

VISUALIZAÇÃO PANORÂMICA GEORREFERENCIADA DE
FOCOS DE MOSQUITO *Aedes Aegypti*

Lucas Silva Meyer

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.

Orientador: Eduardo Antonio Barros da Silva

Coorientador: Wesley Lobato Passos

Rio de Janeiro

Fevereiro de 2025

VISUALIZAÇÃO PANORÂMICA GEORREFERENCIADA DE
FOCOS DE MOSQUITO *Aedes Aegypti*

Lucas Silva Meyer

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

Autor:



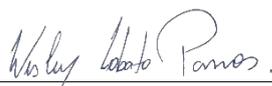
Lucas Silva Meyer

Orientador:



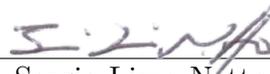
Prof. Eduardo Antônio Barros da Silva, Ph. D.

Coorientador:



Prof. Wesley Lobato Passos, D. Sc.

Examinador:



Prof. Sergio Lima Netto, Ph. D.

Examinador:



Thadeu Luiz Barbosa Dias, M. Sc.

Rio de Janeiro

Fevereiro de 2025

Declaração de Autoria e de Direitos

Eu, *Lucas Silva Meyer* CPF 176.505.487-75, autor da monografia *Aprendizado de Máquina e Processamento Digital de Imagens Aplicados à Detecção Automática de Focos de Reprodução do Mosquito 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. Excetuam-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.



Lucas Silva Meyer

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

Dedico este trabalho ao meu filho, João Vitor, a quem amo de forma que jamais imaginei ser possível.

AGRADECIMENTO

Agradeço à minha mãe, por todo o sacrifício em prol da nossa família, por sempre me motivar a seguir em frente, mesmo diante das dificuldades, e por sempre incentivar minha dedicação aos estudos. Ao meu pai, sou grato por ter plantado em mim o desejo de estudar nesta instituição de excelência e por, tantas vezes, me ajudar a superar desafios que eu acreditava serem impossíveis. Agradeço também aos meus irmãos, que sempre demonstraram, com orgulho, seu apoio a mim.

RESUMO

O mosquito *Aedes aegypti* é transmissor de um conjunto de doenças graves - dengue, zika, chikungunya e febre amarela - que, anualmente, afetam milhões de pessoas. Evitar a reprodução desse mosquito ainda é considerada a forma mais eficiente de reduzir o número de infecções por essas doenças, tarefa dispendiosa e que demanda um contingente significativo de pessoas. Para auxiliar na eficiência dessa tarefa, trabalhos predecessores propõem sistemas de detecção automática dos possíveis criadouros do mosquito utilizando visão computacional e aprendizado de máquina em vídeos capturados por veículos aéreos não tripulados. Neste contexto, o presente projeto busca complementar o trabalho de detecção com um sistema capaz de apresentar as áreas monitoradas de forma fiel, com alta qualidade e destacando os locais contendo os objetos de interesse detectados. Para atender aos objetivos do projeto, foi realizada a costura das imagens capturadas de uma determinada região, apresentando os potenciais focos de reprodução do mosquito através de caixas delimitadoras. Também foi desenvolvido um sistema de georreferenciamento dos objetos detectados nos vídeos. Com o georreferenciamento dos criadouros, foi gerado um mapa capaz de apresentar pinos com links de redirecionamento para rotas no Google Maps que levem até o objeto detectado e um mapa de calor para fácil identificação das regiões mais críticas em relação à quantidade de potenciais focos presentes.

Palavras-Chave: visão computacional, costura de imagens, georreferenciamento, *Aedes aegypti*.

ABSTRACT

The *Aedes aegypti* mosquito is a transmitter of several severe diseases – Dengue, Zika, Chikungunya, and Yellow Fever – which annually affect millions of people. Preventing the reproduction of this mosquito is still considered the most efficient way to reduce the number of infections caused by these diseases, a labor-intensive task that demands a significant workforce. To aid in the efficiency of this task, previous works propose automatic detection systems for possible mosquito breeding sites using computer vision and machine learning in videos captured by unmanned aerial vehicles. In this context, the present project aims to complement the detection work with a system capable of presenting the monitored areas accurately, with high quality, and highlighting the locations containing the detected objects of interest. To achieve the project’s objectives, image stitching of a specific region was performed, presenting potential mosquito breeding sites through bounding boxes. Additionally, a geo-referencing system for the objects detected in the videos was developed. With the geo-referencing of breeding sites, a map was generated capable of presenting pins with links redirecting to routes in Google Maps leading to the detected objects, along with a heatmap for easy identification of the most critical areas in relation to the number of potential breeding sites present.

Key-words: computer vision, image stitching, geo-referencing, *Aedes aegypti*.

Sumário

1	Introdução	1
2	Fundamentação teórica	4
2.1	Identificação de focos de mosquito utilizando machine learning	5
2.2	Costura de imagens	6
2.3	<i>Feature Detection e Matching</i>	7
2.4	<i>Homografia</i>	9
2.5	<i>Structure from Motion</i>	10
2.5.1	O COLMAP como ferramenta de <i>Structure from Motion</i>	12
2.6	RANSAC	13
2.7	Conclusão	14
3	Metodologia e validação de resultados	15
3.1	Introdução	15
3.2	Base de dados	16
3.2.1	Drone	16
3.2.2	Google Earth	17
3.2.3	Litchi for DJI Drones	17
3.2.4	Descrição da base MBG_V2	18
3.3	Costura de imagens	21
3.3.1	Correlação de fase	21
3.3.2	COLMAP	30
3.4	Caixas delimitadoras	42
3.5	Georreferenciamento de objetos e mapa de calor	44
3.6	Conclusão	50

4 Conclusão e trabalhos futuros	51
Bibliografia	53

Lista de Figuras

2.1	Exemplo de costura entre duas imagens.	6
2.2	Imagem com pontos vermelhos representando <i>features</i> detectados em uma imagem.	8
2.3	Costa de duas imagens utilizando a matriz de homografia.	10
2.4	Modelo de reta estimado através método RANSAC. Fonte: [1]	14
3.1	Imagem do quadricóptero DJI Phantom 4 Pro. Fonte: [2].	16
3.2	Ilustração do ângulo de câmera. Fonte: [3]	17
3.3	Imagens contidas nos vídeos da base de dados MBG_V2, com e sem representação dos objetos anotados.	19
3.4	Costura de aparência bem comportada.	23
3.5	Suavização de área sobreposta através de janela cossenoidal.	24
3.6	Comparação do panorama com janela fixa e janela variável.	26
3.7	Gráfico comparativo de posição discreta e posição real dos quadros do panorama.	27
3.8	Exemplos de costuras com quantidade diferentes de salto entre quadros na computação da correlação de fase.	29
3.9	Modelo 3D gerado pelo COLMAP com apresentação das poses de câmera em vermelho para os quadros usados no vídeo retificado e apresentação dos pontos esparsos.	31
3.10	Modelo 3D gerado pelo COLMAP com apresentação das poses de câmera em vermelho para os quadros usados do vídeo não retificado e apresentação dos pontos esparsos.	32
3.11	Reconstruções densas do COLMAP.	34
3.12	Homografia entre par de imagens através do método RANSAC	35
3.13	Quatro quadros costurados pela técnica de homografias relativas.	37

3.14	Costura distorcida.	37
3.15	Perspectivas da nova câmera para computação das homografias diretas.	38
3.16	Projeção 2D dos pontos 3D do modelo esparsa.	40
3.17	Homografia dos pontos projetados considerando uma única pose de câmera.	41
3.18	Caixas delimitadoras anotadas na base de dados apresentadas em um quadro retificado e em um quadro não retificado.	42
3.19	Apresentação de novas posições das caixas delimitadoras conside- rando os métodos de remapeamento de pontos entre o quadro re- tificado e não retificado propostos.	43
3.20	Imagem ampliada de um trecho da costura apresentando as <i>bouding</i> <i>boxes</i> de caixas d'água anotadas na base de dados.	44
3.21	Direção dos versores de latitude e longitude em um quadro.	47
3.22	Sobreposição de um quadro e apresentação das coordenadas dos vértices do quadro.	48
3.23	Mapa georreferenciado.	49

Lista de Tabelas

3.1	Detalhes das regiões de gravação de vídeo.	20
-----	--	----

Capítulo 1

Introdução

O mosquito *Aedes aegypti* é transmissor de arboviroses como dengue, zika, chikungunya e febre amarela [4]. Em 2020, foram reportados mais de 2.3 milhões de casos de dengue somente nas Américas, a maioria deles no Brasil [5]. Em 2023, o Brasil liderou o número de casos de dengue no mundo de acordo com a OMS, registrando 2,9 milhões de infecções pelo vírus, mais da metade do valor registrado mundialmente, que foi de 5 milhões [6]. Já no ano de 2024, somente no Brasil, foram registrados 6,5 milhões de casos prováveis de dengue até o período de 7 de outubro de 2024, de acordo com o Painel de Monitoramento das Arboviroses do Ministério da Saúde, quantidade 400% maior que no ano de 2023 [7]. Além do aumento de infecções registrado ao longo dos anos, as infecções e a presença do mosquito tem se espalhado para locais sem histórico de casos anteriores [7, 8]. A dengue, zika, chikungunya e febre amarela são doenças com potencial de danos graves aos infectados. No período de 2024, pelo menos 5.536 pessoas morreram em decorrência da dengue [7]. O nascimento de bebês com microcefalia tem associação com a infecção pelo vírus zika no primeiro trimestre da gestação [9], condição que pode prejudicar significativamente no desenvolvimento infantil.

Além do impacto de saúde, as doenças transmitidas pelo *Aedes aegypti* geram impactos econômicos significativos. Uma pesquisa conduzida em dezessete países da América do Sul e Central, no ano de 2018, estimou que o custo de epidemias de dengue nesses países excedem três bilhões de dólares anualmente, 1,4 bilhão somente no Brasil [10]. O Ministério da Saúde investiu mais de 1,5 bilhão de reais para o

período sazonal de 2024/2025, o que representa um aumento de 50% em relação ao período sazonal de 2023/2024 [11].

Com o passar dos anos, diferentes estratégias para a redução das infecções por arboviroses vem sendo empregadas, mas, de acordo com o Ministério da Saúde, a forma mais eficaz de preveni-las é através da eliminação dos focos do mosquito *Aedes aegypti* [11].

Esses fatos evidenciam as arboviroses transmitidas pelo *Aedes aegypti* como um problema grave e de caráter global. No Brasil há uma campanha de combate à reprodução do mosquito *Aedes aegypti* através da busca de potenciais focos. Os focos do mosquito são ambientes de água limpa, parada e com sombra: caixas d'água, pneus, garrafas, etc. Para identificar a presença desses objetos, os agentes de saúde fazem visitas residenciais em busca desses objetos. Neste sentido, o presente projeto pode colaborar na identificação da presença de focos do mosquito, direcionando os agentes de saúde para locais onde o sistema já identificou a presença de objetos propícios para a reprodução do *Aedes aegypti*. A proposta tem grande potencial de melhorar a eficiência no combate a reprodução do mosquito e, também, reduzir custo operacional causado pelo deslocamento dos agentes de saúde à locais de ausência dos potenciais focos.

A integração de tecnologias de sensoriamento remoto, processamento de imagens e aprendizado de máquina tem revolucionado a análise espacial e ambiental. A proposta do trabalho é utilizar essas tecnologias para desenvolver um sistema capaz de apresentar mapas indicando focos detectados por sistemas automáticos. Para isso, o projeto utiliza um conjunto de dados contendo vídeos capturados por veículos aéreos não-tripulados (VANTs) durante sobrevoos programados sobre uma região específica. Além dos vídeos, são gerados arquivos de registro de voo no formato CSV, que documentam informações do voo como latitude, longitude e altitude. Esses dados permitem a localização precisa do VANT em cada quadro do vídeo.

A metodologia envolve a realização de uma costura de imagens – um processo que combina múltiplos quadros com campos de visão sobrepostos para criar uma

imagem panorâmica de alta resolução. O objetivo do uso dessa técnica é gerar uma representação única e contínua da área monitorada, proporcionando uma visualização detalhada da região, viabilizando maior acurácia no encontro dos focos pelos agentes de saúde. Através da combinação do resultado obtido com a costura com informações contidas nos arquivos de registro de voo, também é realizado o georreferenciamento dos objetos para produção de um Mapa de Calor, permitindo rápida identificação de áreas críticas pela presença de cores mais intensas no mapa, bem como *pins* com links de redirecionamento para rotas no Google Maps, facilitando a chegada em locais com criadouros detectados pelo sistema.

Este trabalho está organizado da seguinte forma: O Capítulo 2 apresenta uma técnica de detecção de focos de mosquito com aprendizado de máquina e alguns conceitos importantes que foram utilizados no projeto. Os detalhes da solução proposta, metodologia, resultados e discussões sobre os resultados estão presentes no Capítulo 3. Finalmente, o Capítulo 4 fala sobre a contribuição do projeto nas áreas de visão computacional, processamento de imagens, na área da saúde como auxiliar na redução da proliferação do mosquito *Aedes aegypti* e discute direções futuras para esse trabalho.

Capítulo 2

Fundamentação teórica

Neste capítulo, serão apresentados uma técnica de detecção de focos de mosquito com aprendizado de máquina e conceitos que foram utilizados ao longo do desenvolvimento do projeto, como *Image Stitching* (Costura de Imagens), *Structure from Motion* [12], o *software opensource* (código aberto) COLMAP [13] e o método RANSAC [14].

O capítulo está organizado da seguinte forma: a Seção 2.1 apresenta um trabalho para a identificação de focos de mosquitos utilizando aprendizado de máquina, motivador para o desenvolvimento do presente trabalho. A Seção 2.2 introduz o conceito de costura de imagens e comenta quais métodos foram escolhidos para implementar a costura no projeto. A Seção 2.3 comenta sobre o processo de identificação de pontos de interesse em uma imagem e o de casamento entre pontos de interesse correspondente entre duas imagens, comentando sobre os dois métodos que o projeto utiliza, SIFT e ORB. A Seção 2.4 apresenta o conceito de homografia de imagens e faz uma descrição geral do método de computação da homografia entre imagens que o projeto utiliza. A Seção 2.5 introduz a técnica de *Structure from Motion*, detalhando sua aplicação na reconstrução tridimensional de cenas e apresenta o COLMAP como uma ferramenta versátil para implementação de tal técnica, descrevendo seu fluxo de trabalho e os principais algoritmos utilizados, detalhando como os resultados dessa técnica podem ser úteis no projeto para a realização de homografia entre imagens. A Seção 2.6 fala sobre a técnica *Random Sample Consensus* e sua utilidade na estimação da homografia entre quadros. Por fim, a Seção 2.7

conclui pontuando que as metodologias e conceitos apresentados nesse capítulo são a base do sistema desenvolvido.

2.1 Identificação de focos de mosquito utilizando machine learning

A aplicação de machine learning na identificação de focos de mosquito tem ganhado relevância nos últimos anos [15, 16, 17, 18, 19]. Modelos de aprendizado de máquina são treinados para reconhecer padrões em imagens capturadas por VANTs, permitindo a detecção automática de áreas propensas à proliferação de mosquitos.

Uma abordagem para a identificação de criadouros de mosquito é apresentada em [16]. Este trabalho propõe um sistema automático que utiliza vídeos aéreos capturados por VANTs para detectar potenciais criadouros do *Aedes aegypti*. Combinando técnicas de visão computacional e aprendizado de máquina, o sistema identifica objetos como caixas d'água e pneus, que são locais críticos para a reprodução do mosquito.

A metodologia desenvolvida emprega a arquitetura Faster R-CNN [20] para detecção de objetos em cada quadro do vídeo, seguida por um módulo de consistência espaço-temporal que correlaciona as detecções entre quadros sucessivos. Essa abordagem melhora a robustez e precisão do sistema, resultando em um desempenho superior às técnicas tradicionais baseadas em mosaicos de quadros.

Os resultados experimentais demonstraram a eficácia do sistema, alcançando F1-scores de 0,65 para a classe pneu e 0,77 para a classe caixa d'água. Além disso, foi introduzida uma nova métrica denominada Spatiotemporal Tube Average Precision (STT-AP), projetada para avaliar o desempenho de detectores de objetos em sequências de vídeo. A abordagem também mostrou que técnicas de aumento de dados podem reduzir o sobreajuste e melhorar o desempenho do modelo, com um incremento de 5,3% no desempenho da Faster R-CNN.

O estudo demonstra como modelos de *deep learning* podem ser treinados com imagens aéreas para reconhecer padrões associados à presença de objetos considerados como potenciais focos de proliferação do mosquito *Aedes aegypti* e motiva o desenvolvimento de sistemas complementares para apresentação dos objetos detectados de forma amigável ao ser humano. As seções seguintes abordam diversos conceitos utilizados no desenvolvimento de tal sistema.

2.2 Costura de imagens

A técnica de *Image Stitching*, ou Costura de Imagens, é um processo computacional que combina múltiplas imagens sobrepostas de uma cena para criar uma única imagem panorâmica ou composta com um campo de visão ampliado. Essa técnica é amplamente utilizada em aplicações como reconstrução de ambientes, visão computacional, monitoramento aéreo, reconstrução de exames de imagem [21] e fotografia panorâmica. Para o sistema desenvolvido nesse trabalho, o *Image Stitching* é utilizado com o intuito de proporcionar uma visualização atualizada e de alta resolução da região monitorada pelo sistema de detecção. A Figura 2.1 apresenta um exemplo de duas imagens costuradas.

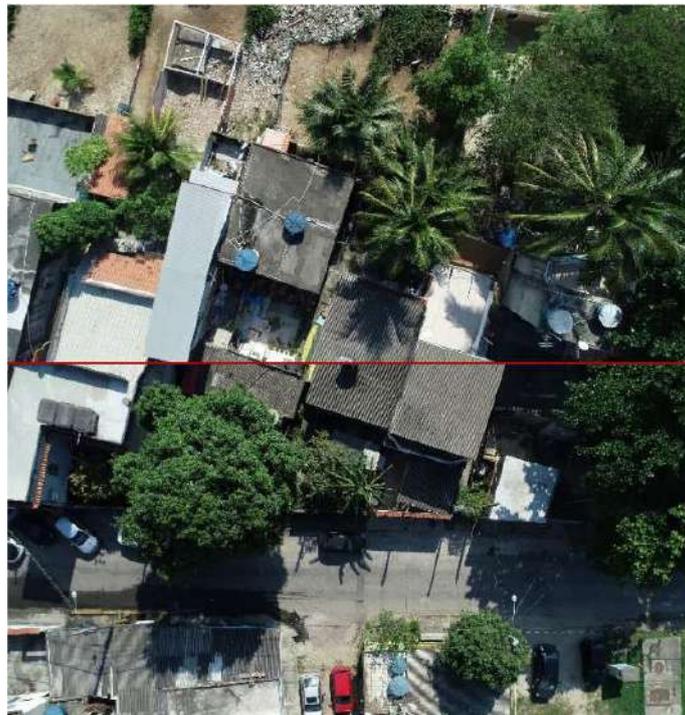


Figura 2.1: Exemplo de costura entre duas imagens.

Há diversas metodologias propostas na literatura para a realização de *Image Stitching* [22, 23, 24, 25, 21, 26, 27]. O *Image Stitching* por correlação de fase [22, 23, 24, 25] é um método útil quando há imagens consecutivas com sobreposição significativa e deseja-se determinar o deslocamento entre elas, porém não consegue lidar com rotações e mudanças de escala. Como a base de dados utilizada no projeto (Seção 3.2.4) foi construída com o intuito de minimizar rotações e diferenças de escala ao longo da captura do vídeo, a costura de imagens por correlação de fase foi uma das opções adotadas ao longo do desenvolvimento do projeto. Outra opção foi a de costurar as imagens através da realização de homografias utilizando-se dados provenientes do uso do software COLMAP, valendo-se das saídas provenientes do *Structure from Motion*, e utilizando o método RANSAC [14] na estimação da homografia entre as imagens. O método de costura através de homografia é mais robusto que o de correlação de fase, pois tem a capacidade de lidar com rotações e mudanças de escala, porém tem maior complexidade computacional.

2.3 *Feature Detection e Matching*

Antes de apresentar os conceitos de homografia e *Structure from Motion* comentados na Seção 2.2, os conceitos de *Feature Detection e Matching*, um conceito importante utilizado por ambas as técnicas, serão introduzidos. *Feature Detection* é um processo utilizado para identificar pontos de interesse em uma imagem a Figura 2.2 ilustra esses pontos em uma imagem. Esses pontos, frequentemente chamados de "features" ou "keypoints", são regiões que contêm informações distintas, como bordas, cantos ou texturas, que podem ser encontrados em diferentes vistas de uma mesma cena ou objeto, sendo altamente informativos para diversas aplicações, como reconstrução 3D e estimação da homografia entre imagens [28]. A etapa de *Matching* consiste no casamento de *features* entre duas imagens distintas, quando essa correspondência entre *features* é imprecisa, ou seja, dois pontos distintos são considerados como sendo o mesmo *feature* nas imagens, esse ponto é considerado discrepante ou "outlier", resultando em erro ao ser considerado para estimar a homografia entre as imagens, justificando o uso do RANSAC, apresentado na Seção 2.6, para mitigar a influência dos *outliers*



Figura 2.2: Imagem com pontos vermelhos representando *features* detectados em uma imagem.

Diversos algoritmos foram desenvolvidos para detectar características em imagens. O projeto faz uso de dois algoritmos diferentes, o SIFT (*Scale-Invariant Feature Transform*, Transformada de Características Invariável em Escala em português) [29], que é utilizado pelo COLMAP para identificação de pontos de interesse em imagens, e o ORB (*Oriented FAST and Rotated BRIEF*, FAST orientado e BRIEF rotacionado em português, onde FAST e BRIEF são as siglas para *Features from Accelerated Segment* e *Binary Robust Independent Elementary Features* respectivamente), onde FAST pode ser traduzido como Características de Segmento Acelerado e BRIEF como Características Binárias Robustas, Independentes e Elementares [30], utilizado para encontrar pontos correspondentes entre as imagens retificadas e não retificadas, no método de remapeamento de caixas delimitadoras por homografia, exemplificado na Seção 3.4.

O SIFT [29] é um clássico algoritmo para detecção e descrição de características. Ele opera em múltiplas escalas e utiliza uma abordagem baseada em diferenças gaussianas para identificar pontos de interesse. Os descritores do SIFT são altamente discriminativos, permitindo a correspondência precisa entre imagens sob variações de escala, rotação e iluminação. O ORB [30] combina os detectores FAST e os

descritores BRIEF. Em relação ao SIFT, o ORB é mais eficiente em termos de custo computacional [31], sendo amplamente utilizado em aplicações onde o desempenho computacional é crítico, como dispositivos móveis.

2.4 Homografia

A homografia é uma transformação geométrica fundamental em visão computacional e processamento de imagens, sendo amplamente utilizada para alinhar e retificar imagens capturadas de diferentes perspectivas. Essa transformação é representada por uma matriz 3×3 e descreve a relação entre dois planos projetivos, mapeando pontos de um plano em outro plano, ou seja, mapeia os pontos de uma imagem para outra sob uma transformação de perspectiva [32]. Uma homografia é formalmente definida pela seguinte equação:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = H \begin{bmatrix} x \\ y \\ w \end{bmatrix}, \quad (2.1)$$

onde:

- H é uma matriz 3×3 não singular,
- (x, y, w) e (x', y', w') são, respectivamente, as coordenadas homogêneas dos pontos antes e depois da transformação.

Dado um conjunto de pelo menos quatro correspondências de pontos entre duas imagens, a matriz H pode ser estimada [28]. A homografia possui diversas aplicações em visão computacional, como retificação de imagens e *Image Stitching*. O método utilizado no projeto envolve:

1. Detecção de *features* em ambas as imagens (SIFT ou ORB).
2. Correspondência de pontos entre as imagens.
3. Cálculo da matriz H utilizando o método RANSAC [14] para lidar com *outliers*.
4. Aplicação da transformação para mapear pontos.

A homografia é uma ferramenta poderosa no processamento de imagens e visão computacional, permitindo a manipulação da perspectiva e a fusão de imagens de maneira precisa. Sua correta aplicação depende da escolha adequada dos pontos de correspondência e do método de estimação, garantindo uma transformação robusta e eficiente para diferentes contextos de análise visual. A Figura 2.3 apresenta um exemplo duas imagens costuradas através da estimação de homografia entre elas.



Figura 2.3: Costa de duas imagens utilizando a matriz de homografia.

2.5 *Structure from Motion*

SfM (*Structure from Motion*, Estrutura a partir do Movimento, em português) é o processo de estimação da estrutura 3D de uma cena a partir de uma composição de imagens 2D [33].

Structure from Motion é uma técnica amplamente utilizada em fotogrametria e visão computacional para reconstruir modelos tridimensionais de objetos ou cenários a partir de imagens bidimensionais [33]. Essa abordagem aproveita a sobreposição

entre imagens capturadas de diferentes perspectivas, para estimar simultaneamente a posição da câmera (pose) e a estrutura tridimensional da cena [32]. O princípio fundamental do SfM baseia-se na triangulação, em que pontos correspondentes em múltiplas imagens 2D permitem determinar a localização espacial desses pontos no espaço 3D [28].

O SfM tem ampla aplicação em diversas áreas, como arqueologia, engenharia, mapeamento urbano e criação de ambientes digitais para jogos e filmes. Sua popularidade é impulsionada por sua capacidade de utilizar câmeras comuns e pela ausência da necessidade de hardware especializado [28]. A técnica é particularmente eficaz quando combinada com métodos de refinamento, como ajustes de feixes (bundle adjustment), uma técnica de otimização iterativa usada para refinar simultaneamente a estrutura tridimensional de uma cena e os parâmetros intrínsecos e extrínsecos da câmera [32], otimizando as estimativas de posições das câmeras e dos pontos 3D [34], diminuindo o erro de reprojeção e garantindo que os pontos 3D projetados nas imagens fiquem o mais próximo possível das suas correspondências detectadas. O erro de reprojeção é uma métrica usada em visão computacional e fotogrametria para avaliar a precisão da reconstrução tridimensional de pontos a partir de múltiplas imagens. Ele mede a discrepância entre a posição projetada de um ponto 3D na imagem, baseada nos parâmetros da câmera e da reconstrução, e a posição real do ponto detectado na imagem. Matematicamente, se um ponto tridimensional P é projetado em uma imagem como p' usando os parâmetros estimados da câmera (matriz de calibração, rotação e translação), o erro de reprojeção é a distância entre p' e a posição real p do ponto na imagem.

Como o SfM estima a posição da câmera e a estrutura tridimensional da cena, ele tem utilidade para analisar a trajetória do VANT durante a gravação, verificando possíveis variações de altitude e inclinações ou rotações da câmera ao longo do voo do quadricóptero.

2.5.1 O COLMAP como ferramenta de *Structure from Motion*

O COLMAP é um software de fotogrametria de código aberto amplamente utilizado para reconstrução 3D a partir de imagens ou vídeos [13]. Ele utiliza algoritmos avançados de SfM [12] e ajuste de múltiplas vistas (*Multi-View Stereo*, MVS) para gerar modelos tridimensionais precisos e texturizados [35]. Sua interface intuitiva e o suporte a configurações personalizáveis tornam-o uma ferramenta versátil para aplicações acadêmicas, industriais e de pesquisa em áreas como mapeamento, arqueologia, robótica e realidade virtual. O COLMAP é reconhecido por sua eficiência, precisão e compatibilidade com diferentes formatos de dados [35], sendo uma escolha popular entre profissionais e pesquisadores para projetos de reconstrução espacial e análise geométrica.

Dessa forma, o processo de reconstrução 3D no COLMAP envolve uma série de etapas interdependentes, que vão desde a identificação de pontos de interesse nas imagens até a geração de modelos tridimensionais detalhados. A seguir, são apresentadas as principais etapas desse fluxo de trabalho, destacando os algoritmos e técnicas empregados em cada uma delas.

1. **Extração e correspondência de características:** O COLMAP começa identificando pontos de interesse em cada imagem, utilizando algoritmos como SIFT. Em seguida, realiza a correspondência desses pontos entre as imagens sobrepostas, estabelecendo relações espaciais iniciais [35].
2. **Reconstrução inicial (SfM incremental ou hierárquico):** O software utiliza dois modos principais para a reconstrução esparsa do conjunto de imagens:
 - **Incremental:** Realiza a reconstrução 3D esparsa a partir de um pequeno conjunto inicial de imagens, expandindo progressivamente a estrutura ao adicionar novas imagens com base em correspondências de características.
 - **Hierárquico:** Divide a cena em submodelos sobrepostos, reconstruindo cada um de forma independente e, posteriormente, mesclando-os em uma

única reconstrução. Esse método paraleliza o processo de SfM e é recomendado para grandes conjuntos de dados, seguido de etapas adicionais de triangulação de pontos e otimização por *Bundle Adjustment* [12].

3. **Ajuste de blocos (Bundle Adjustment):** Após determinar as poses das câmeras e os pontos 3D, o COLMAP aplica otimizações iterativas para minimizar os erros de reprojeção e refinar a qualidade da reconstrução [35].
4. **Reconstrução densa:** O COLMAP pode criar modelos tridimensionais densos ao estimar a profundidade em cada pixel das imagens, usando técnicas avançadas de correspondência estéreo [12].
5. **Exportação e visualização:** Os resultados podem ser exportados em formatos compatíveis com outras ferramentas de modelagem 3D ou utilizados diretamente para análise e visualização [13].

O COLMAP oferece, como opção de saída, arquivos contendo dados sobre a pose das câmeras utilizadas na reconstrução, incluindo informações detalhadas sobre suas posições e orientações, a correspondência entre pontos 3D da reconstrução e pixels de um determinado quadro, além de associar, para cada imagem, quais pixels dessa imagem correspondem e quais pontos 3D da reconstrução. Com essas informações, é possível relacionar pixels em comum entre os quadros da reconstrução e utilizá-los para encontrar homografias entre as imagens, facilitando a construção do panorama da cena.

2.6 RANSAC

O RANSAC (*Random Sample Consensus*, ou Consenso de Amostra Aleatória, em português) é um método iterativo utilizado para ajustar modelos a conjuntos de dados, reduzindo a influência de valores discrepantes [14]. O método funciona selecionando aleatoriamente um pequeno subconjunto de pontos para estimar um modelo candidato. Em seguida, verifica-se quantos pontos do conjunto total são consistentes com o modelo candidato dentro de uma certa margem de erro. O processo é repetido diversas vezes, com a seleção de diferentes subconjuntos aleatórios, e o modelo que

melhor se ajusta à maior quantidade de pontos consistentes é considerado o modelo final. A Figura 2.4 apresenta um modelo de reta estimado com o método RANSAC diferenciando *inliers* em azul de *outliers* em vermelho.

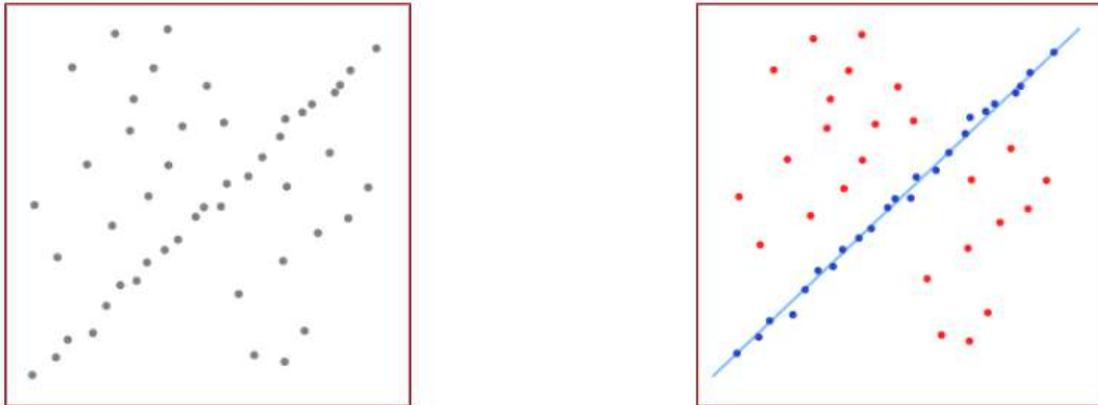


Figura 2.4: Modelo de reta estimado através método RANSAC. Fonte: [1]

Esse método é útil para estimar a homografia entre quadros que mantém a maior quantidade de pontos em comum aos quadros dentro de uma determinada margem de erro.

2.7 Conclusão

Este capítulo introduziu a existência de um sistema automático de detecção de focos de mosquito utilizando aprendizado de máquina, cujas contribuições motivaram o desenvolvimento de um sistema de visualização dos focos complementar ao de detecção. Além disso, ao longo do capítulo foram abordados os principais conceitos técnicos utilizados ao longo do trabalho, explicando suas utilidades e motivações para o uso. Através do uso desses conceitos técnicos, busca-se desenvolver um sistema capaz de utilizar as detecções de criadouros de mosquito em vídeos capturados por VANTs para gerar um mapa que auxilie os agentes de saúde no combate às arboviroses.

Capítulo 3

Metodologia e validação de resultados

3.1 Introdução

Nesse capítulo é feita uma descrição detalhada da solução proposta, a base de dados utilizada, implementação do sistema e validação dos resultados obtidos ao longo do desenvolvimento do projeto.

A Seção 3.2 apresenta a base de dados utilizada, informações sobre os equipamentos e ferramentas de software utilizados nesse processo, descreve a organização e estruturação dos arquivos obtidos pelo VANT e o método utilizado para a anotação dos objetos considerados como potenciais focos de reprodução do mosquito *Aedes aegypti*. A Seção 3.3 descreve o desenvolvimento do *Image Stitching* (Costura de Imagens), uma técnica que consiste em criar imagens com aparência natural composta da junção de múltiplas imagens de uma determinada cena com o objetivo de ampliar a visualização da cena mantendo um bom detalhamento [27, 26]. A Seção 3.4 descreve o método utilizado para destacar os potenciais focos de mosquito presentes na costura, comenta sobre os desafios encontrados nesse processo e os métodos propostos para superar esses desafios, bem como apresenta os resultados obtidos pelos métodos. A Seção 3.5 descreve o desenvolvimento do sistema de georreferenciamento dos objetos anotados e apresentação desses objetos em um mapa de calor. A Seção 3.6 finaliza o capítulo comentando sobre como a metodologia apre-

sentada ao longo desse capítulo pode colaborar com políticas públicas de combate a proliferação de mosquitos transmissores de arboviroses.

3.2 Base de dados

Essa seção apresenta informações sobre a base de dados MBG_V2 (Mosquito Breeding Grounds, Criadouros de Mosquito Versão 2 em português) [15], utilizada no projeto, resume como ela foi construída e a descreve. Dados técnicos sobre o VANT utilizado para a construção da base de dados, comenta sobre as ferramentas de software utilizadas para a geração do plano de voo do VANT e descreve a organização e estruturação dos arquivos que compõem a base de dados MBG_V2.



Figura 3.1: Imagem do quadricóptero DJI Phantom 4 Pro. Fonte: [2].

3.2.1 Drone

O DJI Phantom 4 Pro, mostrado na Figura 3.1, possui um sensor CMOS de 1 polegada, grava vídeos em 4K de até 60 fps e registra fotos de 20 megapixels com nitidez e clareza. A câmera também é equipada com um obturador mecânico, foco automático e a aeronave conta com um estabilizador de 3 eixos, proporcionando imagens e vídeos nítidos e estáveis, que pode inclinar a câmera dentro de uma faixa de 120 graus conforme demonstrado na Figura 3.2.



Figura 3.2: Ilustração do ângulo de câmera. Fonte: [3]

3.2.2 Google Earth

O Google Earth é uma ferramenta de mapeamento e visualização geoespacial desenvolvida pela Google. Ele permite aos usuários explorar virtualmente o planeta, de maneira detalhada e interativa, através de imagens de satélite, imagens aéreas e modelos tridimensionais de terrenos e edifícios.

O Google Earth tem utilização em diferentes contextos, desde a exploração e educação geográfica até o planejamento urbano e pesquisa científica. Nesse projeto, essa ferramenta foi utilizada no contexto de pesquisa científica com o intuito de contribuir na criação do banco de dados de vídeos.

Com essa ferramenta, é possível gerar um arquivo KML (*Keyhole Markup Language*. Keyhole refere-se à empresa *Keyhole, Inc.*, portanto a tradução usual para o português é Linguagem de Marcação *Keyhole*), contendo marcadores, rotas e áreas delimitadas. A partir desse arquivo KML, arquivos CSV de caminho serão gerados para serem exportados ao Litchi.

3.2.3 Litchi for DJI Drones

O Litchi for DJI Drones é um aplicativo desenvolvido pela VC Technology LTD., projetado para oferecer uma experiência de voo personalizável e repleta de recursos para drones da DJI.

O aplicativo oferece uma ampla gama de recursos que aprimoram a experiência de voo, como: planejamento de missões de voo, modos de voo automatizados, controle de câmera e funcionalidades específicas para fotografia.

Para o projeto, foi utilizado o *Mission Hub* do Litchi para geração do caminho de voo do VANT. A geração do caminho dentro do Mission Hub é feita através da importação do arquivo CSV que foi gerado através do KML importado do Google Earth.

3.2.4 Descrição da base MBG_V2

Realizando um processamento nos vídeos obtidos para correção de distorções causadas pela lente da câmera do VANT [16] e, posteriormente, anotando os potenciais focos de reprodução do mosquito a base de dados MBG_V2 foi construída. Uma breve descrição desse processo é dada a seguir, maior detalhamento do processo pode ser consultado na Referência [16].

- **Vídeos:** Os arquivos de vídeo capturados pelo VANT durante o sobrevoo das regiões são armazenados com resolução de 4096 x 2160 p, assegurando alta qualidade e nitidez às imagens registradas.
- **Registro de voo:** além dos vídeos, o VANT disponibiliza *flightlogs* (registros de voo) contendo informações adicionais sobre o VANT durante o voo, como coordenadas GPS, altitude, ângulo de câmera e diversos outros parâmetros de voo. Essas informações são salvas em arquivos CSV.
- **Pré-processamento** Os vídeos são retificados para correção de distorções da lente, utilizando o algoritmo de Zhang [36].
- **Anotação dos objetos:** A base de dados considera 12 potenciais criadouros (garrafa PET, balde, piscina, poça, pneu, caixa d'água, caçamba de lixo, lixeira grande, lixeira pequena, saco plástico, vaso de planta e bueiro) para anotação. Cada anotação inclui informações como a classe do objeto, posição e o quadro correspondente. As anotações foram feitas usando CVAT (*Computer Vision Annotation Tool*, Ferramenta de Anotação de Visão Computacional em português) e salvos em formato XML fornecido pela própria ferramenta.

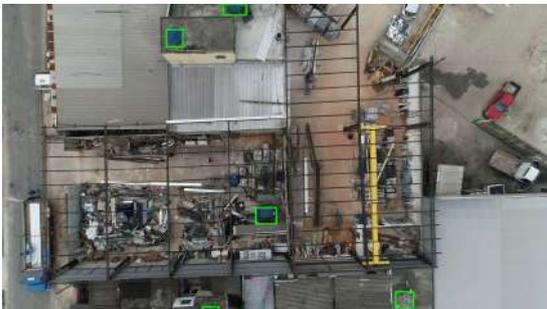
A Tabela 3.1, reproduzida de [16], apresenta detalhes sobre os vídeos disponíveis na base de dados MBG_V2. Essa base de dados tem as informações necessárias para a construção de um sistema de visualização de criadouros de mosquito em uma região.



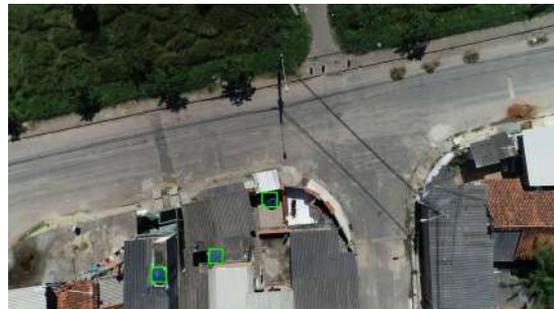
(a) Vizinhança de Tubiacanga.



(b) Vizinhança de Paciência.



(c) Vizinhança de Santa Cruz.



(d) Vizinhança de Sepetiba.

Figura 3.3: Imagens contidas nos vídeos da base de dados MBG_V2, com e sem representação dos objetos anotados.

Tabela 3.1: Detalhes das regiões de gravação de vídeo.

Região	Vizinhança	Área (m ²)	Data de Gravação (dia-mês-ano)	Vídeo #	Duração (min:seg)
A	Tubiacanga	239,317	01-06-2019	01	5:27
				02	5:27
B	Paciência	97,483	28-10-2020	03	3:44
				04	3:29
C	Paciência	113,623	25-02-2021	05	4:18
				06	4:02
D	Paciência	90,199	03-02-2021	07	4:07
				08	3:54
E	Paciência	55,823	03-02-2021	09	2:39
				10	2:28
F	Santa Cruz	81,596	25-02-2021	11	4:25
				12	2:19
G	Santa Cruz	77,431	25-02-2021	13	3:31
				14	3:43
H	Santa Cruz	107,751	03-02-2021	15	4:43
				16	4:29
I	Santa Cruz	100,844	17-03-2021	17	5:37
				18	5:25
J	Sepetiba	75,038	17-03-2021	19	3:08
				20	3:22
K	Sepetiba	53,269	27-01-2021	21	2:35
				22	2:48
L	Sepetiba	131,356	27-01-2021	23	6:03
				24	5:49
M	Sepetiba	149,687	27-01-2021	25	6:15
				26	6:01

3.3 Costura de imagens

Para realização do stitching, primeiro foi proposto o uso da correlação de fase para calcular o deslocamento entre quadros e realizar o alinhamento entre eles [23], com o objetivo de gerar um mosaico com cada imagem transladada. O segundo método proposto faz uso do COLMAP [13], para um processo de costura por homografias.

3.3.1 Correlação de fase

A correlação de fase é um método amplamente utilizado em aplicações de processamento de imagens para determinar o deslocamento entre duas imagens. Este método baseia-se na transformação para o domínio da frequência, onde o deslocamento espacial entre imagens é identificado como uma translação nas fases das suas transformadas de Fourier. Introduzida inicialmente por Kuglin [23], a técnica apresenta vantagens notáveis, como a robustez contra variações de iluminação e a alta precisão para deslocamentos pequenos, o que a torna uma ferramenta útil para o alinhamento de imagens, incluindo sua aplicação na costura de panoramas.

3.3.1.1 Algoritmo Proposto

O algoritmo desenvolvido para a costura de imagens visa processar vídeos gravados por VANTs, gerando panoramas contínuos que representem fielmente a área sobrevoada. Este processo é essencial em aplicações que requerem a visualização ampla e integrada de uma região, como mapeamento de terrenos e monitoramento ambiental.

O método de costura através da correlação de fase tem como princípio o cálculo do deslocamento entre pares de quadros consecutivos extraídos dos vídeos - a escolha de quadros consecutivos maximiza a sobreposição entre as imagens e minimiza as distorções causadas por variações de perspectiva, aumentando a confiança nos deslocamentos calculados. Para encontrar o deslocamento entre dois quadros não consecutivos A e D , utiliza-se a transitividade dos deslocamentos intermediários, acumulando suas contribuições. Por exemplo, se os quadros intermediários entre A e D forem $A \rightarrow B \rightarrow C \rightarrow D$, a correlação de fase cumulativa é dada por

$$Ph_c(A \rightarrow D) = Ph_c(A \rightarrow B) + Ph_c(B \rightarrow C) + Ph_c(C \rightarrow D), \quad (3.1)$$

onde:

- $Ph_c(A \rightarrow D)$: Correlação de fase entre os quadros A e D .
- $Ph_c(A \rightarrow B)$: Correlação de fase entre os quadros A e B .
- $Ph_c(B \rightarrow C)$: Correlação de fase entre os quadros B e C .
- $Ph_c(C \rightarrow D)$: Correlação de fase entre os quadros C e D .

O processo de costura de imagens desenvolvido envolve diversas etapas. Essas etapas incluem desde a extração dos quadros dos vídeos até a geração do panorama final, envolvendo a determinação dos deslocamentos e a aplicação de métodos para garantir a transição suave entre as imagens.

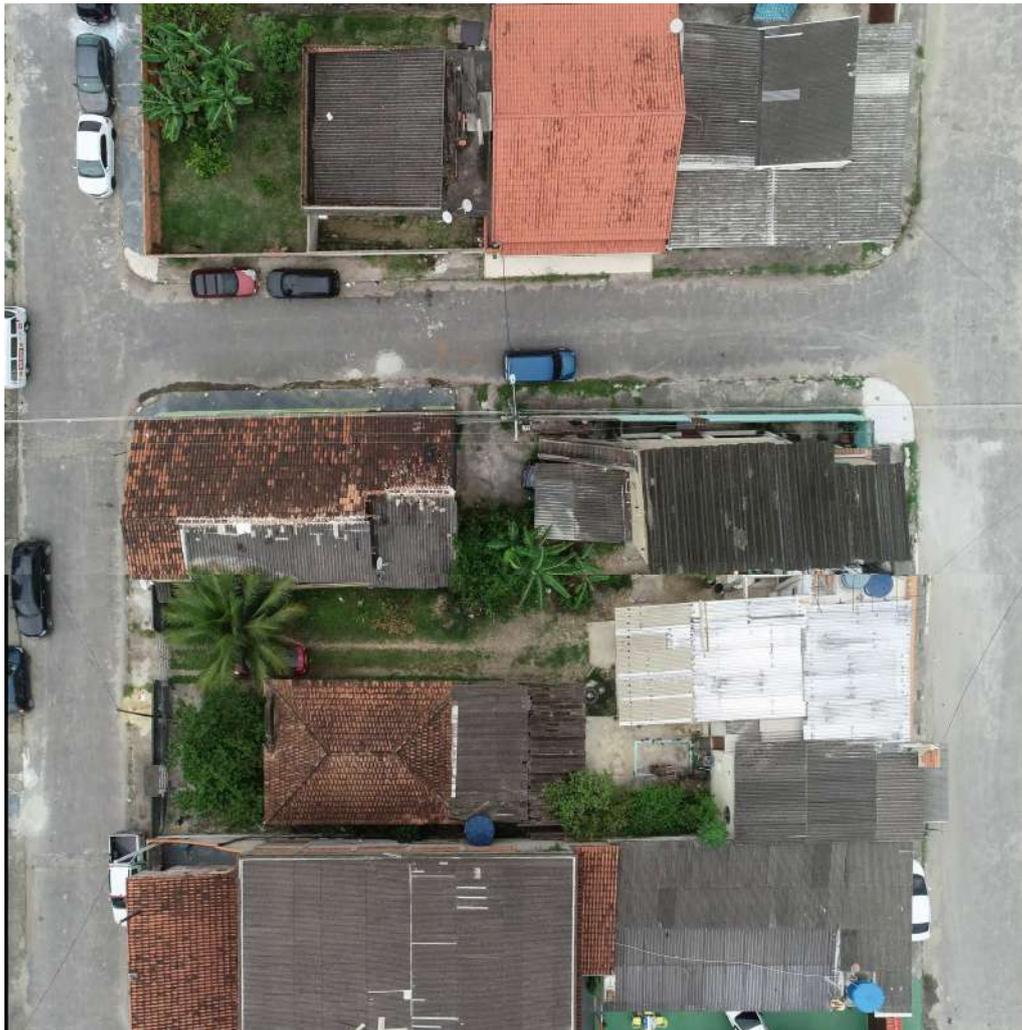
Utilizando o método de cálculo de correlação de fase cumulativa, escolhendo uma sobreposição de 20% entre quadros consecutivos na costura, observou-se alguns casos bem comportados (Figura 3.4), em que a simples sobreposição de um dos quadros sobre o outro gera uma aparência natural na costura, porém também foi observado que em outros casos ocorria uma transição perceptivelmente abrupta entre as imagens (Figura 3.5c), pois a translação calculada através da correlação de fase, não resultou no alinhamento perfeito entre os quadros. Com o intuito de melhorar a aparência da costura, foi proposto a suavização através da aplicação de uma janela cossenoidal sobre a região de pixels a ser sobreposta na costura, com o objetivo de suavizar a percepção do desalinhamento.



(a) Imagem que será costurada na parte de cima.



(b) Imagem que será costurada na parte de baixo.



(c) Resultado da costura.

Figura 3.4: Costura de aparência bem comportada.

O processo de suavização de bordas sobrepostas consiste em, durante a colagem de um quadro, considerar quais regiões desse quadro sofrerão sobreposição por outros quadros ao longo do processo de costura. Considerando um panorama entre dois

quadros transladados verticalmente, o quadro que será costurado na parte superior do panorama terá sua parte inferior sobreposta com a parte superior do quadro que será colado na região inferior do panorama, então uma janela de uma função cosseno ao quadrado será aplicada sobre esse quadro de forma a minimizar a influência dos pixels mais próximos da borda inferior desse quadro, enquanto que a imagem inferior do panorama, terá uma função cosseno ao quadrado que minimiza a influência dos pixels mais próximos da borda superior dessa imagem. Quando essas duas regiões forem somadas, ocorrerá uma soma entre os pixels das imagens, mas devido as janelas cossenoidais aplicadas, o pixel de cada imagem nessa região terá um peso diferente, de tal forma que a soma dos pesos seja um. Esse processo é exemplificado pela Figura 3.5.



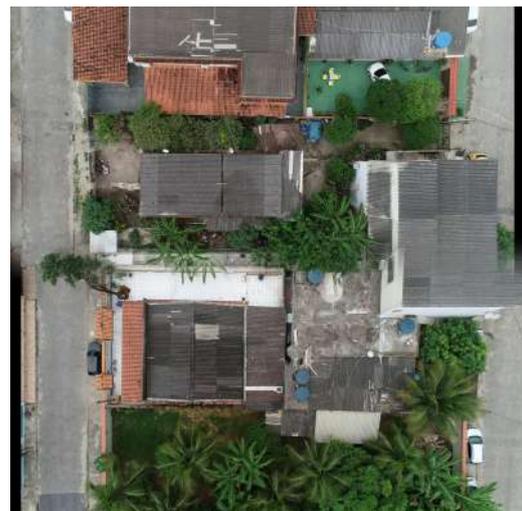
(a) Imagem que será costurada na parte de cima com suavização de área sobreposta.



(b) Imagem que será costurada na parte de baixo com suavização de área sobreposta.



(c) Costura sem aplicação de suavização de área sobreposta.



(d) Costura com aplicação de suavização de área sobreposta.

Figura 3.5: Suavização de área sobreposta através de janela cossenoidal.

O algoritmo proposto para a costura de imagens extraídas de vídeos gravados por VANTs utilizando correlação de fase foi desenvolvido em *Python* e estruturado de maneira modular, com o objetivo de garantir a flexibilidade e a escalabilidade do processo. Esse algoritmo é composto por várias etapas interdependentes, cada uma delas responsável por uma tarefa específica no fluxo de trabalho. A seguir, são detalhadas as etapas principais do algoritmo, abrangendo desde a extração dos quadros dos vídeos até a geração do panorama final, passando por cálculos de correlação de fase, posicionamento de quadros, e aplicação de técnicas de suavização para a integração das imagens.

Inicialmente, selecionados os vídeos que representam uma região monitorada, os quadros individuais desses vídeos são extraídos para a computação da correlação de fase entre pares de quadros consecutivos da região. Em seguida, um limite máximo de sobreposição entre os quadros é estabelecido e é calculada a translação relativa entre quadros não consecutivos, conforme a Equação 3.1, até que seja encontrada uma translação vertical ou horizontal entre dois quadros de tal forma que a sobreposição entre eles seja menor ou igual ao limite de sobreposição arbitrado.

Depois de selecionados os quadros com sobreposição desejada, sabendo a translação entre os quadros, é calculado o valor de translação máximo horizontal e vertical para encontrar a dimensão da matriz de pixels que irá comportar costura, assim como é estabelecido o posicionamento de cada quadro dentro dessa matriz.

Para cada quadro a ser costurado, considera-se a posição dos outros quadros na matriz da costura para saber quais regiões do quadro atual sofrerão sobreposição por outros quadros da costura, e para cada região considerada, é aplicada a janela cossenoidal de suavização.

3.3.1.2 Validação de Resultados

Seguindo o algoritmo apresentado, o resultado obtido mostrado na Figura 3.6a deixa a desejar e algumas análises e observações foram feitas buscando a melhoria desse resultado. A primeira observação feita foi a de que utilizar uma janela cossenoidal fixa não se mostrou promissor, pois o deslocamento entre quadros vizinhos

da costura não necessariamente será o valor estabelecido de sobreposição entre eles. Uma imagem considerando janelas cossenoidais variáveis de acordo com o tamanho da região sobreposta está mostrado na Figura 3.6b, essa modificação apresenta uma melhora nítida na visualização do panorama, mas ainda é possível observar que outros ajustes são necessários. Observando a parte inferior dessa imagem, nota-se que há um descasamento horizontal maior nas imagens com vizinhança vertical no lado esquerdo que nas imagens com vizinhança vertical no lado direito. Considerando que para esse panorama o VANT inicia a trajetória de voo no canto superior direito, move-se para a esquerda da imagem, translada para baixo, move-se para a direita, translada para baixo novamente e então desloca-se para a esquerda, suspeita-se que esteja ocorrendo propagação de erro ao longo do posicionamento dos quadros.



(a) Costura do panorama aplicando janela cossenoidal de tamanho fixo.



(b) Costura do panorama aplicando janela cossenoidal de tamanho variável.

Figura 3.6: Comparação do panorama com janela fixa e janela variável.

A primeira hipótese sobre a propagação de erro no alinhamento da imagem foi de que possa ter ocorrido devido a discretização das posições de cada quadro. Quando o algoritmo calcula a posição dos quadros utilizando o resultado da correlação de fase, esse resultado apresenta translação de pixels com valores reais, não necessariamente valores inteiros, mas para posicionar o quadro no mosaico deve-se colocar o quadro

numa posição de valor inteiro da matriz. Para averiguar essa hipótese construiu-se um gráfico apresentando as posições discretas e as posições reais de cada quadro da costura, conforme apresentado na Figura 3.7. O gráfico mostra que as posições discretas estão alinhadas com as posições reais, portanto a hipótese foi descartada.

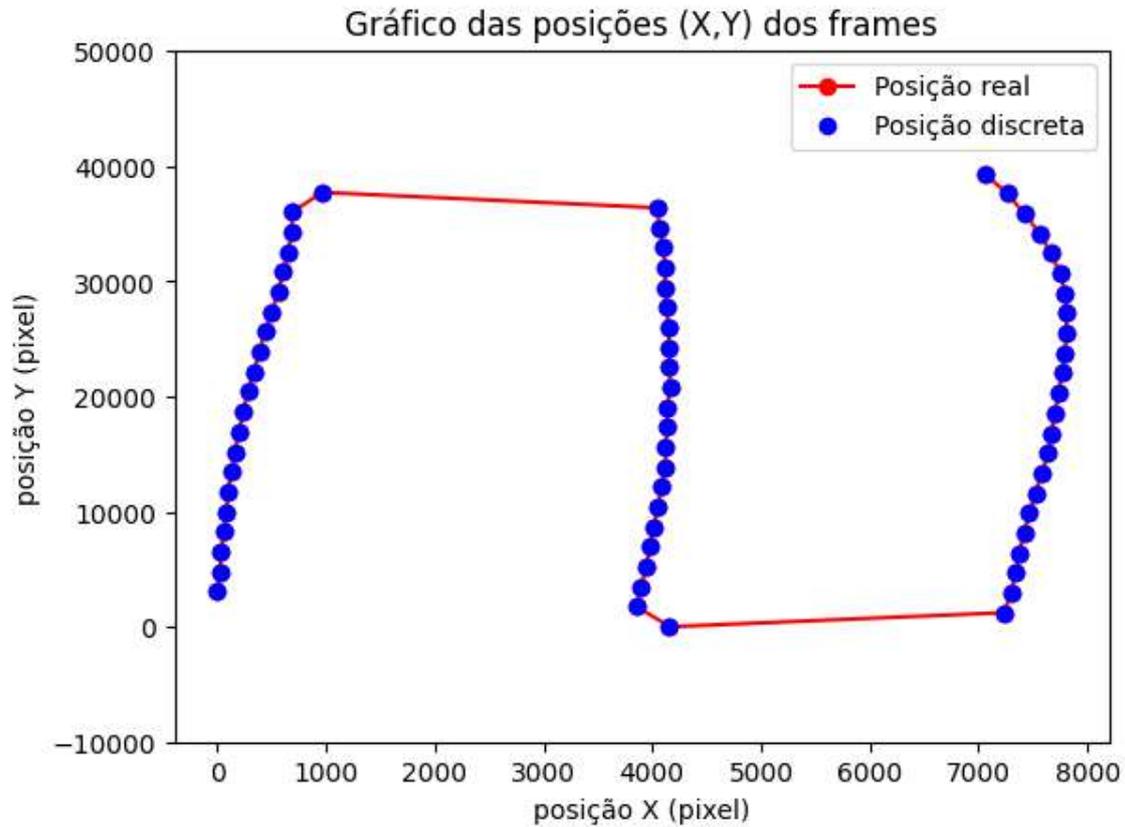


Figura 3.7: Gráfico comparativo de posição discreta e posição real dos quadros do panorama.

Uma segunda hipótese levantada foi de que o cálculo da correlação de fase possuía uma imprecisão, como o algoritmo calcula as posições dos quadros somando os resultados obtidos pela correlação de fase entre pares consecutivos de quadros, essa imprecisão pode se propagar ao longo dessa soma. Para verificar essa hipótese foi proposto que o método utiliza maior espaçamento entre quadros no cálculo da correlação de fase, fazendo com que menos somas fossem necessárias ao logo do posicionamento dos quadros. O objetivo dessa proposta é verificar se ocorre uma diminuição no desalinhamento do panorama.

Os resultados obtidos pela implementação dessa hipótese estão apresentados na Figura 3.8. O Passo é a distância entre os dois quadros que serão utilizados para o cálculo da correlação de fase, portanto quanto maior o passo menor a sobreposição entre os quadros utilizados para contar a correlação de fase e maior é a diferença de perspectiva em relação a um mesmo objeto em cada quadro, diminuindo a precisão do resultado, porém menor é a quantidade de somas que precisam ser feitas para encontrar as posições dos quadros. Observou-se que o aumento muito grande de distância entre os quadros do cálculo da correlação de fase gera um panorama muito desalinhado (Figura 3.8f) e que a alteração no passo não refletiu em modificação no descasamento das imagens do canto superior esquerdo da costura.

Levantou-se, então, a hipótese de que os vídeos talvez não fossem tão comportados como o esperado, o plano de voo foi montado com o objetivo de que o drone mantivesse uma altura fixa, que não ocorresse inclinações e nem rotação da câmera durante o registro dos vídeos, para que as imagens sofressem apenas translações que poderiam facilmente ser calculadas pela correlação de fase. Caso durante a gravação dos vídeos o drone alterasse a sua altura ou a câmera dele sofresse algum tipo de rotação ou inclinação, a correlação de fase não seria capaz de interpretar essas transformações, gerando um comportamento inesperado no cálculo da posição do quadro. Com o intuito de investigar a trajetória do VANT durante a gravação dos vídeos, foi proposto o uso de Structure-from-Motion [12], que estima a posição das câmeras e reconstrói a estrutura 3D de uma cena a partir de imagens 2D, através do COLMAP.



(a) Passo = 3.



(b) Passo = 5.



(c) Passo = 10.



(d) Passo = 12.



(e) Passo = 25.



(f) Passo = 50.

Figura 3.8: Exemplos de costuras com quantidade diferentes de salto entre quadros na computação da correlação de fase.

3.3.2 COLMAP

Quando proposto, o objetivo do uso do COLMAP no projeto era de utilizar as informações que o software disponibiliza sobre a pose da câmera. Optando por utilizar os mesmos quadros utilizados nos panoramas presentes na Figura 3.6, não foi possível gerar um modelo esparso da cena, pois a sobreposição entre os quadros era muito pequena, dificultando a geração de correspondência entre as imagens. Portanto, foi considerado um novo grupo de quadros, obtido através da utilização de todos os quadros de numeração múltipla de 50, adicionando o primeiro e o último quadro, *frame_00001.jpeg* *frame_12113.jpeg* no caso da região utilizada como exemplo na costura da Figura 3.6.

Após a seleção dos quadros utilizados para computar o Structure-from-Motion da região, utilizou-se o *COLMAP-CLI* (COLMAP Command-line Interface, interface de linha de comando do COLMAP em português) para criar o modelo 3D esparso do panorama, as informações observadas no *COLMAP-GUI* (COLMAP Graphical User Interface, interface gráfica do COLMAP em português) são apresentadas na Figura 3.9, essa figura evidencia um comportamento inesperado em relação à trajetória do VANT e o posicionamento dos quadros. Levando em conta que os quadros utilizados para essa reconstrução 3D foram extraídos de um vídeo retificado e que o vídeo não retificado também encontra-se disponível no conjunto de dados, optou-se por realizar a reconstrução 3D através de quadros extraídos do vídeo não retificado. Essa reconstrução apresentou resultados interessantes, conforme mostrado na Figura 3.10, instigando a realização de uma reconstrução densa da cena, Figura 3.11.

O COLMAP gera diferentes modelos de reconstrução densa, incluindo a nuvem de pontos densa (*fused.ply*), que representa a cena por meio de um conjunto de pontos tridimensionais [37]. Além disso, disponibiliza malhas 3D, como a malha Poisson (*meshed_poisson.ply*), que utiliza um método de reconstrução orientado à suavização de superfícies [38], e a malha Delaunay (*meshed_delaunay.ply*), baseada na triangulação Delaunay, ideal para geometrias com maior definição [39]. Esses arquivos possibilitam análises detalhadas e visualizações variadas da cena reconstruída.



(a) Perspectiva superior das poses de câmera.



(b) Perspectiva frontal das poses de câmera.



(c) Perspectiva lateral das poses de câmera.



(d) Perspectiva frontal do modelo esparsos da região.

Figura 3.9: Modelo 3D gerado pelo COLMAP com apresentação das poses de câmera em vermelho para os quadros usados no vídeo retificado e apresentação dos pontos esparsos.

A Figura 3.11a apresenta a nuvem de pontos densa do arquivo *fused.ply* e a Figura 3.11b apresenta a malha Poisson do arquivo *meshed_poisson.ply*. Essas figuras apresentam uma visualização da cena bastante fiel a realidade, e motivaram o uso dos dados que o COLMAP disponibiliza nos arquivos *cameras.txt*, *images.txt* e *points3D.txt*, referentes ao modelo esparsos, como mostrado na Figura 3.10, para a realização da costura de imagens utilizando homografia.

Por padrão, o COLMAP usa arquivos de formato binário para armazenar informações do modelo esparsos de pontos. Porém, oferece a possibilidade de exportar essas informações em arquivos de texto, esses arquivos são: *cameras.txt*, *images.txt* e *points3D.txt*. O arquivo *cameras.txt* contém informações sobre os parâmetros intrínsecos das câmeras, no caso do projeto todas as imagens foram adquiridas pela mesma câmera, então informamos ao COLMAP para que considerasse apenas um



(a) Perspectiva superior das poses de câmara.



(b) Perspectiva frontal das poses de câmara.



(c) Perspectiva superior das poses de câmara.



(d) Perspectiva lateral do modelo esparsos da região.

Figura 3.10: Modelo 3D gerado pelo COLMAP com apresentação das poses de câmara em vermelho para os quadros usados do vídeo não retificado e apresentação dos pontos esparsos.

modelo de câmara para todos os quadros. O arquivo *images.txt* contém informações sobre a pose da câmara e os keypoints (pontos-chaves) usados na construção do modelo esparsa para todas as imagens utilizadas no modelo. O arquivo *points3D.txt* contém informações sobre todos os pontos 3D do modelo esparsa.

A homografia de imagens é uma transformação projetiva que mapeia pontos entre dois planos correspondentes em diferentes perspectivas, preservando a colinearidade e a razão cruzada [32, 40]. É amplamente utilizada em visão computacional para tarefas como alinhamento de imagens, mosaicos panorâmicos e retificação, pois permite transformar imagens capturadas de diferentes pontos de vista em um plano comum. A matriz de homografia é geralmente estimada utilizando correspondências entre pontos correspondentes nas imagens (keypoints), empregando métodos como o RANSAC para maior robustez contra outliers [14]. Com os dados contidos nesses arquivos, é possível realizar a costura através da computação de homografias para cada imagem costurada.



(a) Nuvem de pontos densa.



(b) Malha de Poisson.

Figura 3.11: Reconstruções densas do COLMAP.

3.3.2.1 Validação de Resultados

Iniciando o processo de validação da proposta, duas imagens em sequencia foram separadas para realização da costura entre elas. O primeiro passo foi a identificação dos *keypoints* em comum que as imagens possuíam, essa informação foi extraída do arquivo *images.txt*, em seguida estima-se a homografia utilizando o RANSAC para reduzir a influência dos *outliers* e a costura das imagens é realizada. As Figuras 3.12a e 3.12b mostram um par de imagens com pontos vermelhos representando os *keypoints* em comum dessas imagens que foram utilizados no modelo 3D esparsa do panorama, a Figura 3.12c mostra o resultado da costura.



(a) Imagem superior destacando keypoints em comum.



(b) Imagem inferior destacando keypoints em comum.



(c) Homografia entre as imagens.

Figura 3.12: Homografia entre par de imagens através do método RANSAC

Foi-se aumentando o número de homografias sucessivamente com o objetivo de observar o comportamento da costura conforme o aumento do número de imagens. Para escalar a costura por homografia, foi feita a computação de homografias relati-

vas, para essa conta, a homografia da terceira imagem a ser costurada no panorama deve ser computada como um produto matricial da homografia entre a terceira imagem e a segunda imagem com a homografia da segunda imagem com a primeira imagem, seguindo a equação dada por

$$H_{1,n} = H_{1,2} \cdot H_{2,3} \cdot H_{3,4} \cdots H_{n-1,n} = \prod_{k=2}^n H_{k-1,k}, \quad (3.2)$$

onde:

- $H_{A,B}$: Matriz de Homografia entre o quadro B e o quadro A .

A Figura 3.13, apresenta o resultado para a costura de quatro imagens, apresentando um resultado promissor, com quadros bem alinhados. Mas conforme o número de imagens vai progredindo, observa-se que as primeiras imagens consideradas na costura vão sendo distorcidas e quanto maior o número de imagens, maior a distorção, esse efeito é apresentado na Figura 3.14.

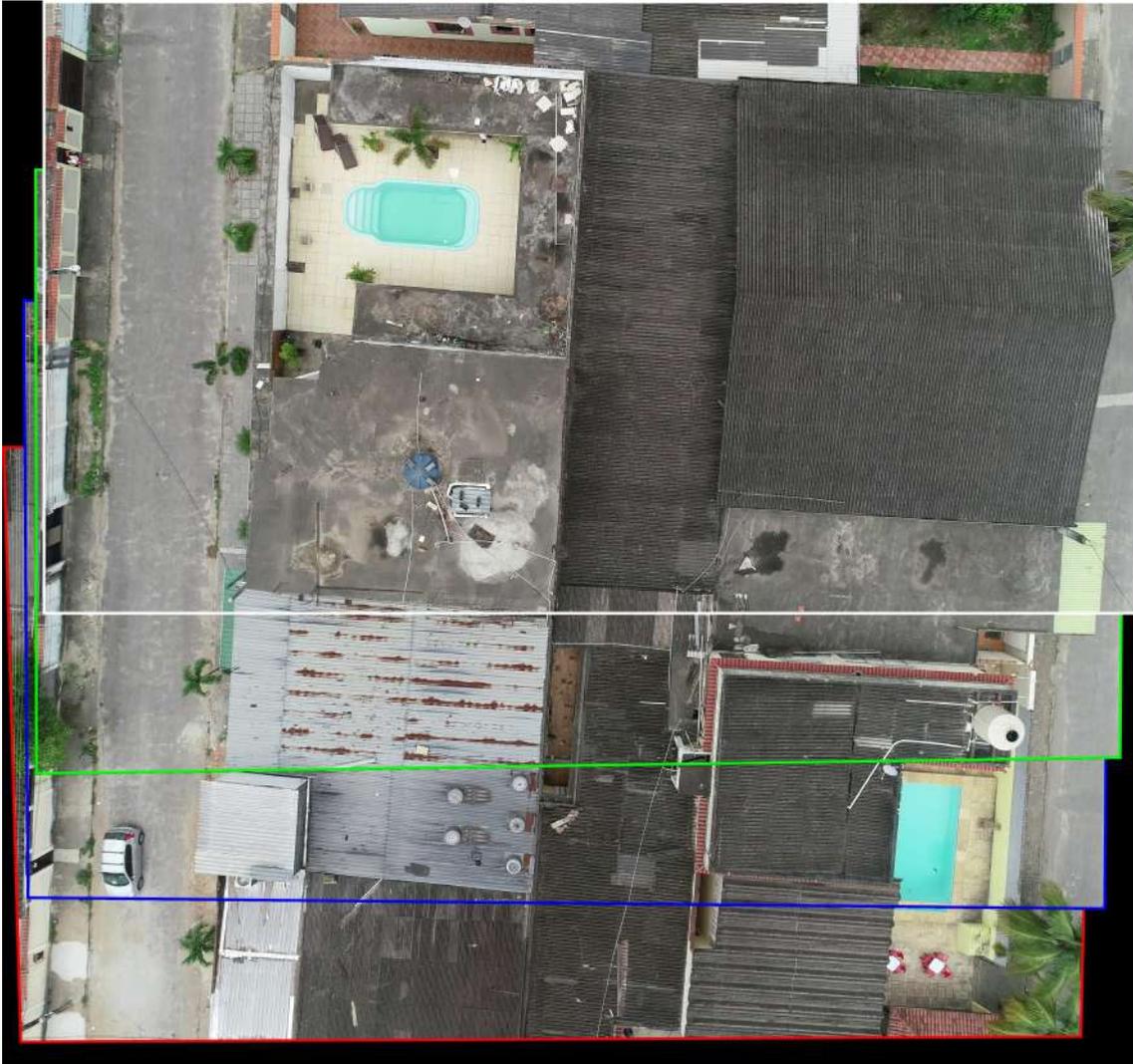


Figura 3.13: Quatro quadros costurados pela técnica de homografias relativas.



Figura 3.14: Costura distorcida.

A hipótese levantada foi a de que o sistema de múltiplas poses de câmera usado causa uma propagação de erro no cálculo das homografias relativas, então criou-se, manualmente, uma nova pose de câmera com o intuito de projetar os pontos 3D do modelo esparsos em sua matriz de pixels, e cada imagem teria sua homografia cal-

culada diretamente nesse novo panorama, evitando a propagação de erro ao longo do produto sucessivo das matrizes de homografia. Revisitando a Figura 3.10, temos que criar uma nova câmera com uma altura superior a altura das câmeras mostradas nessa figura, a distância focal deve aumentar proporcionalmente a altura e as dimensões da imagem vista por essa câmera devem ser suficientes para a projeção de todos os pontos 3D do modelo esparso. A câmera em questão está apresentada na Figura 3.15



(a) Perspectiva superior das poses de câmera.



(b) Perspectiva frontal das poses de câmera.



(c) Perspectiva lateral das poses de câmera.

Figura 3.15: Perspectivas da nova câmera para computação das homografias diretas.

A projeção dos pontos 3D na matriz de pixels dessa nova câmera é realizada com base nos dados fornecidos pelo arquivo *points3D.txt* do COLMAP e envolve uma sequência de transformações matemáticas. Primeiramente a matriz de rotação da câmera é construída a partir dos quatérnios, um número de quatro partes usado em rotações e orientações tridimensionais associados à pose da câmera [41], matriz de rotação é dada por

$$R = \begin{bmatrix} 1 - 2(q_y^2 + q_z^2) & 2(q_x q_y - q_z q_w) & 2(q_x q_z + q_y q_w) \\ 2(q_x q_y + q_z q_w) & 1 - 2(q_x^2 + q_z^2) & 2(q_y q_z - q_x q_w) \\ 2(q_x q_z - q_y q_w) & 2(q_y q_z + q_x q_w) & 1 - 2(q_x^2 + q_y^2) \end{bmatrix}, \quad (3.3)$$

onde q_x, q_y, q_z e q_w são, respectivamente, as componentes x, y, z e w do quatérnio contido no arquivo *images.txt*.

As coordenadas 3D dos pontos, inicialmente no sistema de referência global (sistema de coordenadas no qual a reconstrução tridimensional é representada), são transformadas para o sistema de coordenadas da câmera por

$$P_c = R \cdot P_w + t, \quad (3.4)$$

onde P_w representa o ponto no sistema de referência global, R é a matriz de rotação derivada dos quatérnios, e t é o vetor de translação.

Posteriormente, as coordenadas 3D no sistema da câmera são projetadas para o plano da imagem utilizando a matriz intrínseca K , que contém parâmetros como a distância focal (f_x, f_y) e o centro óptico (c_x, c_y). A projeção é dada por

$$p = K \cdot P_c, \quad (3.5)$$

onde p representa as coordenadas homogêneas no plano da imagem. O resultado final são coordenadas de pixels que incluem informações RGB - também extraída do arquivo *points3D.txt* - associadas a cada ponto projetado.

Após encontrar as coordenadas de projeção de cada pixel do modelo esparso na matriz da câmera, os pontos foram pintados com um raio de 20 pixels, resultando na Figura 3.16.

Com a projeção dos pontos 3D na matriz, basta identificar quais pontos estão presentes em um determinado quadro para computar a homografia desse quadro na matriz. O resultado das homografias usando essa técnica é mostrado na Figura 3.17. O resultado é muito superior ao da correlação de fase, apesar de ainda ter espaço para aperfeiçoamento. No geral, a costura está bem alinhada, o método corrigiu o desalinhamento observado na solução por correlação de fase e produziu um



Figura 3.16: Projeção 2D dos pontos 3D do modelo esparso.

panorama que, no geral, representa bem a região monitorada. Porém é perceptível na imagem que a homografia de alguns quadros estão não correspondentes com a costura, causando pequenos desalinhamentos em algumas regiões da costura. Apesar da costura ainda ter pontos de melhoria, o resultado foi considerado suficiente para o avanço do projeto com os próximos passos do sistema proposto, a apresentação dos criadouros de mosquito.



Figura 3.17: Homografia dos pontos projetados considerando uma única pose de câmera.

3.4 Caixas delimitadoras

Uma *bouding box* (caixa delimitadora, em português) é um retângulo que envolve um objeto de interesse dentro de uma imagem, definida pelos pontos dos cantos superior esquerdo (*top left*) e inferior direito (*bottom right*). Com o intuito de apresentar na costura os criadouros de mosquito identificados, foi incluído na costura as *bouding boxes* referentes às caixas d'água anotadas para a região da costura.

Durante o processo de inserção das *bouding boxes* na costura, percebeu-se que, como as anotações foram feitas utilizando o vídeo retificado, enquanto a costura foi gerada a partir do vídeo não retificado, as caixas delimitadoras ficariam posicionadas de forma incorreta no panorama, conforme ilustrado na Figura 3.18, e que, para posicionamento correto das caixas delimitadoras dos objetos anotados no panorama, seria necessário realizar um remapeamento dos pontos da imagem retificada para seus pontos correspondentes na imagem não retificada.



(a) Caixa delimitadora apresentada em um quadro retificado.



(b) Caixa delimitadora apresentada em um quadro não retificado.

Figura 3.18: Caixas delimitadoras anotadas na base de dados apresentadas em um quadro retificado e em um quadro não retificado.

Em um primeiro momento, sabendo que a retificação dos vídeos foi feita através do método de calibração Zhang [36], com um vídeo contendo diferentes poses de um tabuleiro de xadrez, foi feita uma tentativa de replicar esse método para obter a matriz de parâmetros da câmera, e através de sua inversa remapear os pontos do quadro retificado para o quadro não retificado. A Figura 3.19a mostra um resultado insatisfatório e uma ideia alternativa de realizar o mapeamento de pontos entre os quadros através da computação de uma homografia entre eles foi proposta. Esse método



(a) Caixas delimitadoras apresentadas em quadro não retificado após uso da matriz inversa de parâmetros de distorção da câmera.



(b) Caixas delimitadoras apresentadas em quadro não retificado após uso da homografia entre o quadro retificado e não retificado.

Figura 3.19: Apresentação de novas posições das caixas delimitadoras considerando os métodos de remapeamento de pontos entre o quadro retificado e não retificado propostos.

alternativo faz uso do ORB como detector de características para obter os pontos correspondentes entre as imagens e poder estimar a matriz de homografia entre elas. Esse método apresentou bom resultado, conforme mostrado na Figura 3.19b.

Sabendo a posição correta das *bounding boxes* dos objetos anotados em cada quadro utilizado na costura, pode-se adicionar as caixas delimitadoras dos objetos na costura panorâmica da região monitorada. Com isso, obtêm-se uma representação de alta resolução da área capturada pelo VANT, com indicações precisas da localização de potenciais focos de mosquito *Aedes aegypti*. A Figura 3.20, mostra um zoom da adição de marcações de caixas delimitadoras no panorama da Figura 3.17.



Figura 3.20: Imagem ampliada de um trecho da costura apresentando as *bouding boxes* de caixas d'água anotadas na base de dados.

3.5 Georreferenciamento de objetos e mapa de calor

Outra forma proposta de apresentação dos focos de mosquito é através do georreferenciamento desses objetos. Utilizando os objetos anotados na base de dados e resultados obtidos pela costura feita com COLMAP e homografias, fez-se a estimação das coordenadas desses objetos para apresentação em um mapa. A apresentação é feita tanto utilizando-se pins para indicar os locais dos objetos como construindo um Mapa de Calor, para facilitar a diferenciação entre locais de maior e menor predominância de potenciais criadouros de mosquito.

Nos arquivos de registro de voo disponíveis no banco de dados MBG_V2, é possível obter informações de coordenadas e altitude do VANT ao longo do voo, porém a taxa de amostragem desses arquivos é inferior a taxa de amostragem dos arquivos de vídeo, sendo necessária a realização de uma interpolação dos resultados, para

estimar o valor de latitude, longitude e altitude do quadricóptero a cada quadro da filmagem.

Com as informações de altitude, latitude e longitude do VANT combinadas com o campo de visão da câmera e a resolução de imagem dos quadros é possível encontrar a variação de coordenada em relação a variação de pixels de um determinado quadro.

O processo inicia-se utilizando o valor de altitude do VANT e campo de visão (FOV) da câmera para calcular o valor da diagonal em metros da imagem registrada

$$d = 2 \cdot h \cdot \tan\left(\frac{FOV}{2}\right), \quad (3.6)$$

onde d é a diagonal da imagem em metros, h é a altitude do VANT em relação ao chão. Em seguida é encontrada a escala e de deslocamento em pixels/metros, utilizando o valor de diagonal em metros e o valor de diagonal em pixels, dada pela altura h_{pixel} e largura w_{pixel} da matriz de pixels da imagem

$$e = \frac{d}{\sqrt{(h_{\text{pixel}}^2 + w_{\text{pixel}}^2)}}. \quad (3.7)$$

Com essa escala, podemos encontrar a distância em metros horizontal e vertical de qualquer pixel da imagem em relação ao centro da imagem, posição onde o quadricóptero se encontra. Porém, os eixos de latitude e longitude não necessariamente correspondem aos eixos verticais e horizontais da imagem, é necessário encontrar uma função que mapeie a direção dos versores de latitude e longitude dentro da matriz de pixels da imagem.

Para encontrar os versores de latitude e longitude dentro do espaço de pixels da imagem fez-se uso do resultado da costura obtida com os arquivos de saída do COLMAP. Utiliza-se a posição do pixel central do primeiro e último quadro costurado no mosaico, bem como as coordenadas obtidas do arquivo de registro de voo, para construir um sistema linear dado por

$$\begin{bmatrix} \Delta x & \Delta y \\ -\Delta y & \Delta x \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Delta \text{lat} \\ \Delta \text{lon} \end{bmatrix}, \quad (3.8)$$

onde:

- $\Delta x, \Delta y$: Variação das coordenadas em pixels entre a primeira e a última imagem costuradas na matriz.
- $\Delta \text{lat}, \Delta \text{lon}$: Diferença das coordenadas geográficas entre a primeira e a última imagem costuradas na matriz.
- a, b : Coeficientes que definem a transformação entre deslocamentos em pixels e variações de latitude e longitude.

Os versores são dados pelos coeficientes desse sistema de acordo com a equação que se segue

$$\hat{v}_{\text{lat}} = \begin{bmatrix} \frac{a}{\sqrt{(a^2+b^2)}} \\ \frac{b}{\sqrt{(a^2+b^2)}} \end{bmatrix}, \hat{v}_{\text{lon}} = \begin{bmatrix} \frac{b}{\sqrt{(a^2+b^2)}} \\ \frac{-a}{\sqrt{(a^2+b^2)}} \end{bmatrix}. \quad (3.9)$$

Encontrados os versores de latitude e longitude, o próximo passo para encontrar a coordenada de um determinado pixel na imagem é fazer o produto vetorial entre a variação horizontal e vertical (em metros) do pixel ao centro do quadro no qual ele se encontra com o versor de latitude, para encontrar a variação de latitude, e o versor de longitude, para encontrar a variação de longitude e dividir por um fator de correção de graus para metros, onde 111320.0 é aproximadamente a quantidade de metros correspondente a variação de um grau (1 grau \approx 111.32km) e $\cos(\text{lat} \cdot \frac{\pi}{180})$ é a correção do fator de escala para longitude devido à convergência dos meridianos. Conforme mostrado pelas seguintes equações

$$\Delta \text{lat} = \frac{\Delta \text{pixels} \cdot \hat{v}_{\text{lat}}}{111320.0}, \quad (3.10)$$

$$\Delta \text{lon} = \frac{\Delta \text{pixels} \cdot \hat{v}_{\text{lon}}}{111320.0 \cdot \cos(\text{lat} \cdot \pi/180)}. \quad (3.11)$$

De posse da variação de latitude e longitude entre o pixel de interesse e o centro do quadro, basta somá-las ao valor de latitude e longitude do centro da imagem para encontrar a coordenada desse pixel. A Figura 3.21 exemplifica a orientação dos versores de latitude e longitude dentro de um quadro, a Figura 3.22a mostra as coordenadas do centro e dos cantos desse quadro no Google Maps e a Figura 3.22b sobrepõem o quadro associado com essas coordenadas no mapa.



Figura 3.21: Direção dos versores de latitude e longitude em um quadro.

Com a possibilidade de mapear a coordenada de qualquer pixel de qualquer imagem utilizada na costura, pode-se utilizar o arquivo de anotações da base de dados para georreferenciar os pixels correspondentes a criadouros de mosquito.

Utilizando o arquivo XML de anotações da base de dados MBG_V2 para encontrar os objetos de caixa d'água da região F da Tabela 3.1, criou-se uma lista de pontos para cada objeto distinto de caixa d'água encontrado nos quadros da costura. Se a mesma caixa d'água estiver presente em mais de um quadro, ela só será considerada uma única vez, associada ao ponto referente ao primeiro quadro em que ela foi identificada. O XML fornece os pontos referentes ao canto superior esquerdo (*top_left*) e ao canto inferior direito (*bottom_right*) da caixa delimitadora do objeto anotado, utiliza-se o pixel do ponto médio entre esses dois pontos para encontrar a coordenada do objeto. O mapa apresentando as coordenadas dos objetos anotados é evidenciado na Figura 3.23, tal qual o Mapa de Calor e a funcionalidade de um link de redirecionamento para o Google Maps com uma rota até a coordenada do objeto selecionado.



(a) Coordenadas do centro e dos cantos de um quadro no Google Maps.

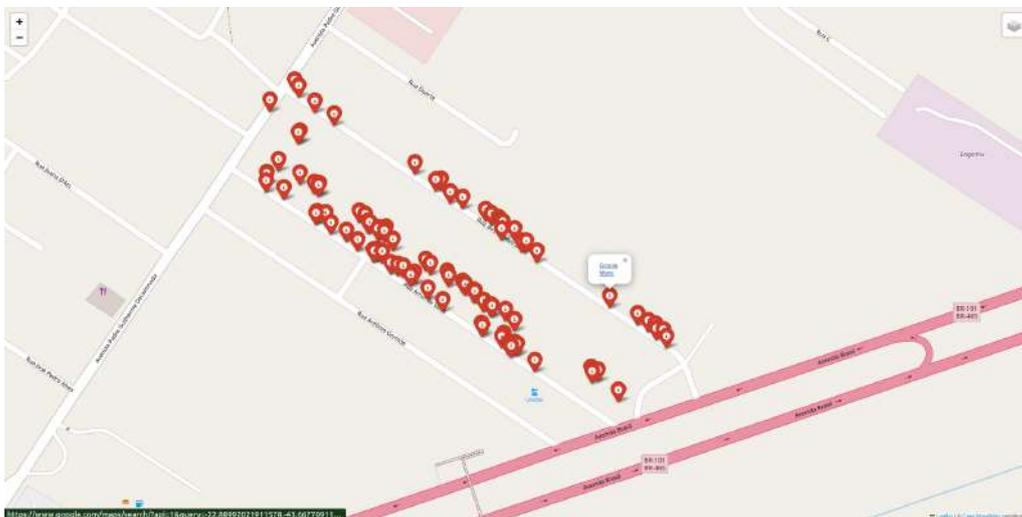


(b) Imagem de um quadro sobreposto ao mapa de satélite do Google Maps.

Figura 3.22: Sobreposição de um quadro e apresentação das coordenadas dos vértices do quadro.



(a) Pinos das coordenadas do centro de cada quadro utilizado na costura.



(b) Pino de localização das caixas d'água com link de redirecionamento para a coordenada no Google Maps.



(c) Mapa de calor das caixas d'água anotadas nos quadros da costura.

Figura 3.23: Mapa georreferenciado.

3.6 Conclusão

O capítulo apresentou a descrição detalhada da solução proposta, incluindo detalhes da base de dados utilizada e também apresentando e discutindo sobre os resultados obtidos ao longo do desenvolvimento do sistema. O capítulo apresentou um sistema capaz de monitorar e apresentar dados de forma eficiente para gerenciamento de políticas de combate a proliferação de mosquitos transmissores de arboviroses, indicar rotas para os locais de criadouros detectados através dos pinos com links de redirecionamento do Google Maps, indicar áreas mais críticas com o mapa de calor e fornecer uma representação detalhada do terreno monitorado, através da costura das imagens obtidas pelo VANT.

Capítulo 4

Conclusão e trabalhos futuros

Nos últimos anos, tem-se observado um crescente desenvolvimento de sistemas voltados para a detecção automática de objetos por meio de Veículos Aéreos Não Tripulados [16, 17, 18, 42, 43]. Esses avanços têm permitido uma análise detalhada de ambientes e a identificação de padrões relevantes para diversas áreas, incluindo a saúde pública. Dentro desse contexto, este trabalho se concentrou na aplicação dessas tecnologias para o combate às arboviroses, doenças de grande impacto sanitário e econômico, tanto no Brasil quanto globalmente.

Ao longo deste projeto, foi introduzido um sistema voltado para a identificação de potenciais focos de reprodução do mosquito *Aedes aegypti* utilizando imagens aéreas capturadas por VANTs [16]. Esse sistema de detecção automática motivou a proposta do presente trabalho como um instrumento complementar à ele, melhorando iniciativas de combate às arboviroses. Com o intuito de oferecer uma forma visual e interativa de monitoramento das áreas afetadas, o presente trabalho foi proposto com a ideia central de proporcionar uma representação fiel e de alta resolução das regiões monitoradas, utilizando a técnica de *Image Stitching* para a geração de panoramas detalhados e a marcação dos criadouros identificados.

A fim de aumentar a usabilidade do sistema, foi desenvolvido um método de georreferenciamento que permite a geolocalização precisa dos objetos detectados. Esse sistema posiciona pinos com links de rota no Google Maps, facilitando o acesso dos agentes de saúde aos locais identificados. Além disso, a incorporação de um mapa de calor permite uma visualização intuitiva das regiões com maior incidência

de criadouros, tornando mais eficiente o planejamento das ações de monitoramento e controle.

Os resultados deste trabalho indicam um grande potencial para a aplicação de VANTs e algoritmos de processamento de imagens na saúde pública. A utilização desse sistema pode auxiliar os governos na implementação de políticas públicas mais eficazes no combate à dengue, chikungunya, zika e febre amarela, otimizando o trabalho dos agentes de saúde e reduzindo custos operacionais. A automação do processo de detecção de potenciais criadouros minimiza a necessidade de inspeções manuais em larga escala, permitindo um direcionamento mais eficiente dos recursos humanos e financeiros.

Para trabalhos futuros, algumas melhorias e extensões podem ser consideradas. A integração de modelos mais avançados de aprendizado de máquina pode aumentar a precisão da detecção de criadouros. Além disso, o aprimoramento do sistema de georreferenciamento pode incluir informações adicionais, como a classificação de risco de cada local identificado. O aprimoramento da técnica atual ou a exploração de novas técnicas de *Image Stitching* também podem contribuir para uma representação ainda mais precisa das regiões analisadas. Pode-se desenvolver um método capaz de integrar o panorama com o mapa de pinos e calor, inserido-o como uma camada desse mapa. E deve-se encontrar um método automático de estimar a pose da câmera (Figura 3.15) a ser usada para projeção dos pontos 3D obtidos pelo COLMAP.

Em suma, este trabalho demonstra como a combinação de técnicas de processamento digital de imagens e aprendizado de máquina pode contribuir significativamente para a saúde pública. A detecção automatizada de criadouros do *Aedes aegypti* aliada a um sistema interativo de visualização e georreferenciamento pode se tornar uma ferramenta valiosa no combate às arboviroses, trazendo benefícios tanto para a população quanto para a gestão pública.

Referências Bibliográficas

- [1] CONTRIBUTORS, W., “Random sample consensus”, https://en.wikipedia.org/wiki/Random_sample_consensus, 2024, (Acesso em 6 Fevereiro 2025).
- [2] DJI Innovations, “DJI Phantom 4 Pro - Página de Suporte”, <https://www.dji.com/br/support/product/phantom-4-pro>, 2025, (Acesso em 29 Janeiro 2025).
- [3] DJI Innovations, *DJI Phantom 4 Pro User Manual*. v2.0 ed. Shenzhen, China, DJI Technology Co., Ltd., 2021. (Acesso em 29 Janeiro 2025).
- [4] RÜCKERT, C., WEGER-LUCARELLI, J., GARCIA-LUNA, S. M., *et al.*, “Impact of simultaneous exposure to arboviruses on infection and transmission by *Aedes aegypti* mosquitoes”, *Nature communications*, v. 8, n. 1, pp. 15412, 2017.
- [5] Pan American Health Organization, “Reported Cases of Dengue Fever In Americas”, <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/dengue-nacional-en/252-dengue-pais-ano-en.html>, 2021, (Acesso em 27 Agosto 2021).
- [6] “Brasil é país com mais casos de dengue no mundo, alerta OMS”, <https://agenciabrasil.ebc.com.br/saude/noticia/2023-12/brasil-e-pais-com-mais-casos-de-dengue-no-mundo-mostra-dados-da-oms>, 2023, (Acesso em 6 Fevereiro 2025).
- [7] “Dengue aumentou 400% no Brasil em 2024 em comparação ao ano passado”, <https://www.cofen.gov.br/dengue-aumentou-400-no-brasil-em-2024-em-comparacao-ao-ano-passado/>, 2024, (Acesso em 6 Fevereiro 2025).

- [8] “Health agencies investigate tropical mosquito found in Jackson County”, <https://www.opb.org/article/2024/10/11/health-agencies-investigate-tropical-mosquito-found-in-jackson-county/>, 2024, (Acesso em 6 Fevereiro 2025).
- [9] “Arboviroses - Microcefalias”, <https://mosquito.saude.es.gov.br/microcefalia>, 2024, (Acesso em 6 Fevereiro 2025).
- [10] LASERNA, A., BARAHONA-CORREA, J., BAQUERO, L., *et al.*, “Economic impact of dengue fever in Latin America and the Caribbean: a systematic review”, *Revista Panamericana de Salud Pública*, v. 42, pp. e111, 2018.
- [11] BORGES, E., “Investimentos no controle da dengue somam R\$ 1,5 bilhão para o período sazonal”, <https://www.gov.br/saude/pt-br/assuntos/noticias/2024/novembro/investimentos-no-controle-da-dengue-somam-r-1-5-bilhao-para-o-periodo-sazonal>, 2024, (Acesso em 6 Fevereiro 2025).
- [12] SCHÖNBERGER, J. L., FRAHM, J.-M., “Structure-from-Motion Revisited”. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [13] SCHÖNBERGER, J. L., FRAHM, J.-M., “Colmap”, <https://colmap.github.io/>, (Acesso em 21 Janeiro 2025).
- [14] FISCHLER, M. A., BOLLES, R. C., “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, *Communications of the ACM*, v. 24, n. 6, pp. 381–395, 1981.
- [15] PASSOS, W. L., ARAUJO, G. M., DE LIMA, A. A., *et al.*, “Automatic detection of *Aedes aegypti* breeding grounds based on deep networks with spatio-temporal consistency”, *Computers, Environment and Urban Systems*, v. 93, pp. 101754, 2022.
- [16] PASSOS, W. L., *Detection of Potential Aedes aegypti Breeding Grounds using Computer Vision and Machine Learning*. Ph.D. dissertation, COPPE, Universidade Federal do Rio de Janeiro, 2024. Tese de Doutorado.

- [17] MENEZES, M. A. G., MAROQUIO, R., MOREIRA, E. M., *et al.*, “Veículos Aéreos Não Tripulados para a Vigilância de Áreas Urbanas em Cidades Inteligentes”, *Anais do III Workshop Brasileiro de Cidades Inteligentes*, , 2022.
- [18] LIMA, G. A., COTRIN, R. O., BELAN, P. A., *et al.*, “Sistema de Visão Computacional para Identificação Automática de Potenciais Focos do Mosquito *Aedes aegypti* Usando Drones”, *Revista Ibérica de Sistemas e Tecnologias de Informação*, v. 43, pp. 93–108, 2021.
- [19] SAYEEDI, M. F. A., HAFIZ, F., RAHMAN, M. A., “MosquitoFusion: A Multiclass Dataset for Real-Time Detection of Mosquitoes, Swarms, and Breeding Sites Using Deep Learning”, 2024.
- [20] REN, S., HE, K., GIRSHICK, R., *et al.*, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 6, pp. 1137–1149, Jun. 2017.
- [21] RANKOV, V., LOCKE, R. J., EDENS, R. J., *et al.*, “An Algorithm for Image Stitching and Blending”. In: *Proceedings of SPIE - The International Society for Optical Engineering*, v. 5701, pp. 190–199, 2005.
- [22] XU, M., LIU, X., WAN, C., “Real-Time Stitching Algorithm of Vehicle Side View Image Based on Multi-Region Fast Phase Correlation”, *IEEE Access*, v. 13, pp. 5076–5091, 2025.
- [23] KUGLIN, C. D., “The phase correlation image alignment method”. In: *IEEE Int. Conf. on Cybernetics and Society, 1975*, pp. 163–165, 1975.
- [24] MENEGHETTI, G., DANELLJAN, M., FELSBURG, M., *et al.*, “Image Alignment for Panorama Stitching in Sparsely Structured Environments”. In: *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, v. 9127, *Lecture Notes in Computer Science*, pp. 428–439, 2015.
- [25] ALBA, A., AGUILAR-PONCE, R. M., VIGUERAS-GÓMEZ, J. F., *et al.*, “Phase Correlation Based Image Alignment with Subpixel Accuracy”. In: *Proceedings of MICAI 2012, Part I, Lecture Notes in Artificial Intelligence*, v. 7629, pp. 171–182, 2013.

- [26] SZELISKI, R., “Image Alignment and Stitching: A Tutorial”, *Foundations and Trends in Computer Graphics and Vision*, v. 2, n. 1, pp. 1–104, 2006.
- [27] BROWN, M., LOWE, D. G., “Automatic Panoramic Image Stitching using Invariant Features”, *International Journal of Computer Vision*, v. 74, n. 1, pp. 59–73, 2007.
- [28] SZELISKI, R., *Computer Vision: Algorithms and Applications*. Springer, 2010.
- [29] LOWE, D. G., “Distinctive image features from scale-invariant keypoints”, *International journal of computer vision*, v. 60, n. 2, pp. 91–110, 2004.
- [30] RUBLEE, E., RABAUD, V., KONOLIGE, K., *et al.*, “ORB: An efficient alternative to SIFT or SURF”. In: *International conference on computer vision (ICCV)*, pp. 2564–2571, IEEE, 2011.
- [31] “Aerial Navigation in GPS Denied Environments”, <https://textzip.github.io/posts/Orthomosaic-SLAM/results>, 2022, (Acesso em 6 Fevereiro 2025).
- [32] HARTLEY, R., ZISSERMAN, A., *Multiple View Geometry in Computer Vision*. 2 ed. Cambridge University Press, 2004.
- [33] MathWorks, “What Is Structure from Motion?”, <https://www.mathworks.com/help/vision/ug/what-is-structure-from-motion.html>, n.d., (Acesso em 26 Janeiro 2025).
- [34] ULLMAN, S., *The Interpretation of Visual Motion*. MIT Press, 1979.
- [35] SCHÖNBERGER, J. L., ZHENG, E., POLLEFEYS, M., *et al.*, “Pixelwise View Selection for Unstructured Multi-View Stereo”. In: *European Conference on Computer Vision (ECCV)*, 2016.
- [36] ZHANG, Z., *A Flexible New Technique for Camera Calibration*, Technical report, Microsoft Research, Redmond, WA, USA, 2008.
- [37] RUSU, R. B., COUSINS, S., *3D is here: Point Cloud Library (PCL)*. IEEE International Conference on Robotics and Automation (ICRA), 2011.

- [38] KAZHDAN, M., HOPPE, H., “Screened Poisson Surface Reconstruction”, *ACM Transactions on Graphics (ToG)*, v. 32, n. 3, pp. 29:1–29:13, 2013.
- [39] DELAUNAY, B., “Sur la sphère vide. A la mémoire de Georges Voronoï”, *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk*, v. 7, pp. 793–800, 1934.
- [40] SZELISKI, R., *Computer Vision: Algorithms and Applications*. 2 ed. Springer, 2022.
- [41] CONTRIBUTORS, W., “Quaternions and spatial rotation”, https://en.wikipedia.org/wiki/Quaternions_and_spatial_rotation, 2025, (Acesso em 6 Fevereiro 2025).
- [42] SILVA, R. T., FERREIRA, J. P., SANTOS, L. M., “Uso de Veículos Aéreos Não Tripulados para Monitoramento Ambiental em Regiões de Difícil Acesso”, *Revista Brasileira de Geotecnologias*, v. 12, pp. 55–72, 2020.
- [43] OLIVEIRA, P. M., SANTOS, R. L., ALMEIDA, F. J., “FoCor: Utilização de Veículos Aéreos Autônomos no Apoio a Fogos Controlados e Detecção de Reacendimentos na Gestão de Incêndios Florestais”, *Anais do Congresso de Engenharia Florestal*, pp. 150–162, 2023.