{v_{\sigma}}$$

onde v_{base} é o bivector composto pela dupla brilho e contraste que descreve o objeto a ser normalizado; v_{medio} é o bivector composto pelos valores médios de brilho e contraste calculados ao longo da inteiridade da base de dados de pneus reais, e v_{σ} é o bivector composto pelo desvio padrão de ambos os descritores, também calculado sobre toda a base de pneus reais.

5.2.2 Casamento de Luminâncias

O *pipeline* completo do casamento de luminâncias mantém os dois adendos trazidos pelo método apresentado no Capítulo 4, mas promove algumas alterações no primeiro deles. Agora, os valores de referência de luminância deixam de ser calculados com base na região ao redor de cada pneu, passando a ser definidos pelo fluxo de comparação de *patches* descrito acima.

O resultado da comparação de *patches* é generalizado, e todos os objetos de aumento que compartilham do mesmo quadro inicial passam a ter valores de referência similares para brilho e contraste. Apenas nos casos em que mais de um casamento é considerado apto para um quadro, uma segunda fase é adicionada nesse processo.

Nesses casos, uma nova rodada de casamentos é realizada entre cada um dos objetos de aumento do quadro e cada um dos valores de referência selecionados pela comparação de *patches*. Assim como no método anterior, esses novos casamentos são feitos entre os descritores de brilho e contraste extraídos da área que cerca o objeto de aumento e os valores de referência extraídos da base – neste caso, apenas a pequena seleção correspondente ao quadro em questão. Nessa nova etapa, os valores de referência considerados como finais passam a ser aqueles que apresentam maior similaridade com cada um dos pneus.

5.2.3 Resultados e problemas

Os resultados obtidos com este método apresentam ganhos visuais de verossimilhança consideráveis em relação aos obtidos nos *pipelines* anteriores. O ajuste

de luminância faz com que os objetos se adequem melhor às diferentes condições de luminosidade encontradas nos vídeos enquanto o método de comparação de *patches* soluciona as discrepâncias locais observadas no Capítulo 4.



Figura 5.7: Exemplo de *quadro* do vídeo 0019 contendo pneus inseridos com o método de casamento de luminâncias e um pneu real



Figura 5.8: Exemplo de *quadro* do vídeo 0038 contendo pneus inseridos com o método de casamento de luminâncias e um pneu real

As Figuras 5.7 e 5.8 mostram resultados do processo de aumento de dados em vídeos gravados em dias mais claros. Ambas as imagens contém pneus reais, que estão marcados. Comparando os pneus de aumento com os reais, é possível perceber que o processo de aumento, se desconsiderarmos o contexto posicional dos objetos, gera resultados convincentes.



Figura 5.9: Exemplo de *quadro* do vídeo 0033 contendo pneus inseridos com o método de casamento de luminâncias

A Figura 5.9, por sua vez, mostra o resultado do processo em um dos vídeos mais escuros. Nesse caso, embora o resultado geral seja convincente e haja consistência entre os diferentes pneus contidos no quadro, ainda é possível identificar artificialidade nos objetos de aumento por motivos que ultrapassam incongruências de posicionamento. Não só é perceptível que os pneus estão um pouco mais claros do que deveriam, mas que também há algo em seus aspectos textural e de coloração que parece destoar do fundo.

Esses resultados parecem nos indicar que, através da construção da base de pneus reais e do método de casamento de *patches*, conseguimos refinar nosso fluxo para que a definição de nossos parâmetros de referência seja satisfatória. Contudo, a manutenção da sensação de artificialidade nos vídeos escuros nos leva novamente para a mesa de projetos. Solucionadas, momentaneamente, as proposições direcionadas especificamente à luminância, decidimos trabalhar o aspecto que mais nos chama

atenção em nossos novos resultados: a baixa correlação entre as características reflexivas e texturais dos objetos de aumento e as que dizem respeito aos demais objetos retratados nas gravações que formam nossa base.

Com esse objetivo, redirecionamos nosso foco para o outro elemento que adicionamos ao nosso *pipeline*: a transformação dos objetos de aumento. Na próxima seção, discorreremos sobre uma metodologia que redefine esse processo. Ela se utiliza de múltiplos descritores, que cobrem outras características além da luminosidade, e é construída em cima de um método de transformação de maior complexidade e flexibilidade.

5.3 Casamento de Histogramas de Cor via Base de Dados

O processo de desenvolvimento desse trabalho perpassa a realização constante de pesquisa bibliográfica. Melhorar os resultados obtidos pelos métodos anteriores incentiva a busca por novas ideias. Esse processo nos pôs em contato com diversos artigos e trabalhos, que ajudaram a direcionar aspectos de nosso desenvolvimento.

A partir de artigos como o em [25], começamos a utilizar histogramas tanto na extração de descritores – e na subsequente comparação de objetos – quanto na transformação final do pneu de aumento. A metodologia desenvolvida sobre este conceito altera, portanto, os dois pontos de contato que fazemos com o *pipeline* existente: a seleção dos valores de referência e o processo de transformação dos objetos de aumento.

5.3.1 Extração de novas *features*

O primeiro ponto de mudança se dá na substituição dos descritores utilizados. Como abordado ao final da seção anterior, buscamos descritores que sejam capazes de interpretar não apenas as relações de iluminação mais imediatas, mas também as de texturização e coloração dos objetos e fundos. Para isso, é necessário que adicionemos novas camadas de profundidade ao nosso método de extração.

Considerados insuficientes, brilho e contraste, dois valores escalares, são substituídos por um conjunto de quatro histogramas. Os três primeiros são construídos

em cima dos três canais da imagem que se deseja analisar; o quarto, em cima de padrões binários locais (*local binary patterns, LBP*).

5.3.1.1 *Local Binary Patterns*

LBP são utilizados há quase duas décadas como descritores texturais de alta eficiência e baixa complexidade computacional. São baseados na construção de representações locais de textura, através da comparação entre cada pixel e sua área de vizinhança imediata. Como são extraídos exclusivamente de imagens em escala de cinza, levam em consideração apenas aspectos texturais e de iluminação, desconsiderando caracterizações de cor [26].

Para cada pixel, seleciona-se uma vizinhança de tamanho e ordenação fixos a seu redor. Cada um dos pixels da vizinhança é, então, comparado – binariamente – com o pixel central: se a intensidade do pixel vizinho for maior do que a do pixel central, assumimos o valor 0; de outra forma, assumimos 1. Concatenando-se todos os pixels binarizados da vizinhança, obtemos um vetor binário. Este vetor é traduzido para um número decimal, e este número passa a ser o valor *LBP* do pixel central.

Esse processo é reproduzido para todos os pixels da imagem de interesse, e é obtida uma imagem em *LBP*. O trabalho apresentado em [26] nos diz que o histograma das *features* obtidas através de *LBP* possui capacidade representativa importantíssima em relação às condições de textura e iluminação da imagem original e, portanto, pode ser utilizado como descritor dessas propriedades.

5.3.1.2 O processo de extração

Para realizarmos a extração de fato dos descritores, é preciso definir primeiro qual espaço de cores utilizar. Essa escolha influi de forma significativa na ênfase que será dada a cada característica da imagem. Por exemplo, utilizar espaços de cores com canais específicos para a representação de luminância ou de luma dá ênfase a representações diretas de iluminação.

O padrão de referência que escolhemos se baseia no trabalho em [25], e começa construindo histogramas normalizados para cada um dos três canais de uma imagem em *HSV*. Esses histogramas são definidos como os descritores desses canais.

Um quarto descritor é construído, então, a partir de *LBP* extraídos da versão em escala de cinza da mesma imagem. Para tanto, é realizado processo de conversão da imagem para escala de cinza e de construção da imagem em LPB; desta é extraído o quarto e último histograma. Os quatro histogramas, conjuntamente, representam a totalidade dos descritores de nossa imagem.

A representação numérica dos descritores é construída da seguinte forma: para um dado histograma, cada um dos *bins* é tratado como uma componente de um vetor n -dimensional, onde n equivale ao número de *bins* utilizado; o n -vetor resultante da concatenação de todas as componentes é considerado como o descritor daquele histograma, e é salvo.

5.3.2 Mudanças na comparação de *patches*

As alterações quantitativas e qualitativas realizadas sobre a extração de descritores não altera a estrutura geral da comparação de *patches*. O processo, compreendido pela comparação dos descritores dos *backgrounds* dos diferentes *patches* com os descritores dos *backgrounds* dos pneus reais contidos na base de dados, se mantém. Se mantém, também, o conceito de generalização a nível de quadro, que busca homogeneizar os resultados do processo de aumento de dados.

A diferença se dá no processo de comparação em si. Primeiro, por conta do maior número de elementos que são utilizados para comparação: passa-se de dois escalares para quatro vetores n -dimensionais. O número de dimensões de cada histograma é variável, e múltiplas combinações foram experimentadas (64 *bins* para cada um dos quatro histogramas; 128 *bins* para cada histograma; 16 *bins* para o *LBP* e tanto 64 quanto 128 *bins* para cada canal da imagem, entre outros), variando de 208 a 512 componentes totais.

Segundo, pela incapacidade de se relacionar diretamente as diferentes componentes que são utilizadas para o cálculo de similaridade: se, antes, brilho e contraste possuíam relação direta, os histogramas construídos não possuem mais essa particularidade, dado que o significado das magnitudes dos canais em *HSV* e das *features* em *LBP* não possuem correlação imediata. Desta forma, para tentarmos minimizar influências indevidas, seguimos novamente [25] e mudamos nossa métrica de similaridade: deixamos de lado a distância euclidiana, e assumimos a similaridade de

cosseno.

A similaridade de cosseno é, em suma, a medida do cosseno do ângulo entre dois vetores, e é representada por

$$\text{Similaridade} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

onde A representa o primeiro vetor da comparação e B , o segundo. Devido à sua definição, abstrai o módulo dos vetores que compara, carregando em si apenas proximidade angular entre os elementos: quanto mais próximos, angularmente, forem os vetores, maior será o valor da similaridade calculada. Considera-se que esta medida seja mais apropriada para casos como o nosso, onde não é possível definir qual deve ser a influência dos módulos de cada um dos elementos descritores.

5.3.3 Casamento de Histogramas

Outra mudança diz respeito à transformação dos pneus de aumento quando de sua inserção nos vídeos da base MBG. Até agora, os objetos passavam por alterações relativas a suas componentes mapeadas no canal de luma ou luminância, através de simples processo de normalização e desnormalização para os novos valores. A partir de agora, é desenvolvido algoritmo vetorial de casamento de histogramas, responsável por aproximar as distribuições observadas nos histogramas dos objetos de aumento daquelas encontradas nos histogramas dos pneus de referência contidos na base de dados de pneus reais e escolhidos por via do casamento de *patches*.

5.3.3.1 Construção teórica

O casamento de histogramas tem como princípio o mapeamento entre um histograma de origem, que se quer mudar, e um histograma de referência, que funciona como alvo. Tratando-se os histogramas de distribuições, pode-se dizer que esse processo é análogo ao mapeamento entre a função de distribuição acumulada (*cumulative distribution function, CDF*) do histograma de origem e a *CDF* do histograma alvo.

Calcular a *CDF* de um histograma configura tarefa trivial. A função de densidade de probabilidade (*probability density function, PDF*) de cada canal de nossa imagem pode ser obtida, a partir do histograma correspondente, através da

relação

$$p_r(r_j) = \frac{n_j}{n}$$

onde n_j representa a quantidade de pixels com valor de intensidade r_j , e n representa o número total de pixels da imagem. Podemos, então, realizar o mapeamento dessa *PDF* em sua *CDF* correspondente seguindo

$$S(r_k) = \sum_{j=0}^k p_r(r_j), \quad k = 0, 1, \dots, L$$

onde L equivale ao número total de níveis de intensidade existentes – em nosso caso, de imagens codificadas em oito *bits*, $L = 255$.

Realizando-se esse processo em ambas as imagens, obtemos duas *CDFs* por canal: uma *CDF* de origem e uma *CDF* de alvo. Com ambas em mãos, o processo de mapeamento consiste em, para cada um dos 256 valores de intensidade contidos na *CDF* de origem, encontrar na *CDF* de alvo o valor de intensidade que compartilha da mesma probabilidade.

Em suma, considerando C_o como a *CDF* de um dos canais da imagem de origem e C_a como a *CDF* do mesmo canal na imagem alvo, para cada nível de intensidade I_{orig} do canal de origem precisamos encontrar um nível de intensidade I_{alvo} no canal alvo para o qual $C_o(I_{orig}) = C_a(I_{alvo})$. Realizando esse processo para todos os 256 níveis de intensidade possíveis para o canal, encontramos uma função de mapeamento M tal que $M(I_{orig}) = I_{alvo}$.

Esse mapeamento requer que realizemos mais uma expansão do conteúdo da base de dados de pneus reais: após os histogramas utilizados para comparação com os *patches*, são concatenados aos vetores de cada elemento da base três histogramas de 256 *bins* (um para cada canal da imagem), para que possa ser realizado o método descrito acima.

5.3.3.2 Transformação dos objetos de aumento

O processo prático começa com o cálculo de três histogramas: um para cada canal da imagem que contém o objeto a ser inserido. Cada um dos histogramas é mapeado através de dois vetores: o primeiro contém os valores do histograma em si, e descreve as incidências de cada intensidade; o segundo contém os limites que definem cada um dos intervalos do histograma. Os histogramas são calculados

apenas na região da imagem que contém o objeto de interesse, através da utilização de máscara. Um esquemático do processo pode ser observado na Figura 5.10. Ao final dessa seção, contamos com seis vetores armazenados.

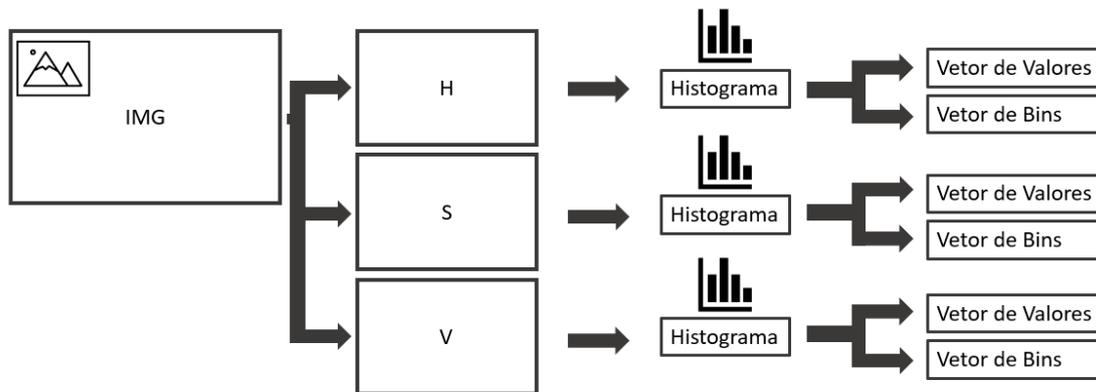


Figura 5.10: Extração dos vetores descritores dos histogramas

Então, são extraídos da base de dados de pneus reais os dados referentes aos mesmos três histogramas, mas da imagem que contém o objeto de referência escolhido pela comparação de *patches*. Novamente, seis vetores são armazenados, totalizando 12 vetores.

Calculam-se as *CDFs* dos seis vetores que representam os valores dos histogramas – três para o objeto de aumento, três para o objeto de referência. Os seis vetores resultantes dos cálculos das *CDFs*, assim como os quatro vetores que representam os limites dos *bins* de cada histograma, são alimentados para uma função que faz o processo de casamento. Junto a eles, são alimentadas também as matrizes que descrevem cada um dos canais da imagem do objeto de aumento. O processo é realizado canal a canal.

A realização do casamento segue os princípios teóricos descritos na subseção anterior. O algoritmo, construído em *Python* com auxílio da biblioteca de operações vetoriais *numpy*, constrói o mapeamento entre as *CDFs* do objeto de aumento e do objeto de referência através de um processo de interpolação linear. Esse processo toma a *CDF* do objeto de aumento como as coordenadas do eixo x sobre o qual a interpolação será montada, a *CDF* do objeto de referência como as coordenadas do eixo x dos dados a serem interpolados e os *bins* do objeto de referência como as coordenadas do eixo y destes mesmos dados.

Com o mapeamento construído, são criados os novos canais da imagem. Esse

processo é realizado através de um *loop*, que percorre apenas os valores de intensidade que aparecem pelo menos uma vez na imagem original do objeto de aumento. Essa restrição no laço de repetição resulta em número menor de iterações e, conseqüentemente, em diminuição do tempo gasto no processo.

Ao final, temos as versões transformadas de cada canal da imagem que contém o objeto de aumento. Essas versões sobrescrevem os canais originais e, a partir delas, é gerada a nova imagem de aumento. Como o método de inserção cola sobre o fundo apenas a porção da imagem relativa ao objeto, possíveis mapeamentos indevidos sobre o *background* da imagem podem ser desconsiderados. A Figura 5.11 apresenta uma representação visual do processo.

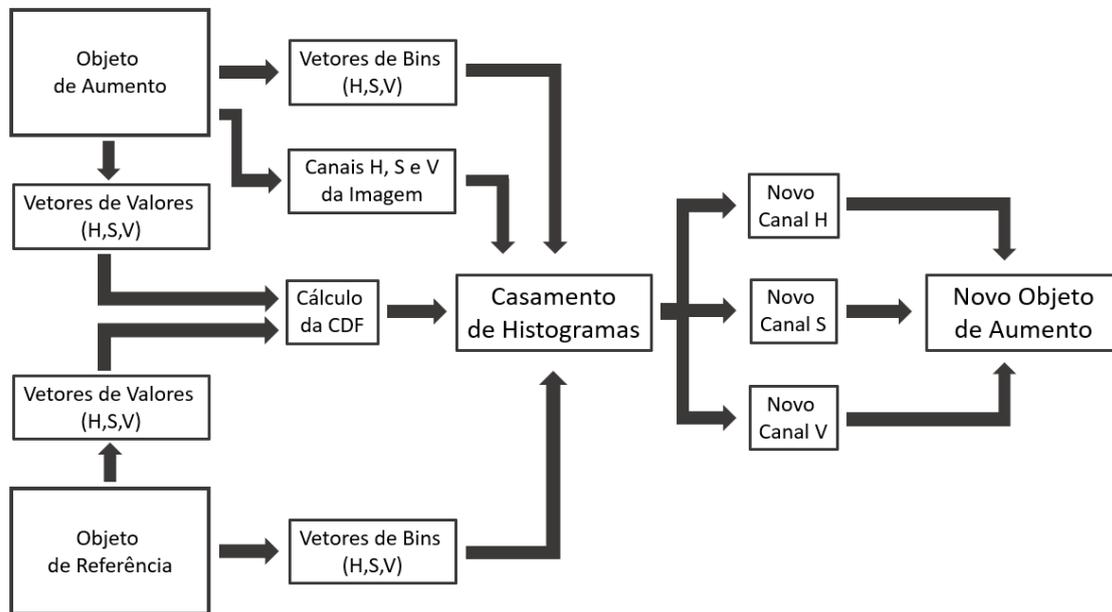


Figura 5.11: Visão geral do processo de transformação

5.3.4 Resultados

O método desenvolvido apresenta resultados que, se destituídos de contexto, parecem inverossímeis: aplicar o método sobre imagens de pneus isolados, colados sobre fundos brancos e separados dos *backgrounds* sobre os quais serão, de fato, inseridos, resulta em imagens de saída inusitadas, como pode ser observado na Figura 5.12.

Contudo, quando observamos sua inserção nos contextos esperados, é possível

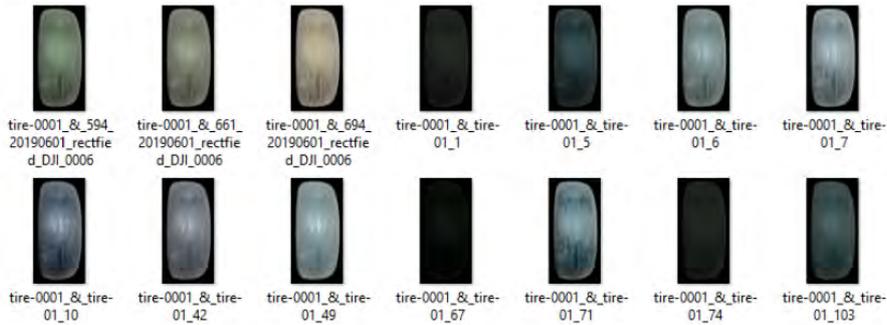


Figura 5.12: Exemplos do método isolado

perceber que as mudanças de coloração, iluminação e de aspecto textural geral dos objetos os aproxima, mais do que nunca, de seu alvo.

Especialmente se levarmos em consideração os vídeos mais escuros contidos na MBG, fica clara a capacidade adaptativa do método. Os objetos inseridos se aproximam dos observáveis nos casos reais, como pode ser constatado na comparação realizada na imagem 5.13. Nela, o primeiro pneu é o real, o segundo é o inserido pelo método original, desenvolvido em [5], e o terceiro corresponde ao método de casamento de histogramas.



Figura 5.13: Comparação em vídeo escuro – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor

Nos vídeos mais iluminados, é possível observar com clareza a sutil influência da mudança de matiz no resultado final. Ante os pneus gerados pelos métodos anteriores, sempre marcados pela mesma coloração acizentada, independentemente de contexto, encontramos agora diferentes tendências de coloração, que variam com o ponto do *background* sobre o qual o objeto é inserido. A Figura 5.14 ilustra um caso com variação tendendo ao verde (com o pneu posicionado sobre ambiente gramado),

enquanto a Figura 5.15 mostra os efeitos de uma variação tendendo ao azul (com o objeto localizado sobre um piso de concreto). Em ambas as imagens, a ordem de aparição das Figuras é semelhante à comparação em vídeo escuro: um pneu real é seguido do método original e, por fim, do casamento de histogramas.



Figura 5.14: Comparação de objetos sobre gramado – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor



Figura 5.15: Comparação de objetos sobre concreto – na ordem, um pneu real, um pneu artificial inserido pelo método desenvolvido em [5] e, por fim, um pneu artificial inserido pelo método de casamento de histogramas de cor

A percepção humana não se configura, como já abordado previamente, em sistema trivial. A combinação de fatores que resulta em uma imagem perpassa elementos que passam despercebidos ao observador, até que (seja necessário que) se preste mais atenção. Da pintura em tela até os compilados de ilusão de ótica encontrados na internet, são múltiplos os exemplos que provam que a composição por detrás de imagens – seja qual for o meio utilizado para sua construção – carrega relações que vão muito além do óbvio. Os resultados obtidos aqui servem de breve evidência dessa afirmação.

Capítulo 6

Resultados Subjetivos

Em todo projeto de pesquisa é necessário que se tomem decisões em relação a qual é o momento certo para realizar o corte que separa o que entrará em um trabalho como este, e o que precisará ser levado como trabalho futuro. Assim, por questões logísticas e de prazo, tomamos a decisão de abdicar, nesse texto, da apresentação de resultados de detecção por parte da rede neural.

Como forma de validar de alguma forma os resultados obtidos concebemos uma metodologia simples para a condução de testes subjetivos. Sua construção se baseou na decisão de projeto inicial que impeliu a criação dos fluxos de aumento de dados contidos nesse texto: a premissa de que a adequação contextual dos objetos de aumento geraria resultados positivos de detecção.

Vale ressaltar que a revisão bibliográfica nos trouxe evidências tanto favoráveis [21] quanto contrárias [20] à adequação contextual de objetos de aumento. Portanto, os testes subjetivos que propomos não têm como objetivo substituir a detecção por Rede Neural que havíamos intencionado inicialmente: não é possível realizar ilação entre os resultados obtidos pela pesquisa subjetiva e aqueles que serão obtidos, em momento futuro, pela rede neural.

Contudo, é possível validar se o objetivo fundamental de nossos métodos de aumento foi cumprido: adicionar informação contextual aos objetos de aumento de forma a torná-los mais verossímeis em relação ao fundo sobre o qual são inseridos.

6.1 Elaboração do Teste Subjetivo

Para a realização do teste subjetivo, foram criados dois questionários base: um para o método de modificação de luminância a partir da base de dados de pneus reais, e um para o método de casamento de histogramas de cor.

Ambos seguem metodologia similar, e foram aplicadas para os mesmos 16 respondentes. Primeiro, são apresentados aos participantes alguns exemplos de pneus reais, em diferentes condições. Para tentar manter o máximo de consistência entre os pneus reais e os pneus artificiais que serão apresentados, foram selecionados casos específicos, onde há pouca influência de sombras e onde os objetos se encontram dispostos horizontalmente. Exemplos podem ser observados na Figura 6.1.

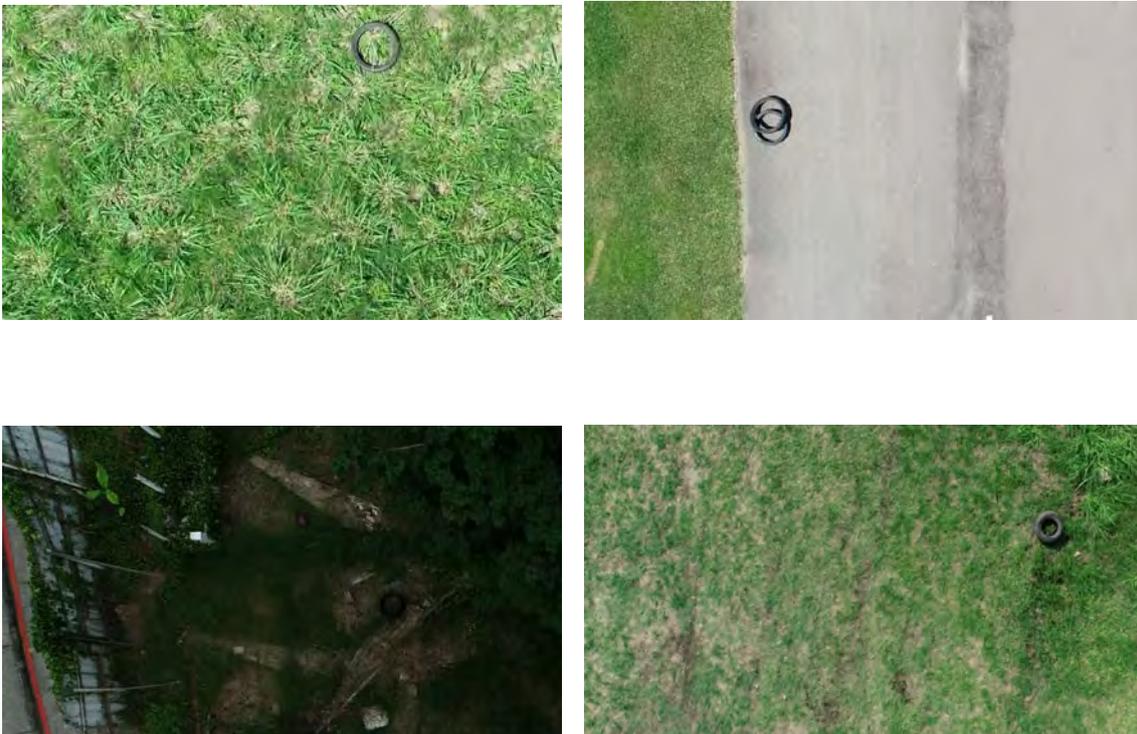


Figura 6.1: Exemplos de Pneus Reais Selecionados

Então, são apresentados, em ordem aleatória, pares de pneus artificiais: um deles correspondente à metodologia desenvolvida em [5], e o outro ao método que se está avaliando. A ordem com a qual o par é apresentado também é aleatória. Como tentativa de diminuição de possíveis tendências avaliativas – dada a importância contextual para a percepção humana – os dois quadros são, salvo a metodologia aplicada, exatamente iguais: o mesmo objeto de aumento é aplicado exatamente no

mesmo ponto sobre o mesmo quadro. Todos os quadros utilizados como fundo para a inserção dos objetos de aumento são também escolhidos de forma a se parecerem, o máximo possível, com as imagens ilustrativas de pneus reais dispostas anteriormente.

Devido às condições de isolamento social impostas pela atual pandemia, os testes foram conduzidos remotamente, nos dispositivos pessoais dos respondentes, sem controle específico das condições de iluminação e visualização dos ambientes, e sem padronização dos dispositivos de exibição das imagens e de distância entre o observador e o monitor.

Quatro exemplos de pares utilizados na pesquisa, sendo dois de cada método, podem ser observados nas figuras ao longo da seção (os pares apresentados nas Figuras 6.2 e 6.3 para o casamento de histogramas, e os apresentados nas Figuras 6.4 e 6.5 para a adequação de luminância utilizando a base de pneus reais).



Figura 6.2: Exemplo de par para casamento de histogramas



Figura 6.3: Outro exemplo de par para casamento de histogramas

Ao participante cabe decidir qual dos dois elementos do par considera mais verossimilhante. Suas respostas são anotadas, de forma anônima, e compiladas para a geração de resultados. Nenhum dos participantes da pesquisa subjetiva teve



Figura 6.4: Exemplo de par para adequação de luminância

qualquer forma de contato prévio com os resultados – ou com qualquer aspecto – dos métodos.



Figura 6.5: Outro exemplo de par para adequação de luminância

6.2 Resultados

Os dois formulários apresentam similaridades nos padrões de resposta obtidos. Em ambos os casos, os objetos de aumento gerados pelas metodologias desenvolvidas neste trabalho se mostraram, para os participantes de nossa pesquisa, mais convincentes do que aqueles gerados pelo método desenvolvido em [5] – que aqui referenciamos como método original.

Para o método de adequação de luminância, os pneus gerados pelo novo método foram escolhidos em cerca de 60% dos casos, contra 40% dos gerados pelo método original. Para o método de casamento de histogramas, os objetos de aumento gerados superaram os do método original em cerca de 64% dos casos.

Os avanços em relação ao método original se aprofundam quando levamos em conta os resultados obtidos nos vídeos mais escuros. Afinal, a discrepância

visual observada entre os objetos de aumento gerados pelo método original e os fundos extraídos de vídeos escuros se configura como principal motivação para o desenvolvimento das novas metodologias – e resolver essa incongruência é o principal objetivo dos novos fluxos.

Para o método de adequação de luminância, os pneus gerados pelo método novo nos quadros extraídos de vídeos escuros foram escolhidos, em média, por cerca de 82% dos participantes. Já para o método de casamento de histogramas, essa proporção sobe para cerca de 91% dos participantes. Um compilado dos resultados pode ser conferido nas tabelas 6.1 e 6.2.

Tabela 6.1: Resultados do teste para o método de casamento de luminância

Teste 1	Todos os quadros	Quadros extraídos de vídeos escuros
Método desenvolvido em [5]	40%	18%
Método de Casamento de Luminâncias	60%	82%

Tabela 6.2: Resultados do teste para o método de casamento de histogramas de cor

Teste 2	Todos os quadros	Quadros extraídos de vídeos escuros
Método desenvolvido em [5]	36%	9%
Método de Casamento de Histogramas de Cor	64%	91%

Vale ressaltar também que, fora os vídeos escuros, todos os demais quadros selecionados envolvem ambientes similares aos que serviram de base para o método original (dias claros, e superfícies planas cobertas principalmente por grama), de forma que a falta de adaptabilidade é pouco penalizada pela pesquisa.

Em suma, as pesquisas subjetivas realizadas validam nossa hipótese inicial: há ganho progressivo de verossimilhança pelos métodos desenvolvidos, especialmente no tocante a sua capacidade adaptativa.

Capítulo 7

Conclusão e Trabalhos Futuros

Começamos esse trabalho com a exposição de nosso objetivo: auxiliar no processo de detecção de criadouros de mosquito *Aedes aegypti* iniciado nos trabalhos [5] e [7].

Então, partimos para uma revisão teórico-metodológica do *pipeline* desenvolvido em [5], começando pelo processo de construção da base de vídeos que é utilizada para treino do sistema de detecção. Descrevemos seu processo de criação, as características de seu conteúdo e seu estado atual.

Prosseguimos com a revisão dos três métodos de inserção de objetos de aumento construídos nesse *pipeline*. Abordamos as referências bibliográficas sobre as quais os métodos foram construídos e destrinchamos o funcionamento dos métodos tanto a nível de quadro quanto a nível de vídeo. Fazemos, então, a avaliação dos resultados obtidos. Identificamos, como ponto de atenção, a não adaptabilidade dos objetos de aumento em relação à iluminação dos *backgrounds* sobre os quais são inseridos.

Com base nessa conclusão, propusemos o desenvolvimento de um primeiro método, que realiza a mudança dos valores de luminância dos objetos de aumento de acordo com a região de *background* sobre a qual são inseridos. Abordamos sua construção, sua fundamentação teórica e sua estrutura. Então, avaliamos os resultados obtidos, e sugerimos a generalização, a nível de *quadro*, dos parâmetros utilizados como referência para a transformação dos objetos de aumento.

Criamos, em seguida, uma base de objetos reais, e descrevemos desde o tratamento das imagens até a extração e armazenamento dos descritores. Propusemos

um método de transformação de luminância análogo ao anterior, com a distinção de usar como referência não mais o *background* do quadro, mas um objeto de referência contido na base construída. Apresentamos, por fim, um algoritmo de seleção que define automaticamente qual elemento da base de objetos reais deve ser utilizado como referência para a transformação.

Novamente, analisamos os resultados, e identificamos novas possibilidades de melhoria. Estas culminaram na proposta de alterar tanto a seleção do objeto de referência quanto o processo de transformação dos objetos de aumento. Para a seleção, propusemos a utilização dos histogramas dos canais da imagem e de *local binary patterns* como descritores dos objetos; para a transformação dos objetos de aumento, construímos um método de casamento de histogramas com o objeto de referência. Foram abordadas a fundamentação teórica e a estrutura algorítmica do método resultante dessas mudanças.

Realizamos, ao final, um teste subjetivo que compara os métodos desenvolvidos nesse trabalho com aquele construído em [5]. O teste valida a hipótese de que os métodos desenvolvidos resultam em ganhos progressivos de verossimilhança.

Como trabalhos futuros, pretendemos realizar testes de detecção com versões diferentes da base de dados: uma versão sem aumento de dados, e uma versão aumentada para cada um dos métodos desenvolvidos. Temos como expectativa extrair métricas comparativas que sejam diretamente relacionáveis a nosso objetivo final.

Então, pretendemos realizar a variação dos parâmetros que regem o algoritmo de seleção, gerando novas versões da base aumentada. Com isso, buscamos identificar que elementos da seleção influem com maior peso na assertividade da detecção. Nesse ponto, pretendemos, dentre outros: buscar outras métricas de similaridade, variar a quantidade de *bins* presentes em cada um dos histogramas utilizados como descritores e variar o tamanho do *background* que circunda os objetos reais que compõem a base de objetos reais.

Por fim, pretendemos integrar o método de aumento de dados, definitivamente, ao fluxo de treinamento do sistema de detecção.

Referências Bibliográficas

- [1] ADJUTO, D., “Mortes por dengue aumentam 50%; casos prováveis sobem 61,5%”, <https://www.cnnbrasil.com.br/saude/2020/03/30/mortes-por-dengue-aumentam-50-casos-provaveis-sobem-61-5>, 2020, (Acesso em 13 de Agosto de 2020).
- [2] BOGAZ, C., “Cresce em 264% o número de casos de dengue no país”, <https://www.saude.gov.br/noticias/agencia-saude/45314-cresce-em-264-o-numero-de-casos-de-dengue-no-pais>, 2019, (Acesso em 13 de Agosto de 2020).
- [3] MINISTÉRIO DA SAÚDE, “Levantamento rápido de índices para *Aedes aegypti* para vigilância etimológica do *Aedes aegypti* no Brasil”, http://bvsmis.saude.gov.br/bvs/publicacoes/manual_liraa_2013.pdf, 2013, (Acesso em 27 de Julho de 2020).
- [4] VCTECHNOLOGY, “Litchi”, <https://flylitchi.com/>, 2018.
- [5] BARROS, B. D., *On Data Augmentation Techniques for the Automatic Detection of Mosquito Breeding Grounds*. M.Sc. dissertation, Universidade Federal do Rio de Janeiro, Fev. 2019.
- [6] LECUN, Y., BOSER, B., DENKER, J. S., *et al.*, “Backpropagation Applied to Handwritten Zip Code Recognition”, *Neural Computation*, v. 1, n. 4, pp. 541–551, 1989.
- [7] PASSOS, W. L., *Automatic Aedes Aegypti Breeding Grounds Detection Using Computer Vision Techniques*. M.Sc. dissertation, Universidade Federal do Rio de Janeiro, Fev. 2019.

- [8] TUN-LIN, W., LENHART, A., NAM, V. S., *et al.*, “Reducing costs and operational constraints of dengue vector control by targeting productive breeding places: a multi-country non-inferiority cluster randomized trial”, *Tropical Medicine & International Health*, v. 14, n. 9, pp. 1143–1153, 2009.
- [9] SHEAD, S., “Chinese residents worry about rise of facial recognition”, <https://www.bbc.com/news/technology-50674909>, 2018, (Acesso em 20 de Agosto de 2020).
- [10] LOPES, A., “FaceApp volta a viralizar em 2020 e reacende preocupação com a privacidade”, <https://veja.abril.com.br/tecnologia/faceapp-volta-a-viralizar-em-2020-e-reacende-preocupacao-com-a-privacidade/>, 2020, (Acesso em 15 de Agosto de 2020).
- [11] GOODFELLOW, I. J., BENGIO, Y., COURVILLE, A., *Deep Learning*. Cambridge, MA, USA, MIT Press, 2016.
- [12] KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E., “ImageNet Classification with Deep Convolutional Neural Networks”, *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 1097–1105, Dez. 2012.
- [13] LOWE, D. G., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision (IJCV)*, v. 60, n. 2, pp. 91–110, Nov. 2004.
- [14] BOSER, B. E., GUYON, I. M., VAPNIK, V. N., “A Training Algorithm for Optimal Margin Classifiers”, *Workshop on Computational Learning Theory (COLT)*, pp. 144–152, Jul. 1992.
- [15] AGARWAL, A., CHAUDHURI, S., CHAUDHURI, U., *et al.*, “Detection of potential mosquito breeding sites based on community sourced geotagged images”. In: *Proc. Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II*, v. 9089, p. 90890M, 2014.
- [16] ZHANG, Z., “A flexible new technique for camera calibration”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 11, pp. 1330–1334, Nov. 2000.

- [17] PEREZ, L., WANG, J., “The Effectiveness of Data Augmentation in Image Classification using Deep Learning”, *Computing Research Repository (CoRR)*, v. abs/1712.04621, 2017.
- [18] LEDIG, C., THEIS, L., HUSZAR, F., *et al.*, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”, *Computing Research Repository (CoRR)*, v. abs/1609.04802, 2016.
- [19] CHOU, Y., CHEN, C., LIU, K., *et al.*, “Changing Background to Foreground: An Augmentation Method Based on Conditional Generative Network for Stin-gray Detection”. In: *25th IEEE International Conference on Image Processing (ICIP)*, pp. 2740–2744, 2018.
- [20] DWIBEDI, D., MISRA, I., HEBERT, M., “Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection”, *Computing Research Repository (CoRR)*, v. abs/1708.01642, 2017.
- [21] DVORNIK, N., MAIRAL, J., SCHMID, C., “On the Importance of Visual Context for Data Augmentation in Scene Understanding”, *Computing Research Repository (CoRR)*, v. abs/1809.02492, 2018.
- [22] FOROOSH, H., ZERUBIA, J. B., BERTHOD, M., “Extension of phase correlation to subpixel registration”, *IEEE Transactions on Image Processing*, v. 11, n. 3, pp. 188–200, 2002.
- [23] REDMON, J., DIVVALA, S. K., GIRSHICK, R. B., *et al.*, “You Only Look Once: Unified, Real-Time Object Detection”, *Computing Research Repository (CoRR)*, v. abs/1506.02640, 2015.
- [24] INTERNATIONAL ELECTROTECHNICAL COMMISSION, “Multimedia Systems and equipment - Colour measurement and management”, Padrão Internacional, 1999.
- [25] RUIZ, D. V., KRINSKI, B. A., TODT, E., “ANDA: A Novel Data Augmentation Technique Applied to Salient Object Detection”, *Computing Research Repository (CoRR)*, v. abs/1910.01256, 2019.

- [26] OJALA, T., PIETIKAINEN, M., MAENPAA, T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, pp. 971–987, 2002.