



# ESCOLA NAVAL



*ta sainte & bi-faire*

Miguel Miranda Bernardes

**Geolocalização de alvos marítimos a partir de  
uma câmara RGB montada num UAV**

Dissertação para obtenção do Grau de Mestre em  
Ciências Militares Navais, na especialidade de Engenharia  
Naval Ramo de Armas e Eletrónica



Alfeite

2025





# ESCOLA NAVAL

*talant de bi-faire*



Miguel Miranda Bernardes

*Geolocalização de alvos marítimos a partir de uma câmara  
RGB montada num UAV*

Dissertação para obtenção do Grau de Mestre em  
Ciências Militares Navais, na especialidade de Engenharia Naval Ramo  
de Armas e Eletrónica

**Orientação de:** Bruno Duarte Damas

**Coorientação de:** Nuno Alexandre Antunes Martins Pessanha Santos

*O Aluno Mestrando,*

*O Orientador,*

---

Miguel Bernardes

---

Bruno Damas

Alfeite

2025



Dedico este trabalho a todos os que me acompanharam ao longo destes 6 anos de  
Escola Naval.



# Agradecimentos

Em primeiro lugar, gostaria de expressar a minha gratidão ao orientador, Professor Bruno Damas, e ao coorientador, Professor Nuno Santos, pelo acompanhamento e ensinamentos transmitidos ao longo desta etapa. A sua contribuição foi fundamental para o meu crescimento pessoal e profissional.

Da mesma forma, gostaria de expressar a mais sincera gratidão à Célula de Inovação e Experimentação Operacional de Sistemas Não Tripulados (CEOV) pela sua total disponibilidade, que foi fundamental para a realização deste trabalho.

Gostaria de expressar o meu profundo agradecimento a toda a minha família, em particular à minha mãe, pai e irmã, pelo apoio incondicional que me foi concedido ao longo de todo o percurso. Sem eles, seria tudo mais difícil.

À minha namorada, o meu mais sentido agradecimento, pela compreensão e pela força que transmitiu. Foi uma peça fundamental para atingir este grande objetivo.

Ao curso “Fernão de Magalhães”, em especial aos Engenheiros, pela camaradagem e entreaajuda que demonstraram ao longo do curso. Foram um pilar importante neste percurso que é a Escola Naval.



# Resumo

Os Veículos Aéreos Não Tripulados (VANTs) têm potencial para se tornarem uma ferramenta valiosa na vigilância marítima, especialmente em cenários onde os navios não utilizam o Sistema de Identificação Automática. Esta dissertação aborda a geolocalização de alvos marítimos a partir de imagens captadas por uma câmara RGB (*Red, Green and Blue*) montada num VANT, utilizando dados de posicionamento (*Global Positioning System, GPS*) e orientação (*Inertial Measurement Unit, IMU*).

A metodologia desenvolvida visa detetar e geolocalizar navios recorrendo a um VANT, contribuindo para o esclarecimento do panorama situacional sem a mobilização de meios adicionais. A abordagem incluiu a calibração da câmara para determinar parâmetros intrínsecos e extrínsecos, a deteção de navios utilizando redes neuronais convolucionais *You Only Look Once* (YOLO) versão 8 com a arquitetura *nano* e a geolocalização direta baseada em transformações de coordenadas, utilizando informações da pose do VANT e da orientação da câmara, com aplicação do Filtro de Kalman Estendido (FKE) para reduzir o erro da estimativa ao longo do tempo.

Um contributo significativo foi a criação de um conjunto de dados original, composto por um vídeo captado por VANT, com dados de pose, dados de transformação de coordenadas do mundo para a câmara, anotações do vídeo e as posições dos navios detetados.

Os resultados experimentais, obtidos em testes reais, demonstraram uma deteção eficaz, alcançando um F1 Score de 0.66. A geolocalização revelou erros médios de 100 m a 1200 m, para alvos próximos (2 km) e distantes (5 km), respetivamente, com o FKE a melhorar a consistência em alvos distantes. Limitações como ruído sensorial, assincronismo temporal e proximidade da deteção ao horizonte impactaram a precisão. A metodologia revelou-se viável para vigilância marítima, mas requer mais testes para alvos a maior distância.

**Palavras-chave:** Veículo Aéreo Não Tripulado, Câmara RGB, Navios, Deteção, Geolocalização



# Abstract

Unmanned Aerial Vehicles (UAVs) have the potential to become a valuable tool in maritime surveillance, particularly in scenarios where ships do not use the Automatic Identification System (AIS). This dissertation addresses the geolocation of maritime targets from images captured by an RGB (Red, Green, and Blue) camera mounted on a UAV, using positioning data (Global Positioning System, GPS) and orientation data (Inertial Measurement Unit, IMU).

The developed methodology aims to detect and geolocate ships using a UAV, contributing to situational awareness without the need for additional resources. The approach included camera calibration to determine intrinsic and extrinsic parameters, ship detection using the You Only Look Once (YOLO) version 8 convolutional neural network with the nano architecture, and direct geolocation based on coordinate transformations, utilizing the UAV's pose and camera orientation data, with the application of the Extended Kalman Filter (EKF) to reduce estimation errors over time.

A significant contribution was the creation of an original dataset, consisting of a video captured by a UAV, including pose data, world-to-camera coordinate transformation data, video annotations, and the positions of detected ships.

Experimental results, obtained from real-world tests, demonstrated effective detection, achieving an F1 Score of 0.66. Geolocation showed average errors of 100 m to 1200 m for nearby (2 km) and distant (5 km) targets, respectively, with the EKF improving consistency for distant targets. Limitations such as sensor noise, temporal asynchrony, and detection proximity to the horizon impacted accuracy. The methodology proved viable for maritime surveillance but requires further testing for targets at greater distances.

**Keywords:** Unmanned Aerial Vehicles, RGB Camera, Ships, Detection, Geolocation



# Declaração

Declaro que, para a realização desta dissertação, foram utilizadas ferramentas de inteligência artificial (*ChatGPT* e *Grok*) para gerar equações e imagens no formato  $\text{\LaTeX}$  para auxílio na programação necessária. Para correções e estrutura frásica, foi utilizada a ferramenta *DeepL Write* no estilo “Acadêmico”.



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Formulação do problema . . . . .	2
1.3	Objetivos . . . . .	2
1.4	Desafios . . . . .	3
1.5	Organização da dissertação . . . . .	4
<b>2</b>	<b>Enquadramento teórico</b>	<b>5</b>
2.1	Modelo da câmara . . . . .	5
2.1.1	Modelo da câmara . . . . .	5
2.1.2	Parâmetros da câmara . . . . .	7
2.2	Sensores do UAV . . . . .	8
2.3	Deteção de objetos . . . . .	8
2.3.1	Deteção em duas fases . . . . .	9
2.3.2	Deteção numa fase . . . . .	11
2.4	Geolocalização . . . . .	12
2.4.1	Transformação de coordenadas espaciais . . . . .	12
2.5	Sistema de Identificação Automática . . . . .	13
2.6	Fusão sensorial . . . . .	14
2.6.1	Filtro de Kalman . . . . .	14
2.6.2	Outros métodos de fusão . . . . .	17
<b>3</b>	<b>Estado da Arte</b>	<b>19</b>
3.1	<i>Dataset</i> . . . . .	19
3.2	Deteção de navios . . . . .	22
3.3	Geolocalização de alvos marítimos . . . . .	23
<b>4</b>	<b>Metodologia</b>	<b>25</b>
4.1	Arquitetura geral . . . . .	25
4.2	Referenciais . . . . .	26

4.3	Descrição dos equipamentos utilizados . . . . .	28
4.4	Calibração da câmara . . . . .	29
4.5	Deteção de navios . . . . .	30
4.5.1	Deteção de navios . . . . .	30
4.5.2	<i>Datasets</i> de treino . . . . .	32
4.6	Geolocalização . . . . .	33
4.7	Aplicação do filtro de Kalman . . . . .	37
<b>5</b>	<b>Resultados experimentais</b>	<b>41</b>
5.1	Calibração da câmara . . . . .	41
5.2	Aquisição de dados para geolocalização . . . . .	42
5.3	Criação de <i>dataset</i> para geolocalização . . . . .	44
5.4	Deteção de navios . . . . .	45
5.5	Geolocalização . . . . .	48
5.5.1	Cinemática inversa . . . . .	48
5.5.2	Simulação . . . . .	49
5.5.3	Aplicação a dados reais . . . . .	53
5.6	Discussão de resultados e dificuldades encontradas . . . . .	58
5.6.1	Calibração da câmara . . . . .	58
5.6.2	<i>Datasets</i> e deteção . . . . .	59
5.6.3	Geolocalização . . . . .	59
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>63</b>
6.1	Conclusões . . . . .	63
6.2	Trabalho Futuro . . . . .	64
	<b>Bibliografia</b>	<b>65</b>
	<b>Apêndices</b>	<b>71</b>
<b>A</b>	<b>Transformação de coordenadas espaciais</b>	<b>71</b>
<b>B</b>	<b>Determinação das matrizes <math>F</math>, <math>Q</math>, <math>h</math> e <math>H</math></b>	<b>75</b>
<b>C</b>	<b>Definição de RMSE e MAE</b>	<b>77</b>

# Lista de Figuras

2.1	Modelo de câmara pinhole. . . . .	6
2.2	Eixos de coordenadas. . . . .	6
2.3	Parâmetros intrínsecos. . . . .	7
2.4	Formas de reconhecimento . . . . .	9
2.5	Deteção em duas fases. . . . .	10
2.6	Deteção em uma fase. . . . .	11
2.7	Comunicações AIS. . . . .	13
3.1	Exemplos de imagens do <i>MarSyn Dataset</i> . . . . .	20
3.2	Exemplos de imagens do <i>Seagull</i> . . . . .	20
3.3	Exemplo de imagens do VESSELimg . . . . .	21
3.4	Exemplo de imagens do <i>ShipDataset</i> . . . . .	21
4.1	<i>Framework</i> adotada para a estimativa da latitude e longitude de um alvo. . . . .	26
4.2	Referenciais do mundo, UAV, câmara e imagem. . . . .	27
4.3	UAV utilizado no voo. . . . .	28
4.4	Câmara <i>SIYI ZR10</i> . . . . .	29
4.5	Tabuleiro de xadrez para calibração da câmara. . . . .	30
4.6	Exemplo de <i>data augmentation</i> no <i>Aerial Vessels Detection Dataset</i> . . . . .	31
4.7	Deteção numa imagem do <i>Aerial Vessels Detection Dataset</i> . . . . .	32
4.8	Imagens pertencentes ao <i>Aerial Vessels Detection Dataset</i> . . . . .	33
4.9	Imagens pertencentes a <i>ship aeriels 2 Computer Vision Dataset</i> . . . . .	34
4.10	Imagens pertencentes a <i>Aerial views of boats Computer Vision Model</i> . . . . .	34
5.1	Imagens utilizadas para calibração. . . . .	41
5.2	Ilustração de voo realizado com a CEOV. . . . .	42
5.3	Imagem capturada durante recolha de dados. . . . .	43
5.4	Exemplo da <i>Ground truth</i> de um frame. . . . .	45
5.5	Imagens da comparação entre <i>datasets</i> A e A+B. . . . .	46
5.6	Imagens da comparação entre sem e com <i>data augmentation</i> . . . . .	47
5.7	Imagens da comparação entre o <i>yolov8 nano</i> e <i>small</i> . . . . .	48

5.8	Modelo <i>Simulink</i> da simulação . . . . .	49
5.9	Simulação em <i>Simulink</i> . . . . .	50
5.10	Simulação da cinemática inversa. . . . .	51
5.11	Simulação da cinemática inversa e EKF. . . . .	51
5.12	Resultado da geolocalização em <i>Simulink</i> . . . . .	52
5.13	Simulação com ruído da cinemática inversa e EKF. . . . .	52
5.14	Simulação com ruído da cinemática inversa vs EKF com limitação do erro. . . . .	53
5.15	Geolocalização com a <i>ground truth</i> das deteções. . . . .	54
5.16	Posição real dos navios. . . . .	54
5.17	Erro da geolocalização pelo método da cinemática inversa referente a cada navio. . . . .	55
5.18	Geolocalização com EKF. . . . .	56
5.19	Erro da geolocalização pelo método do EKF referente a cada navio. . . . .	56
5.20	Geolocalização com navios detetados. . . . .	57
5.21	Geolocalização com navios detetados aproximada à área de interesse. . . . .	58
5.22	Centro das <i>Bounding Box</i> . . . . .	60

# Lista de Tabelas

4.1	Distribuição do <i>Dataset</i> . . . . .	33
5.1	Exemplo de dados de posição e orientação do UAV. . . . .	43
5.2	Exemplo de dados de orientação da câmara. . . . .	43
5.3	Exemplo de dados AIS obtidos durante testes práticos. . . . .	44
5.4	Comparação entre <i>datasets</i> A e A+B. . . . .	46
5.5	Comparação entre com e sem <i>data augmentation</i> . . . . .	46
5.6	Comparação entre o <i>yolov8 nano</i> e <i>small</i> . . . . .	47
5.7	Erros de geolocalização da simulação. . . . .	53
5.8	Erros da geolocalização referente a cada navio. . . . .	55
5.9	Erros da geolocalização com EKF referente a cada navio. . . . .	57



# Lista de Abreviaturas

<b>AIS</b>	<b>A</b> utomatic <b>I</b> dentification <b>S</b> ystem
<b>BB</b>	<b>B</b> ounding <b>B</b> ox
<b>CEOV</b>	<b>C</b> élula de <b>I</b> novação e <b>E</b> xperimentação <b>O</b> peracional de <b>S</b> istemas <b>N</b> ão <b>T</b> ripulados
<b>CNN</b>	<b>C</b> onvolutional <b>N</b> eural <b>N</b> etwork
<b>CVAT</b>	<b>C</b> omputer <b>V</b> ision <b>A</b> notation <b>T</b> ool
<b>ECEF</b>	<b>E</b> arth- <b>C</b> enter <b>E</b> arth- <b>F</b> ixed
<b>EKF</b>	<b>E</b> xtended <b>K</b> alman <b>F</b> ilter
<b>ENU</b>	<b>E</b> ast- <b>N</b> orth- <b>U</b> p
<b>FKE</b>	<b>F</b> iltro de <b>K</b> alman <b>E</b> stendido
<b>GNSS</b>	<b>G</b> lobal <b>N</b> avigation <b>S</b> atellite <b>S</b> ystem
<b>GPS</b>	<b>G</b> lobal <b>P</b> ositioning <b>S</b> ystem
<b>IMU</b>	<b>I</b> nertial <b>M</b> easurement <b>U</b> nit
<b>IoU</b>	<b>I</b> ntersection <b>o</b> ver <b>U</b> nion
<b>KF</b>	<b>K</b> alman <b>F</b> ilter
<b>MAE</b>	<b>M</b> ean <b>A</b> bsolute <b>E</b> rror
<b>MATLAB</b>	<b>M</b> atrix <b>L</b> aboratory
<b>NED</b>	<b>N</b> orth- <b>E</b> ast- <b>D</b> own
<b>OpenCV</b>	<b>O</b> pen <b>S</b> ource <b>C</b> omputer <b>V</b> ision
<b>PF</b>	<b>P</b> article <b>F</b> ilter
<b>R-CNN</b>	<b>R</b> egion-based <b>C</b> onvolutional <b>N</b> eural <b>N</b> etworks
<b>RGB</b>	<b>R</b> ed, <b>G</b> reen and <b>B</b> lue
<b>RMSE</b>	<b>R</b> oot <b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>RPN</b>	<b>R</b> egion <b>P</b> roposal <b>N</b> etwork
<b>SAR</b>	<b>S</b> earch <b>A</b> nd <b>R</b> escue
<b>SSD</b>	<b>S</b> ingle <b>S</b> hot <b>M</b> ulti <b>B</b> ox <b>D</b> etector
<b>UAV</b>	<b>U</b> nmanned <b>A</b> erial <b>V</b> ehicle
<b>UKF</b>	<b>U</b> nscented <b>K</b> alman <b>F</b> ilter
<b>VANT</b>	<b>V</b> eículo <b>A</b> éreo <b>N</b> ão <b>T</b> ripulado
<b>VHF</b>	<b>V</b> ery <b>H</b> igh <b>F</b> requency
<b>YOLO</b>	<b>Y</b> ou <b>O</b> nly <b>L</b> ook <b>O</b> nce
<b>ZEE</b>	<b>Z</b> ona <b>E</b> conómica <b>E</b> xclusiva



# Capítulo 1

## Introdução

### 1.1 Motivação

A vigilância e o controlo marítimo constituem missões de crescente importância para a Marinha Portuguesa, sobretudo perante os desafios associados à vasta área marítima sob jurisdição nacional (Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos, 2025). A extensão deste território requer soluções tecnologicamente avançadas que assegurem uma cobertura eficiente e em tempo real.

Neste contexto, os veículos autónomos não tripulados (*Unmanned Aerial Vehicle*, UAV) têm-se afirmado como uma ferramenta promissora em missões de vigilância marítima e de busca e salvamento (*Search and Rescue*, SAR). A agilidade, flexibilidade e acessibilidade, tornam-os especialmente adequados para operar em ambientes marítimos complexos, demonstrando resultados cada vez mais eficazes (Lyu et al., 2023). Para aumentar a eficácia destas missões, é essencial não só detetar os alvos de interesse com elevada fiabilidade, mas também determinar com precisão a sua localização geográfica.

Esta capacidade assume particular relevância em cenários como a pesca ilegal, o tráfico de estupefacientes ou outras atividades marítimas clandestinas, nas quais os alvos frequentemente operam sem o sistema AIS (*Automatic Identification System*) ativo (Kurekin et al., 2019). Adicionalmente, em operações de busca e salvamento, a precisão da geolocalização pode ser determinante para o sucesso da missão, onde um UAV, ao detetar um alvo, indica de imediato a respetiva posição.

A opção por um UAV como plataforma base é justificada pela sua capacidade de operação autónoma e prolongada, perfil de voo a grandes altitudes, reduzida probabilidade de deteção e flexibilidade no planeamento de rotas e zonas de voo. Estas características conferem aos UAV uma eficácia ímpar para além da cobertura proporcionada pelos radares costeiros, atualmente disponíveis em Portugal.

Para além disso, apesar do investimento inicial associado à sua aquisição, os UAV representam uma solução significativamente menos dispendiosa do que a utilização de navios da Marinha Portuguesa em patrulhas de longa duração, sobretudo tendo em conta a vasta Zona Económica Exclusiva (ZEE) nacional.

## 1.2 Formulação do problema

A presente dissertação emerge com o propósito de mitigar a ausência de informação acerca dos alvos marítimos em tempo real. O conhecimento da localização dos alvos constitui um elemento necessário para o esclarecimento do panorama situacional marítimo.

Para tal, é imperativo que o UAV esteja equipado com um sistema de navegação por satélite (*Global Navigation Satellite System*, GNSS), mais concretamente o *Global Positioning System* (GPS), para a aquisição da posição, uma unidade de medida inercial (*Inertial Measurement Unit*, IMU) para a obtenção da atitude e uma câmara *Red, Green and Blue* (RGB), para a captura de imagens.

Com base nos dados recolhidos, é possível estimar a posição geográfica dos alvos marítimos, uma vez que se assume que todos os alvos estão a uma altitude de zero metros. A estimativa é composta por duas etapas principais: a deteção dos alvos nas imagens captadas pelo UAV, através do treino de um detetor de navios baseado numa arquitetura de rede neuronal e a determinação da posição dos alvos no sistema de coordenadas geográficas, com base na posição e orientação do UAV.

## 1.3 Objetivos

O objetivo principal desta dissertação consiste no desenvolvimento de uma metodologia para a geolocalização de alvos marítimos (navios), com base em imagens aéreas captadas por um UAV equipado com uma câmara RGB, GPS e IMU. Esta metodologia visa contribuir para a obtenção da localização geográfica (latitude e longitude) de alvos marítimos em tempo real, facilitando o esclarecimento do panorama situacional sem a mobilização de meios humanos, aéreos ou navais.

A deteção de alvos constitui o primeiro procedimento a ser realizado, sendo imperativo para evitar falsas deteções ou alvos não detetados. A implementação do procedimento em questão será efetuada através do desenvolvimento de um sistema de deteção automática de navios a partir de imagens captadas pelo UAV, utilizando redes neuronais para localizar a posição dos alvos na imagem.

Relativamente à geolocalização, que constitui a base da presente dissertação, o objetivo reside na criação de um algoritmo de geolocalização que, com base na posição e orientação (pose) estimada do UAV, determine as coordenadas geográficas dos navios detetados. A determinação da geolocalização dos alvos será auxiliada por técnicas de filtragem temporal.

Por fim, a forma ideal de validação dos resultados é efetuada através dos dados do Sistema de Identificação Automática (*Automatic Identification System*, AIS), que detém a localização dos navios. Ademais, é igualmente plausível a validação dos resultados com o navio atracado, ainda que se trate de um cenário menos realista, mas igualmente válido.

## 1.4 Desafios

O desenvolvimento desta dissertação enfrenta diversos desafios técnicos e práticos. Em primeiro lugar, a deteção precisa de alvos marítimos revela-se complexa devido à presença de reflexos e brilhos indesejados na superfície oceânica, que podem comprometer a qualidade das imagens capturadas. Além disso, a existência de pequenas cristas de espuma formadas pela rebentação das ondas, podem resultar em falsas deteções, dificultando a identificação correta dos navios.

Outro desafio consiste na identificação de um *dataset* que integre simultaneamente imagens aéreas, informações de pose do UAV, parâmetros da câmara e a localização geográfica dos navios. Esta necessidade torna imperativa a criação de um conjunto de dados específico e adequado aos objetivos do presente estudo.

No que diz respeito ao sistema de câmara, a estimação incorreta dos seus parâmetros pode introduzir erros sistemáticos na geometria da imagem, afetando negativamente a precisão da geolocalização. A análise e correção destes desvios representam uma tarefa crucial para o sucesso da metodologia.

Por fim, a sincronização temporal entre os diferentes sensores a bordo do UAV, nomeadamente a câmara, o GPS e a IMU, é fundamental para garantir que os dados recolhidos correspondem ao mesmo instante. Uma sincronização inadequada pode resultar em desfasamentos que comprometam a estimativa da geolocalização dos alvos.

## 1.5 Organização da dissertação

A presente dissertação encontra-se organizada em diversos capítulos, sendo que cada um deles contribui para uma análise detalhada dos elementos-chave, conforme descrito:

- **Capítulo 2 - Enquadramento teórico:** O presente capítulo aborda os principais temas que se revelam fundamentais para a compreensão dos objetos de estudo, nomeadamente o modelo da câmara, sensores do UAV, deteção de objetos, geolocalização, sistema de identificação automática e fusão sensorial;
- **Capítulo 3 - Estado da arte:** O capítulo descreve o estado da arte das temáticas em estudo, que consistem em *dataset*, deteção de navios e geolocalização;
- **Capítulo 4 - Metodologias:** Este capítulo descreve a metodologia utilizada, descreve os referenciais utilizados, a calibração da câmara, a deteção de navios, a geolocalização e a aplicação do Filtro de Kalman;
- **Capítulo 5 - Resultados experimentais:** O capítulo exhibe os resultados obtidos através dos métodos utilizados para a deteção e para a geolocalização, os testes foram conduzidos em ambiente simulado e real;
- **Capítulo 6 - Conclusões:** O capítulo final aborda as possíveis discrepâncias entre os métodos utilizados e propõe direções para potenciais trabalhos futuros no âmbito do tema em estudo.

# Capítulo 2

## Enquadramento teórico

Para que seja possível atingir o objetivo desta dissertação, que consiste na geolocalização de alvos marítimos por meio de uma câmara RGB montada num UAV, é necessário explorar conceitos-chave previamente desenvolvidos, nomeadamente o modelo da câmara, os sensores do UAV, a deteção de objetos, a geolocalização, a informação AIS e a fusão sensorial.

### 2.1 Modelo da câmara

A principal função de uma câmara assenta na projeção de um ambiente tridimensional (3D) numa imagem bidimensional (2D). O comportamento da câmara é geralmente modelado através do modelo pinhole, ao qual se podem adicionar os efeitos das lentes e os resultados de uma calibração exata, permitindo descrever matematicamente o processo de formação da imagem.

#### 2.1.1 Modelo da câmara

O modelo *pinhole* pode ser concebido através da colocação de uma barreira com um pequeno orifício entre o ambiente 3D e um material fotossensível, como ilustrado na Figura 2.1. O princípio subjacente ao seu funcionamento baseia-se numa única passagem de cada feixe de luz através de um pequeno orifício, eliminando a influência dos outros feixes (Hata e Savarese, 2017).

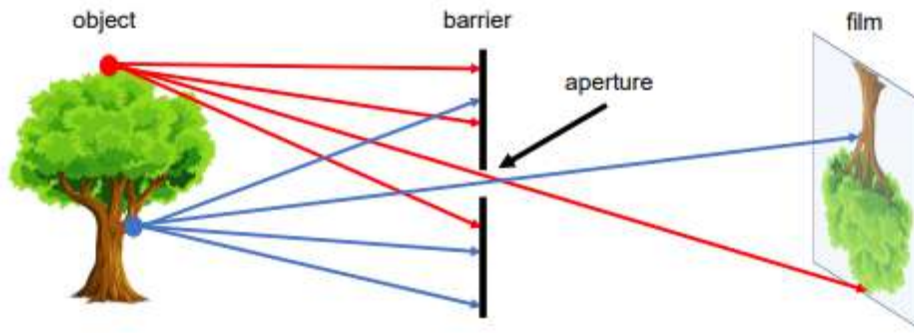


FIGURA 2.1: Modelo de câmera pinhole (Hata e Savarese, 2017).

Na Figura 2.2 são apresentados os principais eixos de coordenadas, sendo que os vetores  $(u, v)$  representam as coordenadas da imagem, tipicamente com origem no canto superior esquerdo, com  $u$  a apontar para a direita e  $v$  a apontar para baixo. As coordenadas da câmera são designadas pela letra “ $C$ ”, onde  $X_C, Y_C$  têm a mesma orientação de  $u, v$ , respetivamente, e  $Z_C$  no sentido da lente da câmera. As coordenadas designadas pela letra “ $W$ ” correspondem ao sistema de coordenadas do mundo. Os eixos são definidos de acordo com a aplicação.

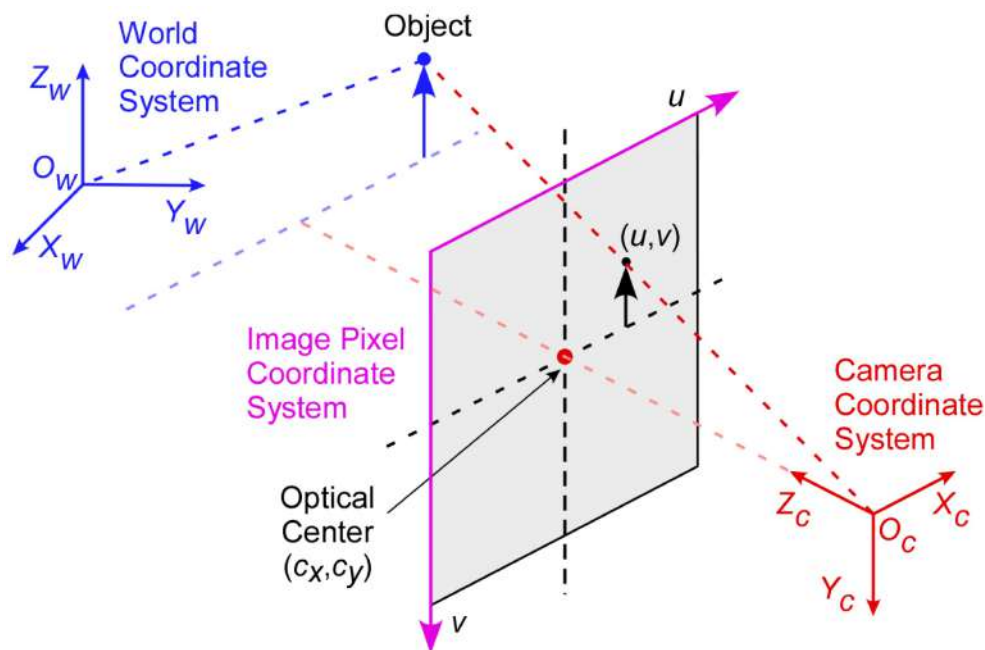


FIGURA 2.2: Eixos de coordenadas (Li e Yoon, 2023).

De modo geral, é utilizado o referencial Este-Norte-Cima (*East-North-Up*, ENU), que define o eixo  $x$  direcionado para este, o eixo  $y$  para norte e o eixo  $z$  para cima. Alternativamente, pode ser utilizado o referencial Norte-Este-Baixo (*North-East-Down*, NED) correspondendo aos eixos  $x$ ,  $y$  e  $z$ , respetivamente.

### 2.1.2 Parâmetros da câmara

As características da câmara são definidas pela matriz dos parâmetros intrínsecos,  $K$ , que dependem da distância focal,  $f=(f_x, f_y)$  e do centro ótico,  $C'=(c_x, c_y)$ , e pela distorção da lente. A origem do referencial,  $O$ , define o centro da câmara, conforme ilustrado na Figura 2.3.

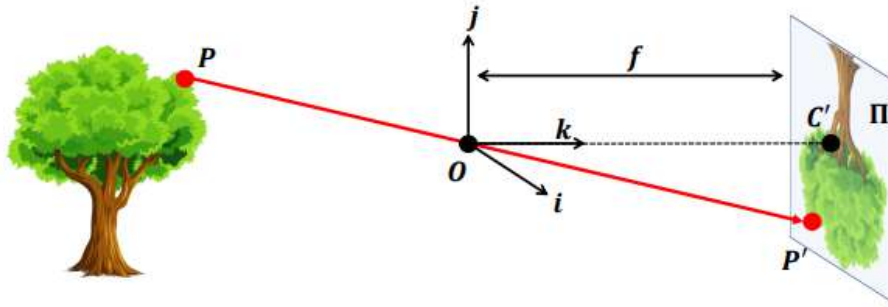


FIGURA 2.3: Parâmetros intrínsecos (Hata e Savarese, 2017).

A distância focal corresponde à distância, em milímetros, entre o plano focal e o centro ótico da lente, quando esta está focada no infinito. O centro ótico representa a posição central em relação ao plano da imagem. Quanto aos coeficientes de distorção, estes medem o grau de distorção das imagens obtidas, tendo em conta o modelo e as lentes utilizadas.

Os coeficientes de distorção podem ser representados pela matriz

$$d = \begin{bmatrix} k_1 & k_2 & p_1 & p_2 & k_3 \end{bmatrix}. \quad (2.1)$$

Estes são parâmetros necessários para, posteriormente, determinar a distorção radial através de:

$$\begin{aligned} x_{\text{radial}} &= x_C(1 + k_1r^2 + k_2r^4 + k_3r^6) \\ y_{\text{radial}} &= y_C(1 + k_1r^2 + k_2r^4 + k_3r^6) \end{aligned} \quad (2.2)$$

e a distorção tangencial através de:

$$\begin{aligned} x_{\text{tangencial}} &= x_C + (2 + p_1r^2y_C + p_2(r^2 + 2x_C^2)) \\ y_{\text{tangencial}} &= y_C + (p_1(r^2 + 2y_C^2) + 2p_2x_Cy_C) \end{aligned} \quad (2.3)$$

Contudo, não se restringe aos parâmetros intrínsecos, sendo igualmente necessário considerar os parâmetros extrínsecos,  ${}^cT_w$ , que correspondem a uma matriz de rotação e translação, de forma a colocar a câmara nas coordenadas do mundo.

Como demonstrado na Equação 2.4, as coordenadas no plano da imagem,  $P_I$ , são definidas pela multiplicação das matrizes dos parâmetros intrínsecos, extrínsecos e pelo ponto no referencial do mundo  $P_w$ . As coordenadas dos pontos no referencial da câmara,  $P_C$ , são definidas por  ${}^cT_w P_w$ , que determinam a posição e orientação da câmara, convertendo os pontos do sistema de coordenadas do mundo para o sistema de coordenadas da imagem.

$$P_I = K {}^cT_w P_w \Leftrightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.4)$$

## 2.2 Sensores do UAV

No caso dos UAV, a telemetria agrega o GPS, que fornece: latitude, longitude e altitude e a IMU, que calcula: velocidade, aceleração, rotação, inclinação e guinada (Sévigny et al., 2021).

A IMU constitui uma das unidades mais relevantes para a navegação inercial de sistemas não tripulados, sendo esta baseada em microcontroladores e sensores. Possui seis graus de liberdade, consistindo estes em três acelerómetros e três giroscópios ortogonais entre si. Estes sensores medem a aceleração em três dimensões e as taxas de rotação em torno dos três eixos ortogonais. Deste modo, é possível determinar as variáveis acima referidas (L. e Jarrah, 2008).

## 2.3 Detecção de objetos

Existem diferentes tipos de reconhecimento: a classificação de objetos, que apenas indica quais os objetos presentes na imagem, a deteção de objetos, que cria uma caixa delimitadora (*Bounding Box*, BB) em torno dos objetos e os classifica, a segmentação semântica, que deteta a forma de todos os elementos da imagem e os classifica, e a segmentação de instâncias de objetos, que deteta a forma dos objetos da imagem com a respetiva classificação, conforme ilustrado na Fig.2.4.

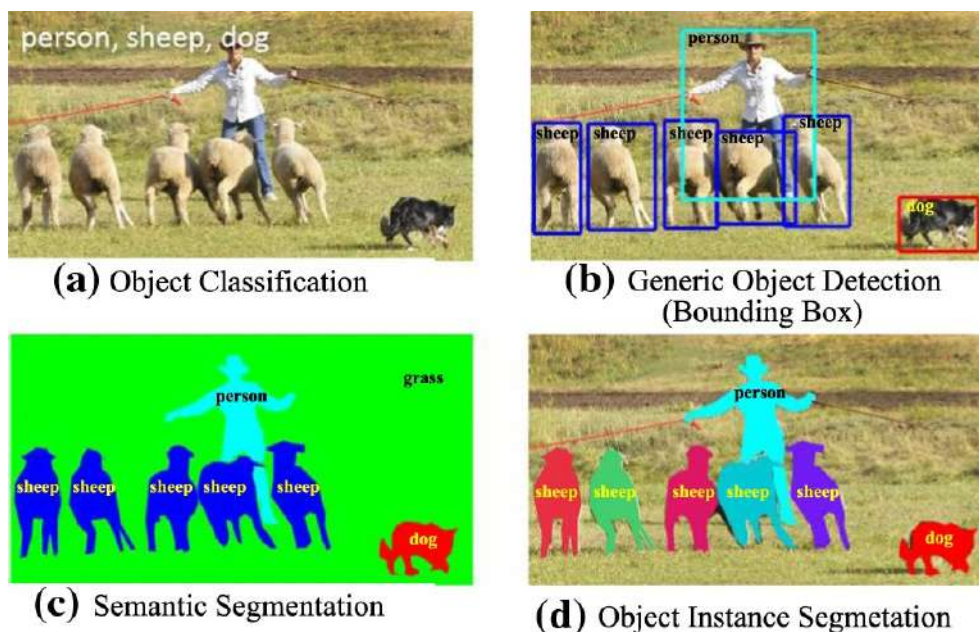


FIGURA 2.4: Formas de reconhecimento: a) Classificação, b) Detecção, c) Segmentação e d) Segmentação de instâncias (L. Liu et al., 2020).

Em virtude da capacidade de representação das redes neurais convolucionais (*Convolutional Neural Network*, CNN), a deteção de objetos através de *deep learning* tem adquirido uma relevância significativa no âmbito da investigação, com resultados bastante positivos (X. Yang et al., 2022).

A deteção de navios constitui uma área de estudo em expansão, existindo diversos métodos para o efeito. Segundo o estudo de Z. Q. Zhao et al. (2019), são apresentados dois métodos de deteção de objetos. A deteção em duas fases consiste na aquisição de zonas delimitadas e na subsequente classificação nas diferentes categorias de objetos. A deteção numa fase trata como um problema de regressão ou classificação, de modo a obter a categoria e a localização dos objetos numa única fase.

### 2.3.1 Deteção em duas fases

Os detetores de duas fases encontram-se divididos em duas etapas: a primeira etapa, que compreende uma proposta de deteção que consiste no reconhecimento da região de interesse (*Region of Interest*, RoI) e geração das BBs e a segunda etapa que se centra na classificação dos objetos das regiões definidas e, por conseguinte, no refinamento das localizações, de acordo com o ilustrado na Figura 2.5 (Jiao et al., 2019).

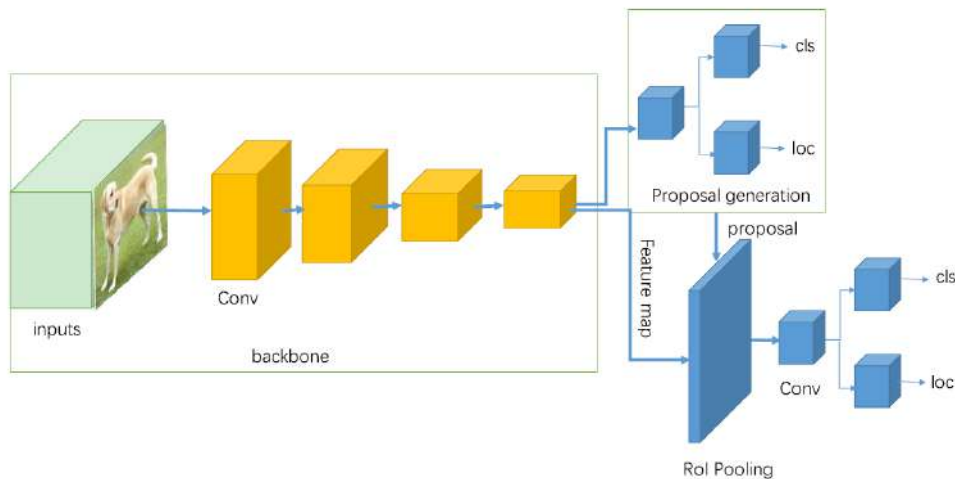


FIGURA 2.5: Detecção em duas fases (Jiao et al., 2019).

Esta arquitetura de rede proporciona uma maior precisão na detecção, graças à sua análise mais detalhada dos objetos, especialmente em cenários mais complexos.

Relativamente ao *Region-based Convolutional Neural Networks* (R-CNN), este começa por seleccionar propostas de BB, de modo a analisar cada uma num modelo pré-treinado (Krizhevsky et al., 2012). A detecção final do objeto é alcançada através da passagem de cada região por um classificador. Apesar da sua elevada precisão, o principal inconveniente do R-CNN reside no elevado tempo de processamento necessário para tratar todas as propostas seleccionadas, uma vez que cada região é processada de forma independente, resultando em redundâncias computacionais (Girshick et al., 2016). Essa limitação motivou o desenvolvimento de arquiteturas mais eficientes, que buscam otimizar o processo de detecção sem comprometer a precisão.

Nesse contexto, emergiram as redes totalmente convolucionais baseadas em regiões (*Region-based Fully Convolutional Networks*, R-FCN), que representam uma evolução significativa em relação ao R-CNN. Diferentemente do R-CNN, que processa cada proposta de região de forma independente, o R-FCN utiliza uma rede totalmente convolucional partilhada para extrair características de toda a imagem, reduzindo redundâncias computacionais e acelerando o processo de detecção. Essa rede é dividida em duas componentes principais: uma rede totalmente convolucional que gera mapas de características compartilhados para toda a imagem e uma rede específica para regiões de interesse (RoI) que classifica e refina as BBs sem compartilhar camadas computacionais adicionais (Z. Q. Zhao et al., 2019). Essa abordagem permite que o R-FCN combine a precisão do R-CNN com uma maior eficiência computacional.

### 2.3.2 Detecção numa fase

De acordo com X. Yang et al. (2022), a principal distinção entre a deteção numa fase e a deteção em duas fases assenta no facto de que esta última realiza em simultâneo a classificação e a regressão dos objetos, como se pode observar na Figura 2.6, tornando este método mais rápido que o anterior. De modo adicional, estes alcançam uma alta velocidade de interferência, o que os torna adequados para aplicações em tempo real.

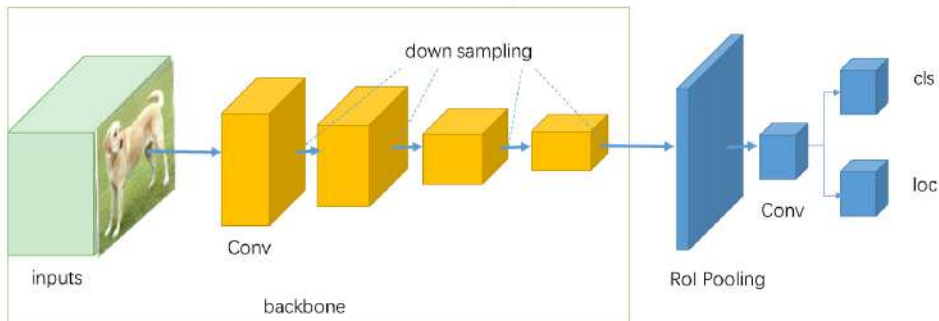


FIGURA 2.6: Deteção em uma fase (Jiao et al., 2019).

O algoritmo *You Only Look Once* (YOLO) apresenta uma arquitetura numa fase e aborda a deteção como um problema de regressão e classificação, no qual a rede aprende a mapear diretamente as BB (Jiang et al., 2021). A sua arquitetura consiste na divisão de cada imagem numa grade e na associação de cada célula ao objeto a detetar, tornando o processo mais eficiente. Embora versões iniciais do YOLO apresentem desempenho limitado na deteção de objetos pequenos devido à baixa resolução da grade e à perda de detalhes finos, versões mais recentes, como YOLOv8, introduziram melhorias, como resoluções de entrada mais altas e fusão de *features* multi-escala, permitindo maior eficácia na deteção de objetos de diferentes dimensões, incluindo os de menor tamanho.

Adicionalmente, destaca-se o detetor de várias BB de disparo único (*Single Shot MultiBox Detector*, SSD), que representa um bom compromisso entre velocidade e precisão (Zou et al., 2019). Comparativamente ao YOLO, o SSD utiliza mapas de características de múltiplas camadas convolucionais, permitindo a identificação de objetos em diferentes escalas, incluindo os de menores dimensões (W. Liu et al., 2016).

## 2.4 Geolocalização

A geolocalização consiste no processo de atribuir coordenadas geográficas, no mundo real, a determinados objetos ou elementos visíveis na imagem. Conforme referido por Helgesen et al. (2019) a ausência de informação sobre a distância ao alvo pretendido constitui um desafio significativo para a geolocalização. O objetivo principal assenta na obtenção das coordenadas de um objeto relativamente ao sistema de referência da Terra, com base nos píxeis da câmara. Em tempo real, acresce o desafio de o realizar apenas com base numa imagem, sem quaisquer pontos de referência no solo. Este processo baseia-se na pose da câmara (latitude, longitude, altitude e orientação) no momento exato em que a imagem é capturada. Por conseguinte, qualquer erro ou desvio na leitura da posição da câmara origina um erro na localização do alvo.

A geolocalização pode ser efetuada através de dois métodos: indireta e direta. A indireta é um método mais intuitivo, no qual são conhecidos pontos de controlo no solo, que servem de referência relativamente à pose da câmara. Contudo, o processo em questão requer pontos conhecidos no solo, o que nem sempre é possível, no caso do oceano, e é moroso no que diz respeito ao processamento, o que se revela desvantajoso para muitas situações. (He et al., 2018). Face às desvantagens apresentadas do método indireto, vários utilizadores optam pela utilização do método direto. Este recorre exclusivamente aos sensores do UAV, para estimar a localização do objeto de interesse (Pfeifer et al., 2012).

### 2.4.1 Transformação de coordenadas espaciais

O sistema geodésico mundial de 1984 (*World Geodetic System 1984*, WGS-84) é um sistema de referência terrestre convencional que inclui na sua definição um sistema de coordenadas geocêntricas e a aproximação da superfície da Terra em forma de elipsoide. Com o decorrer do tempo, foram implementadas diversas melhorias no sistema. No entanto, foi o GPS que mais influenciou a sua evolução.

É possível realizar a transformação de WGS-84 (latitude, longitude e altitude), para *Earth-Center*, *Earth-Fixed (ECEF)* ( $x, y, z$ ), pelo método de Agency, 1987 e posteriormente para ENU. No sentido inverso, pelos métodos de Vermeille, 2004, para a longitude e de Heikkinen, 1982, para a latitude e altitude. Tais conversões são possíveis consultar no Apêndice A.

## 2.5 Sistema de Identificação Automática

O AIS consiste num sistema de seguimento automático instalado na maioria dos navios. Opera numa banda dedicada de *Very High Frequency* (VHF) que, através de transmissores e recetores, torna possível a partilha e receção de dados de modo a obter um panorama local do tráfego marítimo, como se pode observar na Figura 2.7.

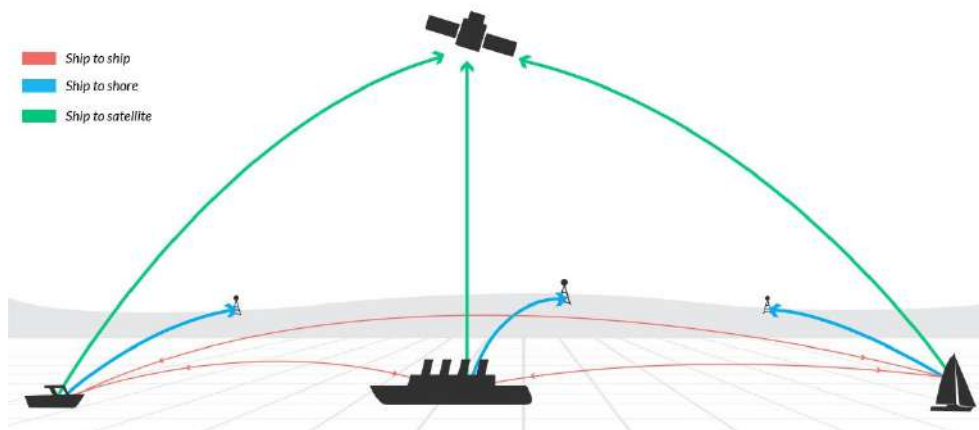


FIGURA 2.7: Comunicações AIS (NATO Shipping Centre, 2021).

Conforme apresentado em Badrudin et al. (2023), a informação fornecida pelo AIS inclui, entre outras informações:

- Número da Organização Marítima Internacional (*International Maritime Organization*, IMO);
- Indicativo de chamada;
- Número de identidade de segurança marítimo móvel (*Maritime Mobile Service Identities*, MMSI);
- Posição do navio, latitude e longitude;
- Tipologia de navio, rumo e velocidade;
- Calado estático e dimensões do navio;
- Origem, destino e estado da navegação.

Em conformidade com o disposto na Regra 19 do Capítulo V da Convenção Internacional para a Salvaguarda da Vida Humana no Mar (*Safety Of Life At Sea*, SOLAS), todos os navios de arqueação bruta igual ou superior a 300 toneladas, afetos a viagens internacionais, bem como os navios de carga de arqueação bruta

igual ou superior a 500 toneladas, não afetos a viagens internacionais e os navios de passageiros, independentemente da sua dimensão, devem estar equipados com AIS. Adicionalmente, de acordo com a alínea a) do nº2 do artigo 10º do Regulamento (CE) nº 1224/2009 do Conselho, de 20 de novembro de 2009, a utilização do AIS é obrigatória para todos os navios de pesca com comprimento fora-a-fora igual ou superior a 15 metros (NATO Shipping Centre, 2021).

## 2.6 Fusão sensorial

A fusão sensorial consiste num processo de combinação de dados provenientes de sensores. Estes, através de modelos matemáticos, clarificam e simplificam a funcionalidade de um sistema, possibilitando medições ou estimativas de medições. O resultado é uma combinação de medições, ou estimativas referentes ao sistema, tais como a velocidade e a distância, que são grandezas físicas, viáveis para utilização em navegações autónomas. Existem diversos métodos de fusão, sendo que neste tópico serão abordados o Filtro de Kalman e as respetivas variantes, estendido e *unscented*, bem como o filtro de partículas.

### 2.6.1 Filtro de Kalman

O Filtro de Kalman (*Kalman Filter*, KF) é um algoritmo que descreve uma solução recursiva para o problema de filtragem linear de dados discretos (Kalman, 1960). Este algoritmo é composto por um conjunto de equações matemáticas capazes de estimar o estado atual de um sistema com base no histórico dos sensores e das ações dos atuadores. A eficácia do KF deve-se, essencialmente, a dois processos distintos, a previsão e a atualização, sendo estes dependentes um do outro (Welch e Bishop, 1995). Para aplicar os processos anteriores, é necessário definir o modelo dinâmico e o modelo de observação subjacentes a estes processos do KF, equações

$$x_k = f(x_{k-1}, u_{k-1}) + w_k \quad (2.5)$$

$$z_k = h(x_k) + v_k, \quad (2.6)$$

respetivamente, onde:

- $x_k$  é o estado instante atual;
- $x_{k-1}$  é o estado anterior

- $f()$  e  $h()$  são funções lineares;
- $u_{k-1}$  é a atuação de controlo;
- $w_k$  são perturbações externas;
- $z_k$  é a observação atual;
- $v_k$  é o ruído de medição.

Uma vez que o Filtro de Kalman assume que as funções de transição de estado  $f()$  e de observação  $h()$  são lineares, é possível representá-las, pelas matrizes  $F$  e  $H$ , respetivamente, correspondentes aos modelos lineares do sistema e da medição. Em conjunto com a matriz de controlo, que representa as atuações exercidas no sistema, e com as matrizes de covariância associadas ao ruído de processo e ao ruído de observação,  $w_k$  e  $v_k$ , é possível derivar as equações fundamentais do filtro, que permitem realizar a estimativa ótima do estado do sistema com base em observações ruidosas.

$$x_k = F_k x_{k-1} + B_k u_k + w_k \quad (2.7)$$

$$z_k = H_k x_k + v_k, \quad (2.8)$$

onde:

- $F_k$  é a matriz de previsão;
- $B_k$  é a matriz de controlo;
- $H_k$  é a matriz de observação.

A fase de previsão baseia-se no modelo dinâmico do sistema, onde  $\hat{x}_k$  representa a previsão do estado atual e conseqüente cálculo da nova incerteza associada,  $P_k$ ,

$$\hat{x}_k = F_k \hat{x}_{k-1} + B_k u_k \quad (2.9)$$

$$P_k = F_k P_{k-1} F_k^T + Q_k, \quad (2.10)$$

respetivamente, em que  $Q_k$  é a matriz de covariância do ruído do processo.

Na atualização, o KF fundamenta-se na previsão previamente calculada do modelo dinâmico, resultando na melhor estima do estado atual,  $\hat{x}'$ , com a inclusão

das observações atuais ( $z_k$ ), e na respetiva previsão da incerteza,  $P'_k$ . A inclusão do ganho de Kalman,  $K'$ , tanto para o estado como para a incerteza, é imperativa para estabelecer uma correlação entre a previsão e a medição, conforme evidenciado nas equações

$$K'_k = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \quad (2.11)$$

$$\hat{x}'_k = \hat{x}_k + K'(z_k - H_k \hat{x}_k) \quad (2.12)$$

$$P'_k = P_k - K' H_k P_k. \quad (2.13)$$

O KF é um estimador recursivo que apenas requer o estado imediatamente anterior, assumindo modelos lineares e ruídos gaussianos. Sob estas condições, garante-se a convergência do algoritmo. Por outro lado, muitos dos problemas não são lineares, surgindo, por conseguinte, variantes do KF, o *Extended Kalman Filter* (EKF) e o *Unscented Kalman Filter* (UKF), que tratam não linearidades, permitindo a aplicação do KF.

O EKF procede à linearização em cada instante das funções  $f()$  e  $h()$ , viabilizando o cálculo das matrizes

$$F_k \approx \left. \frac{\partial f}{\partial \mathbf{x}_k} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k-1}} \quad (2.14)$$

$$H_k \approx \left. \frac{\partial h}{\partial \mathbf{x}_k} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k-1}}. \quad (2.15)$$

A linearização é efetuada por meio da aplicação de derivadas parciais, obtendo-se, deste modo, a matriz Jacobiana, que descreve a relação entre as variáveis do sistema. Trata-se de uma técnica baseada numa expansão da série de *Taylor* de primeira ordem das funções de estado e medição não lineares (N. Yang et al., 2005). As aproximações realizadas já não permitem o pressuposto de otimalidade do filtro e as garantias de convergência.

O UKF constitui um método que trata as não linearidades, recorrendo aos denominados *sigma-points*. Este filtro pressupõe que a incerteza segue uma distribuição normal centrada na estimativa mais recente. Para lidar com funções não lineares, o UKF avalia essas funções em vários *sigma-points*, que são distribuídos ao longo dos eixos principais da elipsóide, que representa a incerteza em torno da estimativa atual. Com base nas avaliações efetuadas nesses pontos, o UKF realiza uma

aproximação linearizada, utilizando uma média ponderada dos resultados obtidos (Xiong et al., 2006).

### 2.6.2 Outros métodos de fusão

O filtro de partículas (*particle filter*, PF) constitui um método sequencial, baseado no método de Monte Carlo, que possibilita uma representação da distribuição de estados, mediante amostragens consecutivas. Em oposição ao EKF, o PF não impõe qualquer restrição quanto à forma das distribuições de probabilidade, sendo, por este motivo, adequado para problemas de estimação totalmente não lineares e não Gaussianos.

A simulação de Monte Carlo envolve a utilização de um conjunto de partículas ponderadas (amostras) para representar a função de densidade posterior necessária, recorrendo a um algoritmo de amostragem de importância sequencial. Essas partículas são posteriormente propagadas através do modelo não linear, permitindo uma aproximação da distribuição posterior mesmo em sistemas altamente não lineares. Assim, é possível estimar de forma precisa momentos estatísticos relevantes, como a média e a covariância do estado. De modo geral, cada iteração do algoritmo PF compreende a atualização das partículas e dos pesos, quando tal se revela necessário.

O número de partículas é crucial para o desempenho do PF, de modo que um maior número de amostras resulta numa densidade superior em cada iteração e, por conseguinte, numa melhor estimativa. Em cada iteração, as partículas com pesos reduzidos são eliminadas e a concentração das partículas é ajustada para uma zona de maior peso (N. Yang et al., 2005).



# Capítulo 3

## Estado da Arte

No presente capítulo, proceder-se-á à análise do estado da arte, com o propósito de explorar as abordagens existentes no que diz respeito ao *dataset*, à deteção de navios e à geolocalização de alvos marítimos.

### 3.1 *Dataset*

Face à evolução dos trabalhos académicos no âmbito da deteção de objetos no mar, foram desenvolvidos diversos *datasets* para a deteção de navios. Atualmente, os UAVs incorporam câmaras de melhor qualidade, pelo que ficam com resolução de imagem superior, comparativamente com as imagens satélite (Geers, 2023).

O *Mid-Air* (Fonder e Droogenbroeck, 2019) é um *dataset* totalmente sintético, enquanto o *Blackbird UAV* (Antonini et al., 2020) combina dados reais de voo com dados sintéticos. Ambos são testados em ambientes simulados, o que permite testar métodos sem recurso a um UAV. Ambos os *datasets* são adequados para o desenvolvimento de algoritmos de visão computacional.

O *dataset Maritime Synthetic* (MarSyn) (M. Ribeiro et al., 2022) é composto por 25 sequências de vídeo fotorrealistas, cada uma com 1000 fotografias, gerados através do *software Blender*, totalizando 25.000 imagens e 34.000 anotações. A simulação inclui uma vasta gama de cenários marítimos, incluindo condições de sol, céu nublado e pôr do sol, imagens de proximidade da costa e reflexos na água, bem como uma diversidade de tipos de embarcações, tais como navios de carga, militares, barcos de pesca e botes de salvamento, com comprimentos variando entre 3 e 125 metros. As imagens obtidas através de uma câmara simulada de um UAV a altitudes compreendidas entre 150 e 1000 metros, como ilustrado na Figura 3.1.



FIGURA 3.1: Exemplos de imagens do *MarSyn Dataset* (M. Ribeiro et al., 2022).

O *Seagull* (R. Ribeiro et al., 2019) é um *dataset* constituído por vídeos capturados através de um UAV equipado com diversas câmaras, das quais duas pertencem ao espectro visível, *GoPro Hero 2* e *Tase 150*. Este apresenta uma vasta quantidade de imagens, capturadas em diversos cenários, incluindo diferentes reflexos, cristas das ondas e perspetivas da imagem. No entanto, é limitado relativamente à tipologia de navios, uma vez que as filmagens apenas incluem um único navio, conforme ilustrado na Figura 3.2.

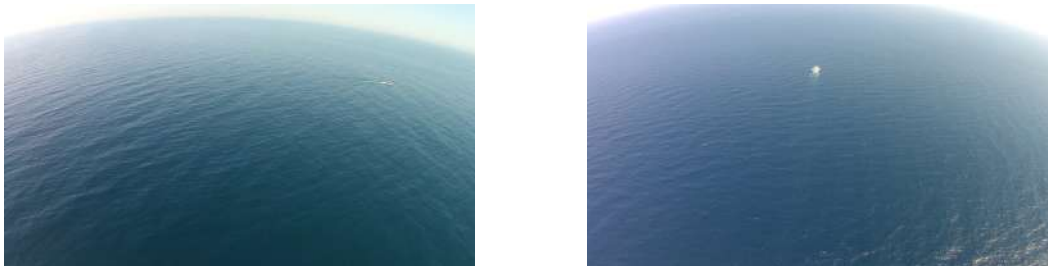


FIGURA 3.2: Exemplos de imagens do *Seagull* (R. Ribeiro et al., 2019).

O *dataset* de Rubí et al., 2024 contém mais de seis mil imagens, contudo, comparativamente ao *Seagull*, apresenta uma limitação superior no que diz respeito à variedade das imagens. Esta limitação deve-se ao facto de o *dataset* estar direcionado para a vigilância de um porto, sendo estas imagens próximas e de baixa altitude, como mostra a Figura 3.3.

Existe também o *Aerial Vessels Detection Dataset* (Makrigiorgis et al., 2022), que apresenta mais de dez mil imagens. Contudo, muitas das imagens são idênticas, uma vez que resultam de *frames* de vídeos. Engloba 3 classes, pessoas, navios de pequenas e médias dimensões e navios de grandes dimensões.

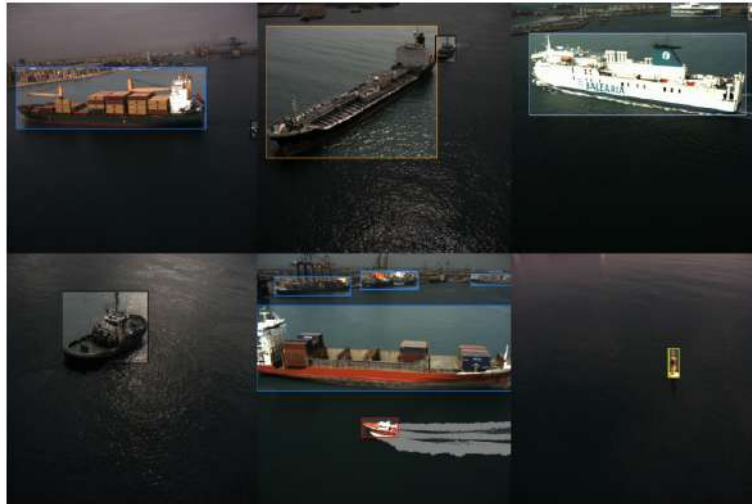


FIGURA 3.3: Exemplo de imagens do VESSELIing (Rubí et al., 2024).

O *ShipDataset* (J. Zhao et al., 2023) abrange cinco cenários com diferentes ângulos de filmagem e condições de iluminação, obtidos através de vídeos capturados por meio de *drone*. O tráfego marítimo nas imagens capturadas apresenta uma situação de congestionamento, conforme ilustrado na Figura 3.4.



FIGURA 3.4: Exemplo de imagens do *ShipDataset* (J. Zhao et al., 2023).

O *ship aeriels 2 Computer Vision Dataset* (Marine ship 2, 2023) e o *Aerial views of boats Computer Vision Model* (Scanu, 2024) são *datasets* depositados no *Roboflow*, que possuem uma vasta diversidade de tipologias de navios, ângulos de captura de imagem e brilhos. Nomeadamente o *ship aeriels 2 Computer Vision Dataset*, que apesar de apenas possuir 60 imagens, representa o cenários ideais para o efeito.

Contudo, no contexto da dissertação, os *datasets* disponíveis são limitados. Frequentemente, os estudos concentram-se em apenas uma tipologia de navios, aplicam um nível elevado de zoom a cada navio ou são imagens capturadas a partir do solo. Seguidamente, serão apresentados *datasets* relevantes para o âmbito do trabalho.

## 3.2 Detecção de navios

A deteção de navios consiste em determinar a localização e a classe dos objetos detetados numa imagem específica. Neste caso, apenas se procura detetar objetos que sejam classificados como navios, obtendo a sua localização na imagem através de BBs. Por conseguinte, é imperativo determinar as coordenadas centrais da deteção, bem como a sua largura e altura, através do tamanho da BB.

No âmbito do estudo desenvolvido por R. W. Liu et al. (2021), foi concebido um método de aprendizagem aprimorado, fundamentado numa CNN, que propõe uma extensão do modelo *yolov3*. As BBs foram redesenhadas para se ajustarem melhor aos navios e foi realizada uma previsão das incertezas de localização das mesmas.

Com a evolução da arquitetura do YOLO, D. Zhao et al. (2023) optou pelo *yolov7* otimizado para a vigilância por UAVs, com uma atenção particular para objetos de dimensões reduzidas. Adicionalmente, inclui um módulo de atenção deformável na rede principal para otimizar a extração de características, bem como convoluções deformáveis e separáveis em profundidade para maximizar a extração semântica e minimizar parâmetros. A função de ativação *Mish* é utilizada, sendo o modelo sujeito a um processo de treino que envolve técnicas de pré-treino e ajuste fino no *dataset VisDrone2019*. O modelo alcançou um *mAP50* de 52,3% no conjunto de validação, apresentando um desempenho superior em condições de oclusão e indefinição.

Além disso, a câmara RGB pode ser utilizada em simultâneo com outros sensores, como a câmara de infravermelho, *RA*dio *D*etection *A*nd *R*anging (RADAR) e *L*ight *D*etection *A*nd *R*anging (LiDAR), como apresentado por Farahnakian et al. (2018). Inicialmente, o método *Selective Search* gera propostas de RoI em imagens RGB, que são posteriormente refinadas, visando aumentar a confiabilidade dos resultados. Uma CNN permite depois a classificação dos objetos (navio, terra) nas

propostas finais. Este método apresenta uma maior precisão e eficiência, ao apresentar um menor número de propostas em comparação com os métodos de sensor único.

O algoritmo YOLACT++ foi implementado por M. Ribeiro et al. (2022) para realizar a detecção e segmentação de embarcações em cada imagem, de forma independente. Posteriormente, um *Conditional Random Field* 3D refina a segmentação, explorando correlações temporais entre imagens para garantir consistência e corrigir erros. O modelo foi treinado com dados sintéticos do conjunto *MarSyn* e filmagens aéreas de outros *datasets* como o *Seagull* e o *Airbus*, devido à escassez de vídeos marítimos rotulados.

De modo a detetar a tipologia de navios que passavam num canal, Akiyama et al., 2018 propõe um detetor baseado em CNN, utilizando o *Visual Geometry Group* 19 (VGG19) pré-treinado no *ImageNet* com transferência de aprendizagem devido à limitação de dados. O sistema em questão efetua a classificação de cinco tipologias de embarcações (recreio, rebocadores, casas-barco, *jet skis* e outras) com base numa amostra de 3152 imagens obtidas a partir de uma câmara de vigilância localizada em Tóquio. O modelo utilizou a função de ativação de *Rectified Linear Unit* (ReLU), a perda de entropia cruzada, o otimizador Adam e *Early Stopping*, alcançando um *F1-Score* médio de 0.70.

### 3.3 Geolocalização de alvos marítimos

A utilização de UAV para efeitos de geolocalização tem-se tornado cada vez mais frequente. Em geral, a geolocalização direta é o método mais comum (Eling et al., 2014). Segundo Gibbins et al. (2004) a geolocalização enfrenta três desafios principais: a estimativa da pose do UAV, a transformação do referencial da câmara para o mundo e a computação para localizar o alvo. O estudo de Johnston (2006) apresenta um método para determinar a localização geográfica de objetos no solo, recorrendo a câmaras de vídeo de UAVs. O processo em questão envolve a utilização de matrizes de rotação e geometria do sistema de câmara para calcular as coordenadas geográficas WGS-84 (latitude, longitude e elevação) de um objeto referenciado dentro do campo de visão da câmara. Os cálculos são efetuados com base na posição GPS e atitude do UAV, bem como na elevação do terreno abaixo do veículo. O artigo descreve os quatro blocos fundamentais para esta abordagem: transformações de dados, transformações de referência, parâmetros da câmara e características da

lente. No estudo realizado por Cacace et al. (2024), a geolocalização direta é utilizada em conjunto com o filtro de Schmidt-Kalman, com o propósito de determinar a melhor estimativa da posição do alvo, tendo em conta a incerteza da posição do UAV.

No estudo de Carrillo-Perez et al. (2022) é adotada a geolocalização indireta, onde a homografia é empregue para determinar a posição dos navios. Neste sentido, é estabelecida uma correlação entre os píxeis da imagem e as respetivas coordenadas geográficas, de modo a obter pontos de referência na imagem. A principal limitação deste método reside no facto das câmaras permanecerem estáticas. Já em, Babinec e Apeltauer (2016) é aplicado o mesmo processo mas com a câmara fixada num UAV a praticar voos a baixas altitudes. Para que esta homografia seja estabelecida, são necessários pontos de referência no solo, cujas coordenadas no mundo real são conhecidas, sendo exigidos pelo menos quatro marcos não colineares para uma estimativa completa da transformação. Em oposição a um sistema de câmara fixa, a plataforma UAV é dinâmica. Assim, o foco deste estudo é a análise da precisão e propagação de erros, resultantes de imprecisões nas posições dos marcos, elevação dos objetos e deformações geométricas da lente da câmara.

# Capítulo 4

## Metodologia

No presente capítulo, é apresentada a metodologia utilizada para a deteção de navios e a respetiva geolocalização. A partir de dados adquiridos, tais como a pose do UAV e a pose da câmara, com uma calibração conhecida da câmara e com um detetor de navios, é possível realizar a geolocalização dos mesmos. Este processo permite determinar a localização dos navios detetados, obtendo-se, deste modo, um panorama situacional de uma determinada área.

### 4.1 Arquitetura geral

O processo é iniciado com a captura de imagens pela câmara do UAV. A partir das imagens obtidas, procede-se à deteção de navios nas mesmas, de modo a obter o pixel central de cada deteção. Em simultâneo com a captura das imagens, são recolhidos todos os dados relevantes da pose do UAV e da câmara. A recolha de dados do UAV foi implementada através das mensagens MAVLink “GLOBAL\_POSITION\_INT” e “ATTITUDE”, obtendo a pose do UAV. Para adquirir a pose da câmara SIYI ZR10 foi estabelecida conexão via Protocolo de Datagrama do Utilizador (*User Datagram Protocol*, UDP).

Previamente à obtenção dos dados, é imprescindível realizar a recolha dos parâmetros intrínsecos da câmara. Estes elementos apresentam uma natureza fixa, uma vez que se pretende evitar variações de zoom durante o processo de captação das imagens. A calibração foi realizada com recurso a ferramentas do *Open Source Computer Vision* (OpenCV) e *Matrix Laboratory* (MATLAB).

Após a recolha de toda a informação anteriormente referida, é possível realizar a geolocalização de alvos marítimos, uma vez que se assume que todos os alvos estão à altitude 0 metros. A determinação das coordenadas WGS-84, designadamente a latitude e a longitude do alvo, é efetuada com base na sua posição no

referencial do mundo. Através das coordenadas do referencial do mundo, é possível converter para coordenadas ENU, coordenadas geodésicas e, por fim, WGS-84, respectivamente. A Figura 4.1 ilustra a *framework* adotada para a detecção da latitude e longitude dos alvos detetados.

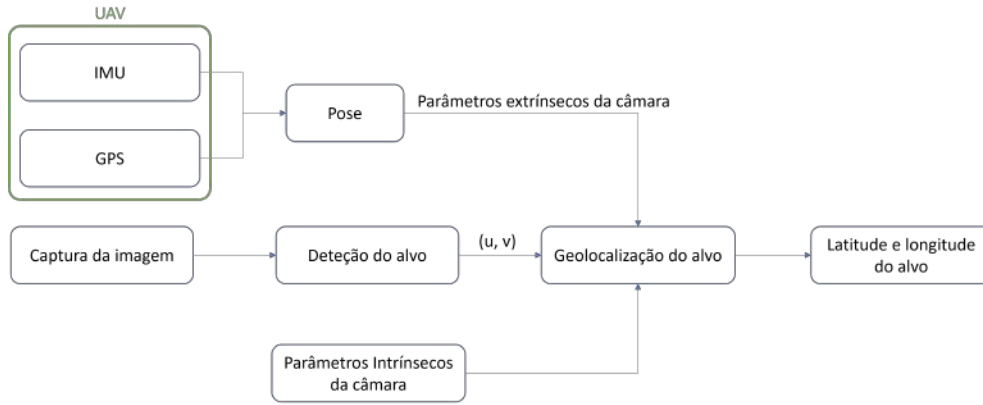


FIGURA 4.1: *Framework* adotada para a estimativa da latitude e longitude de um alvo.

## 4.2 Referenciais

O referencial do mundo terá origem no local da descolagem do UAV na altitude zero, definindo o eixo  $x$  para norte, o  $y$  para oeste e o  $z$  para cima. Esta definição foi estabelecida com o intuito de facilitar a transformação para o referencial do UAV, uma vez que este é definido com  $x$  a apontar para a frente,  $y$  para bombordo e  $z$  para cima. Em circunstâncias convencionais, as coordenadas ENU seriam habitualmente utilizadas. No entanto, a referida transformação é aplicada exclusivamente no final, com o propósito de obter as coordenadas no formato WGS-84.

Partindo do princípio de que a distância entre a origem do referencial do UAV e o centro da câmara é irrelevante, assume-se que a origem dos referenciais é a mesma, alterando apenas os eixos. Neste caso, o eixo do  $z$  orientado para fora da lente da câmara, o  $x$  para estibordo e o  $y$  para baixo, perpendicular a  $z$  e  $x$ , como ilustrado na Figura 4.2.

No que diz respeito à imagem, a origem do referencial encontra-se no canto superior esquerdo da imagem, com  $u$  a apontar para a direita e  $v$  para baixo. De notar que o eixo  $z$  da câmara cruza a imagem no centro desta, conforme demonstrado na reta cinzenta representada na Figura 4.2.

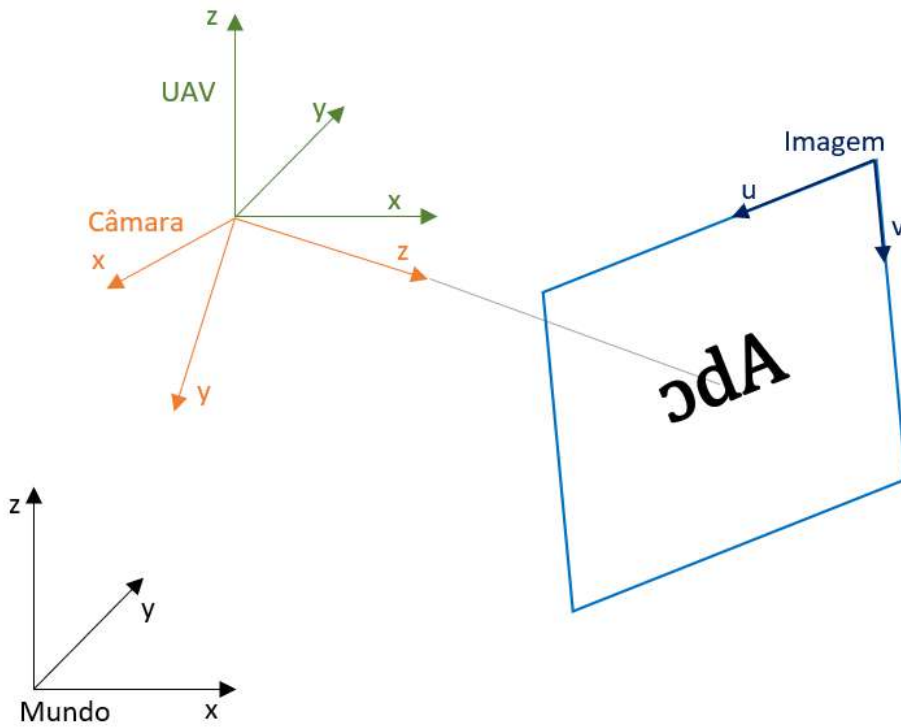


FIGURA 4.2: Referenciais do mundo, UAV, câmara e imagem.

Para projetar um ponto do mundo real na imagem captada por uma câmara acoplada a um UAV, é necessário encadear uma sequência de transformações homogêneas entre diferentes referenciais. O processo inicia-se com as coordenadas no referencial do mundo, passando pelo referencial do UAV, pelo da câmara e, finalmente, pelo referencial da imagem, de acordo com:

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \rightarrow \begin{bmatrix} x_{uav} \\ y_{uav} \\ z_{uav} \end{bmatrix} \rightarrow \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \quad (4.1)$$

A composição das duas primeiras transformações resulta na transformação direta do mundo para a câmara, designada por:

$${}^cT_w = {}^cT_{uav} \cdot {}^{uav}T_w. \quad (4.2)$$

A notação adotada segue a convenção de transformações homogêneas entre referenciais distintos:

- ${}^{uav}T_w$ : transformação do referencial do mundo para o referencial do UAV;
- ${}^cT_{uav}$ : transformação do referencial do UAV para o referencial da câmara;
- $K$ : matriz de parâmetros intrínsecos da câmara, que realiza a transformação do referencial da câmara para o plano da imagem.

Finalmente, ao aplicar a matriz intrínseca  $K$ , obtém-se a projeção de um ponto expresso em coordenadas do mundo no espaço da imagem, como descrito na Equação 2.4.

### 4.3 Descrição dos equipamentos utilizados

A recolha de dados foi efetuada através da realização de um voo com um UAV equipado com uma câmara, contando com o apoio da Célula de Inovação e Experimentação Operacional de Sistemas Não Tripulados (CEOV) da Marinha Portuguesa. O voo foi realizado no Mar da Palha, no estuário do rio Tejo. O UAV utilizado foi um hexacóptero, conforme ilustrado na Figura 4.3, que possui seis braços e rotores. Este possui seis graus de liberdade e destaca-se pela sua estabilidade, precisão, robustez e eficiência. A principal vantagem desta tipologia de UAV reside na sua redundância, que permite a realização de voos seguros mesmo em caso de falha de um motor.



FIGURA 4.3: UAV utilizado no voo.

A câmara acoplada ao UAV é uma *SIYI ZR10 2K* com 30x zoom híbrido e 10x zoom ótico. Esta câmara possui estabilizadores nos três eixos, conforme ilustra a Figura 4.4.



FIGURA 4.4: Câmara *SIYYI ZR10* (SIYYI Technology, 2024).

A recolha de dados AIS foi efetuada recorrendo a uma antena AIS através do programa *OpenCPN*. Os dados originais recolhidos fornecem o número de identificação do serviço móvel marítimo (*Maritime Mobile Service Identity*, MMSI). Todavia, com o intuito de facilitar a organização de dados e de não revelar a identidade dos navios presentes, este dado foi alterado para “Ship”, atribuindo a cada navio um número de identificação. Através destes dados, obtém-se também a velocidade relativa ao solo, *Speed Over Ground* (SOG), o rumo relativo ao solo, *Course Over Ground* (COG) e a proa, *Heading*.

## 4.4 Calibração da câmara

A calibração da câmara é um processo que visa estabelecer a correspondência entre um ponto específico no espaço e a sua representação numa imagem bidimensional (Chen et al., 2023). De referir que é através da matriz dos parâmetros intrínsecos,  $K$ , que se efetua a transição do referencial da câmara para o plano da imagem.

Para este efeito, recorre-se ao método proposto por Zhang (2000), o qual, através de um padrão xadrez, como ilustrado na Figura 4.5, permite determinar os parâmetros intrínsecos (Equação 4.3) e os coeficientes de distorção da lente (Equação 4.4).

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (4.3)$$

$$d = [k_1 \quad k_2 \quad p_1 \quad p_2 \quad k_3] \quad (4.4)$$

O padrão xadrez utilizado é aconselhado pela *OpenCV*, este consiste num tabuleiro designado por 9x6 em que todos os quadrados têm as mesmas dimensões. O tabuleiro utilizado foi retirado de bibliotecas da *OpenCV* (OpenCV, 2018).



FIGURA 4.5: Tabuleiro de xadrez para calibração da câmara.

A determinação dos parâmetros intrínsecos e da distorção da câmara foi realizada com recurso à ferramenta *Camera Calibrator* do MATLAB, que para entrada recomenda, entre 10 e 20 imagens. No decorrer do trabalho, verificou-se que seria mais vantajoso utilizar um maior número de imagens do que as recomendadas. Nesse sentido, foram utilizadas 34 imagens com diferentes ângulos, distâncias e fontes de iluminação.

## 4.5 Detecção de navios

### 4.5.1 Detecção de navios

A deteção de navios implica uma vasta diversidade de imagens para treino da rede neuronal. Essas imagens têm de conter diversas tipologias de navios, de diversos ângulos e com diferentes brilhos, uma vez que essa variabilidade é essencial para que a rede adquira uma capacidade de generalização adequada. Sem este nível de diversidade, o modelo tende a sobreajustar-se a cenários específicos, comprometendo o seu desempenho em situações reais, nas quais a aparência dos navios pode variar significativamente em função da iluminação, da perspectiva ou do tipo de embarcação.

Para o treino da rede neuronal recorreu-se ao modelo pré-treinado *yolov8*, uma arquitetura de rede neuronal convolucional (CNN) otimizada para tarefas de deteção em tempo real. A versão do YOLO selecionada apresenta uma vantagem significativa em termos de documentação existente, um bom suporte da *Ultralytics*, versatilidade e disponibilidade de modelos pré-treinados.

A rede foi treinada com 100 épocas, utilizando imagens com resolução de 640x640 píxeis e lotes de 16 imagens.

#### 4.5. Detecção de navios

A seleção do modelo será efetuada em três etapas. Primeiramente, aferir a necessidade ou não de adicionar *datasets* de modo a complementar as similaridades entre imagens, esse estudo será realizado com a adição de dois *datasets* de menores dimensões.

De seguida, será avaliada a necessidade de realizar um *data augmentation*, recorrendo ao *yolov8n*. Posteriormente, será verificada qual a variante do YOLOv8 que apresenta melhor equilíbrio entre a precisão e o poder de computação. Esta avaliação será realizada exclusivamente entre as variantes *nano* e a *small* do *yolov8*.

O *data augmentation* ou, em português, aumento de dados, é realizado por meio de ferramentas do Ultralytics (2025), nomeadamente, ajustes de cores, saturação, brilho, rotação, escala, cisalhamento, perspetiva e inversão esquerda-direita. Todos estes foram utilizados nos valores predefinidos pelo autor. Desta forma, é possível variar ângulos dos navios, brilhos das imagens, entre outros, de forma aleatória. A Figura 4.6 ilustra a aplicação de *data augmentation*.

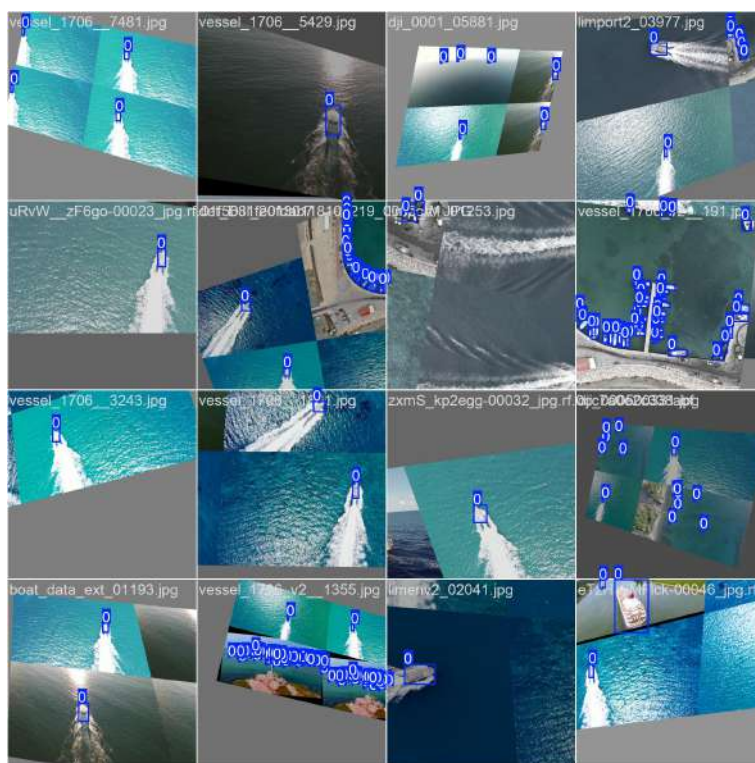


FIGURA 4.6: Exemplo de *data augmentation* no *Aerial Vessels Detection Dataset*.

Como resultado do processo de deteção, são gerados ficheiros de texto no formato YOLO para cada imagem. O arquivo de texto é constituído por uma linha por objeto, apresentando o seguinte formato: `classe xcentro ycentro largura`

altura. Os valores em questão encontram-se normalizados entre 0 e 1. É importante salientar que, para o presente trabalho, é suficiente ter conhecimento do ponto central da detecção. Na Figura 4.7, é apresentado um exemplo de uma BB. A construção desta BB foi orientada pelos seguintes valores: 0 0.480469 0.367188 0.232813 0.377757. Considerando que a imagem apresenta dimensões de 1920 x 1088, infere-se que o centro da detecção se localiza no píxel [923 400].



FIGURA 4.7: Detecção numa imagem do *Aerial Vessels Detection Dataset*.

## 4.5.2 *Datasets* de treino

O *dataset* principal será o *Aerial Vessels Detection Dataset* este contempla 10 252 imagens (Makrigiorgis et al., 2022). Estas possuem as características pretendidas pois foi um *dataset* propositadamente criado para vigilância marítima, como se pode ver na Figura 4.8.

O conjunto de dados está dividido em três subconjuntos distintos: 1 024 imagens destinadas à validação, 1 025 imagens reservadas para teste e as restantes 8 203 imagens utilizadas para treino. No que diz respeito às instâncias, todas as detecções de pessoas foram desconsideradas, e a distinção entre barcos, de menores dimensões, e navios, de maiores dimensões, não foi mantida. Em vez disso, optou-se por unificar estas categorias numa única classe designada por “Navio”. Assim, foram utilizadas 49 470 instâncias para o treino, 6 033 para validação e 5 356 para teste. A Tabela 4.1 apresenta a composição detalhada do *dataset*.

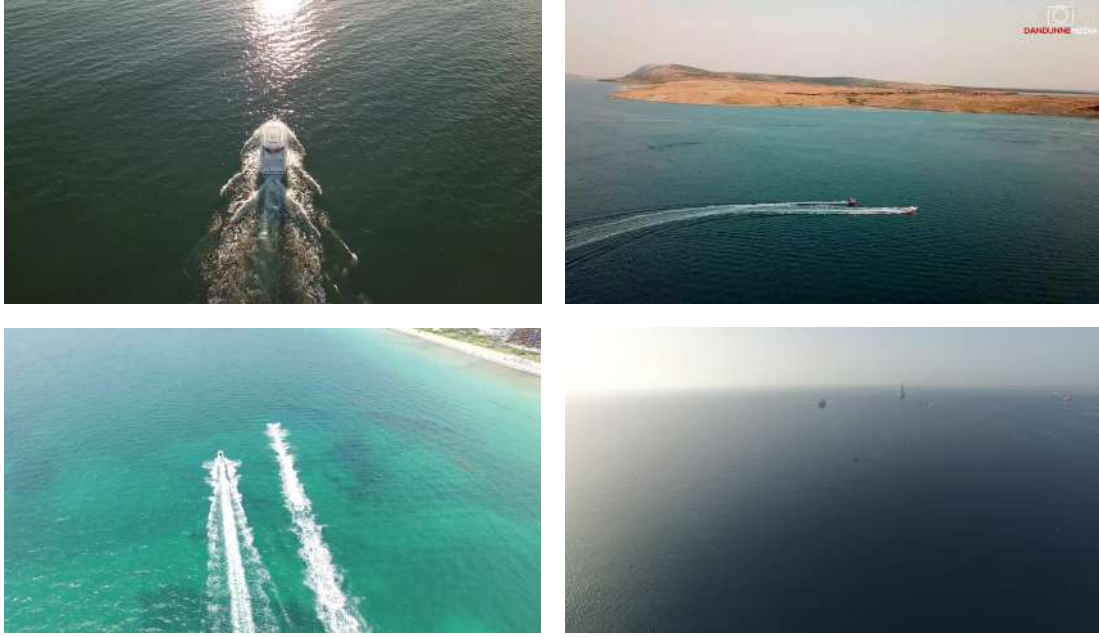


FIGURA 4.8: Imagens pertencentes ao *Aerial Vessels Detection Dataset*.

TABELA 4.1: Distribuição do *Dataset*.

Subconjunto	Imagens	Pessoas	Barcos	Navios
Treino	8 203	219	48 550	920
Validação	1 024	7	5 890	143
Teste	1 025	13	5 247	109

Os *datasets* referidos na primeira etapa da seleção do modelo são o *ship aerals 2 Computer Vision Dataset* (Marine ship 2, 2023), Figura 4.9 e o *Aerial views of boats Computer Vision Model* (Scanu, 2024), Figura 4.10, de menores dimensões, de 60 e 1 469 imagens, respetivamente. Estes acrescentam tipologias de navios e ângulos de incidência diferentes ao *dataset* principal.

## 4.6 Geolocalização

As variáveis necessárias para a geolocalização são a pose do UAV, definida pela sua posição geográfica (latitude, longitude e altitude) e orientação (*roll*, *pitch* e *yaw*), bem como a atitude da câmara em relação ao UAV (*pitch* e *yaw*). A partir destas variáveis é possível determinar a matriz de transformação extrínseca que relaciona o referencial do mundo com o referencial da câmara:

$${}^cT_w = {}^cT_{uav} \cdot {}^{uav}T_w. \quad (4.5)$$

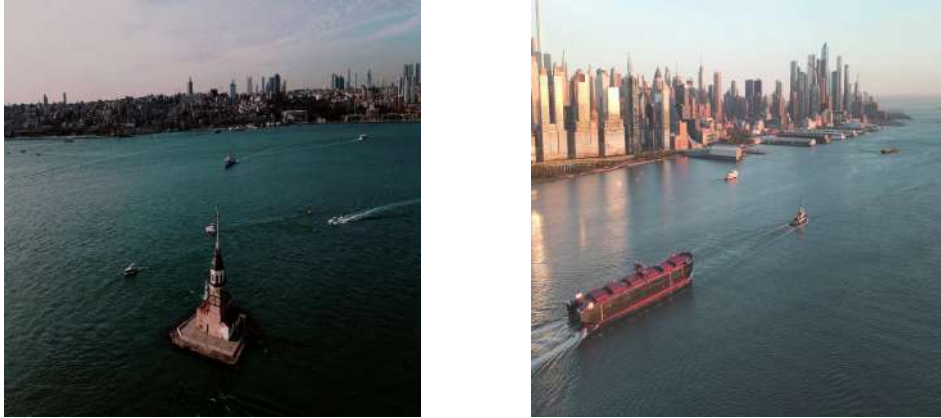


FIGURA 4.9: Imagens pertencentes a *ship aerals 2 Computer Vision Dataset*.



FIGURA 4.10: Imagens pertencentes a *Aerial views of boats Computer Vision Model*.

No primeiro passo da implementação, assumindo-se conhecida a posição real do navio no referencial do mundo, determinam-se as suas coordenadas na imagem. Para tal, consideram-se as coordenadas homogêneas de um ponto no mundo:

$$P_w = [x_w \ y_w \ z_w \ 1]^T, \quad (4.6)$$

que são transformadas para o referencial da câmara através de  ${}^cT_w$ :

$$P_c = {}^cT_w \cdot P_w. \quad (4.7)$$

De seguida, aplica-se a projeção no plano da imagem através da matriz intrínseca  $K$ :

$$P_I = K \cdot P_c, \quad (4.8)$$

e, por fim, procede-se à normalização, de modo a obter as coordenadas homogêneas da imagem:

$$[u \ v \ 1]^T. \quad (4.9)$$

Na etapa seguinte, pretende-se resolver o problema inverso: determinar a posição do navio no mundo a partir das suas coordenadas na imagem. Neste caso, apenas o vetor de direção da projeção é conhecido (na forma normalizada), surgindo, portanto, a necessidade de introduzir o fator de escala  $\lambda$ . Assim, obtém-se:

$$P_c = \lambda \cdot K^{-1}[u \ v \ 1]^T, \quad (4.10)$$

e, posteriormente, reconstróem-se as coordenadas no mundo a partir da transformação inversa:

$$P_w = ({}^cT_w)^{-1} \cdot P_c. \quad (4.11)$$

Como o valor de  $z_w$  é previamente conhecido, os navios encontram-se à superfície,  $z_w = 0$ , o sistema pode ser resolvido para obter apenas  $(x_w, y_w)$  e o correspondente  $\lambda$ .

Retomando a Equação 4.10, é possível expandir a transformação entre o referencial da câmara e o do mundo, obtendo:

$$\lambda K^{-1}P_I = {}^cT_w P_w, \quad (4.12)$$

onde  $P_I = [u \ v \ 1]^T$  representa as coordenadas homogêneas da imagem. Para simplificar a notação, define-se:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad (4.13)$$

sendo, portanto,  $\alpha = [\alpha_1 \ \alpha_2 \ \alpha_3]^T$  o vetor de direção no referencial da câmara associado ao píxel  $(u, v)$ .

Substituindo 4.13 em 4.12, obtém-se:

$$\lambda \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 0 \\ 1 \end{bmatrix}. \quad (4.14)$$

A expressão anterior pode ser reescrita sob a forma matricial equivalente:

$$\begin{bmatrix} r_{11} & r_{12} & -\alpha_1 \\ r_{21} & r_{22} & -\alpha_2 \\ r_{31} & r_{32} & -\alpha_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ \lambda \end{bmatrix} = \begin{bmatrix} -t_x \\ -t_y \\ -t_z \end{bmatrix}, \quad (4.15)$$

cuja solução é dada por:

$$\begin{bmatrix} x_w \\ y_w \\ \lambda \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & -\alpha_1 \\ r_{21} & r_{22} & -\alpha_2 \\ r_{31} & r_{32} & -\alpha_3 \end{bmatrix}^{-1} \begin{bmatrix} -t_x \\ -t_y \\ -t_z \end{bmatrix}. \quad (4.16)$$

Após a determinação da posição do navio no referencial do mundo ( $P_w$ ), através da cinemática inversa, é necessário converter estas coordenadas para o sistema geodésico global WGS-84 (latitude, longitude e altitude). Para tal, seguem-se os seguintes passos:

1. Primeiramente, a primeira posição do UAV na altitude zero metros, considerada como origem do referencial do mundo, é convertida para coordenadas no sistema *East-North-Up* (ENU), a partir das respectivas coordenadas geodésicas  $(\varphi_d, \lambda_d, h_d)$ , correspondentes à latitude, longitude e altitude do UAV.
2. A posição relativa do navio, obtida no referencial cartesiano  $(x, y, z)$ , é convertida para o sistema ENU através da matriz de rotação:

$$M_{xyz \rightarrow ENU} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.17)$$

Assim, obtém-se:

$$P_w^{ENU} = M_{xyz \rightarrow ENU} \cdot P_w. \quad (4.18)$$

3. A posição absoluta do navio em ENU é então dada pela soma entre a origem do referencial do mundo em ENU,  $[e \ n \ u]^T$ , e a posição relativa do navio:

$$P_{ENU} = \begin{bmatrix} e \\ n \\ u \end{bmatrix} + P_w^{ENU}. \quad (4.19)$$

4. Por fim,  $P_{ENU}$  é convertido para o sistema de coordenadas *Earth-Centered, Earth-Fixed* (ECEF) e, posteriormente, para coordenadas geodésicas WGS-84, correspondentes à latitude e longitude do navio.

## 4.7 Aplicação do filtro de Kalman

A aplicação do Filtro de Kalman (KF) foi realizada para estimar a posição bidimensional  $(x, y)$  de um alvo a partir de medições de coordenadas de imagem  $(u, v)$  obtidas por uma câmara, integrando um modelo dinâmico com incertezas nas medições e no processo.

As variáveis  $P_x$ ,  $P_y$  e  $P_z$  representam, respetivamente, a posição tridimensional em  $x$ ,  $y$  e  $z$ . Contudo, uma vez que se pressupõe que a posição do navio em  $z$  é constante e igual a zero, a sua posição ( $P$ ) e velocidade ( $V$ ) nesse eixo não serão mencionadas. Consequentemente, o vetor de estado do navio ( $x$ ) sofrerá alterações, como demonstra a equação

$$x_k = \begin{bmatrix} P_x \\ P_y \\ P_z \\ V_x \\ V_y \\ V_z \end{bmatrix} \rightarrow x_k = \begin{bmatrix} P_x \\ P_y \\ V_x \\ V_y \end{bmatrix}. \quad (4.20)$$

A observação do KF corresponde às coordenadas da imagem  $(u, v)$ , obtidas através das deteções dos alvos na imagem:

$$z = \begin{bmatrix} u \\ v \end{bmatrix} \quad (4.21)$$

Podemos verificar as equações cinemáticas para as dimensões referidas acima na equação

$$\begin{cases} v_x(k) = v_x(k-1) + \Delta T a_x(k-1) \\ p_x(k) = p_x(k-1) + \Delta T v_x(k-1) + \frac{1}{2}\Delta T^2 a_x(k-1) \\ v_y(k) = v_y(k-1) + \Delta T a_y(k-1) \\ p_y(k) = p_y(k-1) + \Delta T v_y(k-1) + \frac{1}{2}\Delta T^2 a_y(k-1) \end{cases}. \quad (4.22)$$

A dinâmica discreta, com passo  $\Delta T$ , apresentada na Equação 2.8, no presente trabalho, a matriz de previsão e de controlo apresenta os seguintes valores, respetivamente:

$$F_k = \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad e \quad B_k = \begin{bmatrix} \frac{\Delta T^2}{2} & 0 \\ 0 & \frac{\Delta T^2}{2} \\ \Delta T & 0 \\ 0 & \Delta T \end{bmatrix}. \quad (4.23)$$

Esta evidência sugere uma maior complexidade na previsão do estado atual. Por conseguinte, a incerteza associada ( $P_k$ ) também sofrerá alterações, conforme a Equação 2.10. Em que a matriz de covariância do ruído do processo,  $Q_k$  é definida por, onde  $\sigma_a^2$  é a variância do processo:

$$Q_k = \sigma_a^2 B_k B_k^T, \quad (4.24)$$

$$Q = \sigma_a^2 \cdot \begin{bmatrix} \frac{\Delta T^4}{4} & 0 & \frac{\Delta T^3}{2} & 0 \\ 0 & \frac{\Delta T^4}{4} & 0 & \frac{\Delta T^3}{2} \\ \frac{\Delta T^3}{2} & 0 & \Delta T^2 & 0 \\ 0 & \frac{\Delta T^3}{2} & 0 & \Delta T^2 \end{bmatrix}. \quad (4.25)$$

Na fase de atualização, os valores estimados durante a etapa de predição são corrigidos com base nas observações disponíveis. No âmbito deste estudo, estas observações correspondem às posições dos navios na imagem ( $P_I$ ). Para esta fase, a câmara terá o seu modelo de observação, com o objetivo de estabelecer uma relação matemática entre as informações extraídas das imagens e as coordenadas no

referencial do mundo real ( $P_w$ ). Desta forma, considera-se a seguinte equação:

$$P_I = h \cdot P_w = (h_{INT} \circ h_{EXT}) \cdot P_w. \quad (4.26)$$

Como já visto anteriormente,  $h_{EXT} = {}^c T_w$ , que corresponde aos parâmetros extrínsecos,  $P_c = h_{EXT} \cdot (P_w)$ . Ou seja,

$$P_I = h_{INT} \cdot P_c. \quad (4.27)$$

A determinação de  $h_{INT}$  encontra-se descrita no Apêndice B na função `calcular_h_e_H`.

Considerando que a função  $h()$  é uma equação não linear, que mapeia a posição do navio  $P_w$  para as coordenadas de imagem  $P_I$  sem depender da sua velocidade, é necessário recorrer ao Filtro de Kalman Estendido (EKF) para lidar com esta não linearidade. O EKF constitui um método estatístico que permite estimar o estado de um sistema dinâmico não linear, combinando um modelo de movimento com medições ruidosas.

Para abordar a não linearidade da função de observação  $h()$  no contexto do EKF, é necessário calcular o jacobiano  $H_k$ , que representa a derivada parcial de  $h$  em relação ao estado estimado  $\hat{x}_{k-1}$ , permitindo a linearização local do sistema. Os elementos desse jacobiano são definidos como se segue:

$$H_{INT} = \left. \frac{\partial h_{INT}}{\partial x} \right|_{x=\hat{x}_{k-1}} = \begin{bmatrix} h_{11} & h_{12} & 0 & 0 \\ h_{21} & h_{22} & 0 & 0 \end{bmatrix}, \quad \text{onde :} \quad (4.28)$$

$$h_{11} = \frac{\partial u}{\partial x_w}, \quad h_{12} = \frac{\partial u}{\partial y_w}, \quad (4.29)$$

$$h_{21} = \frac{\partial v}{\partial x_w}, \quad h_{22} = \frac{\partial v}{\partial y_w}. \quad (4.30)$$

A resolução de cada equação encontra-se devidamente resolvida e demonstrada em Damas et al., 2024.

Para obter o jacobiano completo  $H$  do sistema, é necessário considerar a composição da função de observação  $h = h_{INT} \circ h_{EXT}$ . A transformação extrínseca  $h_{EXT}$ , definida pela matriz  ${}^c T_w$ , mapeia as coordenadas do mundo  $P_w = [P_x, P_y, 0]^T$

para o referencial da câmara  $P_c = [x_c, y_c, z_c]^T$ :

$$P_c = {}^c T_w \cdot [P_x \ P_y \ 0 \ 1]. \quad (4.31)$$

A função intrínseca  $h_{INT}$  projeta  $P_c$  para as coordenadas de imagem  $P_I = [u, v]^T$  através de uma relação não linear, aproximada por:

$$u = f \frac{x_c}{z_c} (1 + k_1 r^2) + u_0, \quad v = f \frac{y_c}{z_c} (1 + k_1 r^2) + v_0, \quad (4.32)$$

onde  $r^2 = (x_c/z_c)^2 + (y_c/z_c)^2$ ,  $f$  é a distância focal,  $(u_0, v_0)$  é o ponto principal, e  $k_1$  é o coeficiente de distorção radial.

O jacobiano  $H$  é derivado aplicando a regra da cadeia, uma vez que  $h$  é composta por  $h_{INT}$  e  $h_{EXT}$ . Assim, o jacobiano total é dado por:

$$H = \frac{\partial h}{\partial x} = \frac{\partial h_{INT}}{\partial P_c} \cdot \frac{\partial h_{EXT}}{\partial x}, \quad (4.33)$$

onde:

$\frac{\partial h_{INT}}{\partial P_c}$  é a matriz  $H_{INT}$  avaliada em  $P_c$ ,  $\frac{\partial h_{EXT}}{\partial x}$  representa a derivada da transformação extrínseca em relação ao estado  $x_k$ .

Como  $h_{EXT}$  é linear e depende apenas de  $P_x$  e  $P_y$ , e  $h_{INT}$  é avaliada em  $P_c$ , o jacobiano  $H$  resulta numa matriz  $2 \times 4$  com a forma:

$$H = \begin{bmatrix} \frac{\partial u}{\partial P_x} & \frac{\partial u}{\partial P_y} & 0 & 0 \\ \frac{\partial v}{\partial P_x} & \frac{\partial v}{\partial P_y} & 0 & 0 \end{bmatrix}, \quad (4.34)$$

onde os termos  $\frac{\partial u}{\partial P_x}$  e  $\frac{\partial v}{\partial P_y}$  são obtidos propagando as derivadas de  $h_{INT}$  através da matriz  ${}^c T_w$ , e as colunas correspondentes a  $V_x$  e  $V_y$  são nulas devido à independência da observação em relação à velocidade.

# Capítulo 5

## Resultados experimentais

No presente capítulo, proceder-se-á à apresentação dos resultados obtidos ao longo do trabalho, com vista à determinação da geolocalização de alvos marítimos. Para tal, é descrita a calibração da câmara, treinar um detetor de navios, realizar e testar a geolocalização e implementar o Filtro de Kalman para a geolocalização.

### 5.1 Calibração da câmara

Para determinar a calibração da câmara, foram utilizadas 34 imagens do tabuleiro xadrez com várias distâncias e ângulos, a Figura 5.1 são dois exemplos das imagens utilizadas.



FIGURA 5.1: Imagens utilizadas para calibração.

Os resultados da calibração da câmara, para imagens com resolução de  $2560 \times 1440$  píxeis, mostram distâncias focais  $f_x = 2655.1517$  e  $f_y = 2656.1203$ , com uma relação  $f_x/f_y \approx 0.9996$ . Esta proporção próxima de 1 indica que os píxeis do sensor são aproximadamente quadrados, o que evidencia a uniformidade do sistema ótico. O centro ótico encontra-se em  $(c_x, c_y) = (1385.3951, 639.5275)$ , deslocado em relação ao centro geométrico da imagem. A matriz dos parâmetros intrínsecos  $K$  da câmara

é dada por:

$$\begin{bmatrix} 2655.1517 & 0 & 1385.3951 \\ 0 & 2656.1203 & 639.5275 \\ 0 & 0 & 1 \end{bmatrix}.$$

A distorção radial da câmara apresenta um coeficiente  $k_1 = -0.28289802$ , indicando distorção do tipo barril, com  $k_2 = 0.33686289$  e  $k_3 = -0.78219721$ , que modelam distorções radiais de ordem superior (quarta e sexta ordens, respectivamente). Estes coeficientes indicam a necessidade de compensar efeitos pronunciados nas regiões periféricas da imagem, especialmente em aplicações que requerem alta precisão geométrica. As distorções tangenciais, com  $p_1 = 0.00040536$  e  $p_2 = 0.00245581$ , são mínimas, sugerindo um alinhamento adequado entre a lente e o sensor.

Estes parâmetros demonstram que a câmara é adequada para a aplicação em causa, uma vez que os píxeis quase quadrados simplificam projeções geométricas e as distorções radiais, embora significativas nas bordas, podem ser corrigidas com precisão utilizando os coeficientes determinados.

## 5.2 Aquisição de dados para geolocalização

Conforme mencionado anteriormente, todos os dados relevantes foram compilados durante um voo realizado em colaboração com a CEOV, conforme ilustrado na Figura 5.2.

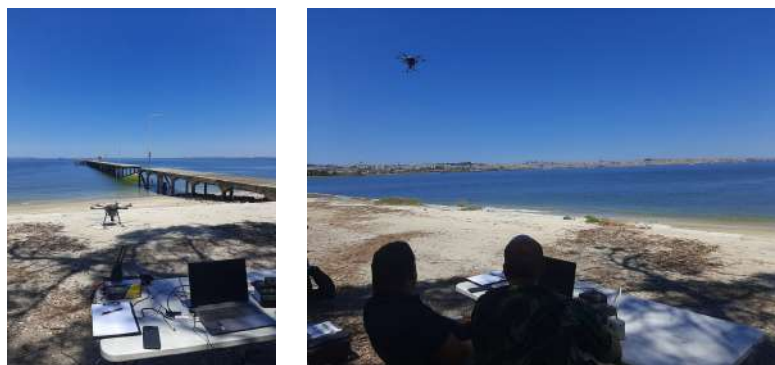


FIGURA 5.2: Ilustração de voo realizado com a CEOV.

A partir do conjunto UAV e câmara, foram obtidas informações sobre a posição e orientação do UAV, bem como a orientação da câmara. Essas informações estão detalhadas nas Tabelas 5.1 e 5.2, que apresentam os primeiros quatro conjuntos de dados adquiridos, respetivamente. A Figura 5.3 ilustra uma imagem capturada

## 5.2. Aquisição de dados para geolocalização

---

durante as filmagens. O vídeo completo utilizado nesta dissertação tem a duração de 2 minutos e 33 segundos e está disponível em <https://youtu.be/Bhsx7U9BHQ8>.

TABELA 5.1: Exemplo de dados de posição e orientação do UAV.

<i>Timestamp</i>	Lat	Lon	Alt (m)	Pitch (°)	Roll (°)	Yaw (°)
2025-06-26 13:36:09.428	38.657916	-9.130303	85.627	1.79	1.36	-10.42
2025-06-26 13:36:09.944	38.657916	-9.130303	85.63	2.0	1.29	-10.42
2025-06-26 13:36:10.460	38.6579159	-9.1303029	85.642	1.81	0.88	-10.48
2025-06-26 13:36:10.979	38.6579158	-9.1303029	85.658	1.82	0.9	-10.41

TABELA 5.2: Exemplo de dados de orientação da câmara.

<i>Timestamp</i>	Pitch (°)	Yaw (°)
2025-06-26 13:36:09.428	-8.4	-1.3
2025-06-26 13:36:09.944	-8.4	-1.4
2025-06-26 13:36:10.460	-8.4	-1.3
2025-06-26 13:36:10.979	-8.4	-1.4



FIGURA 5.3: Imagem capturada durante recolha de dados.

No período de captura das imagens aéreas, foram identificados seis navios: dois em navegação e quatro fundeados. Após a recolha de dados AIS, os navios de interesse serão os previamente referidos. É importante salientar que a cadência de transmissão dos sinais AIS, que é mais elevada para navios em navegação e reduzida para navios fundeados ou atracados, influenciou a disponibilidade de dados durante o período de observação.

Das posições dos quatro navios fundeados, três foram confirmadas com base nos dados AIS recolhidos. No entanto, para dois destes navios, os sinais AIS foram emitidos em momentos não coincidentes com o intervalo temporal do vídeo, dificultando a correspondência direta entre os dados visuais e os dados AIS. O quarto navio fundeado não foi detetado pela antena AIS, possivelmente devido à ausência de transmissão de sinal ou a limitações na cobertura da antena. Estes resultados

destacam a importância de sincronizar os dados AIS com as observações visuais para garantir uma identificação precisa dos navios.

Na Tabela 5.3 são apresentados os quatro primeiros dados adquiridos pela antena AIS. No âmbito deste estudo, apenas é pertinente considerar a posição dos navios.

TABELA 5.3: Exemplo de dados AIS obtidos durante testes práticos.

<i>Timestamp</i>	Ship	Lat	Lon	SOG	COG	Heading
12:36:08.000	Ship 5	38.663867	-9.109817	19.1	292.0	284
12:36:14.000	Ship 5	38.664067	-9.110433	19.0	292.0	284
12:36:16.000	Ship 6	38.674933	-9.127533	23.2	178.0	171
12:36:26.000	Ship 6	38.67385	-9.127383	23.4	174.0	168

### 5.3 Criação de *dataset* para geolocalização

Com base nos dados recolhidos, foi criado um *dataset* original, intitulado *Ship Geolocation*. Este conjunto de dados está disponível publicamente na plataforma *Kaggle*, em <https://www.kaggle.com/datasets/miguelbernardes/ship-geolocation/>.

O *dataset* inclui um ficheiro README, com a descrição detalhada do seu conteúdo, o vídeo captado pelo UAV, os dados AIS dos navios, os parâmetros intrínsecos da câmara e uma matriz no seguinte formato:

```
t lat lon h r11 r21 r31 0 r12 r22 r32 0 r13 r23 r33 0
```

t1 t2 t3 1 u1 v1 u2 v2 u3 v3 u4 v4 u5 v5 u6 v6 em que,  $t$  representa o tempo decorrido,  $lat$   $lon$   $h$  correspondem à latitude, longitude e altitude do UAV. De seguida, encontra-se a transformação homogénea do referencial do mundo para o referencial da câmara,  ${}^cT_w$ . Esta é determinada com base na Equação 4.5, onde, para cada referencial, é necessário efetuar as correspondentes translações, rotações e mudanças de eixo, quando aplicável. Por fim, as coordenadas  $u$  e  $v$  dos navios detetados. Estas foram determinadas mediante a utilização do *software* de anotação *Computer Vision Annotation Tool* (CVAT) (CVAT.ai Corporation, 2023), onde todos os navios foram anotados manualmente em cada *frame*. Na Figura 5.4 pode observar-se um exemplo da *ground truth* de uma imagem adquirida.



FIGURA 5.4: Exemplo da *Ground truth* de um frame.

## 5.4 Detecção de navios

O processo de deteção de navios foi treinado com base no *dataset Aerial Vessels Detection Dataset (Dataset A)* e no modelo *yolov8n* de Ultralytics, 2025. Para complementar e otimizar este ponto de partida, foram considerados os *datasets ship aerals 2 Computer Vision Dataset* e o *Aerial views of boats Computer Vision Model*, os quais serão sempre utilizados em conjunto e designados como *Dataset B*. Adicionalmente, empregou-se a versão *small* do *yolov8*.

A avaliação do modelo de deteção de objetos foi realizada mediante comparação com a *ground truth* das deteções. Foi definido um limiar de 0.5 referente ao valor mínimo de *Intersection over Union (IoU)*, uma métrica que avalia a precisão das deteções em tarefas de reconhecimento de objetos. O IoU mede a sobreposição entre a caixa delimitadora prevista pelo modelo e a caixa verdadeira (*ground truth*). Este valor é calculado como a razão entre a área de interseção e a área de união das duas caixas. Um limiar de IoU igual a 0.5 indica que uma deteção é considerada correta se a área de sobreposição for pelo menos 50% da área total da união. Este limiar é um padrão comum em tarefas de deteção de objetos, pois equilibra precisão e flexibilidade na avaliação.

A análise da deteção foi conduzida com base nas seguintes métricas: Verdadeiros Positivos (VP), que correspondem a deteções corretas de navios; Falsos Positivos (FP), que ocorrem quando um navio é detetado, mas não existe na realidade; Falsos Negativos (FN), que se referem a navios existentes que não foram detetados; Precisão (P), definida como a proporção de deteções corretas em relação

ao total de deteções realizadas; Sensibilidade (S), que representa a proporção de casos reais detetados em relação ao total de casos reais existentes; e o F1 *Score* (F1), que corresponde à média harmónica entre a precisão e a sensibilidade.

Na primeira fase de averiguação, realizou-se a comparação entre o *dataset* A, Figura 5.5A, e a combinação dos *datasets* A + B, Figura 5.5B. Esta etapa tem como objetivo complementar o *dataset* A, devido à elevada similaridade das imagens, pertencentes a *frames* consecutivos. Os resultados desta comparação são apresentados na Tabela 5.4.



FIGURA 5.5: Imagens da comparação entre *datasets* A e A+B.

TABELA 5.4: Comparação entre *datasets* A e A+B.

<i>Dataset</i>	VP	FP	FN	P	S	F1
A	2 762	371	20 673	0.8816	0.1179	0.2079
A+B	7 172	2 267	16 263	0.7598	0.306	0.4363

Como esperado, apesar da reduzida dimensão do *dataset* B, a sua heterogeneidade permite observar melhorias nas deteções. Constata-se um aumento de FP e uma diminuição de FN, o que se traduz num aumento de, aproximadamente, 0.23 no valor de F1 *Score*. Assim, no estudo seguinte será tido em consideração o conjunto de dados A + B, o qual obteve os melhores resultados.

O estudo subsequente consistirá na avaliação da influência do *data augmentation*, Tabela 5.5, conforme mencionado no Subcapítulo 4.5. Considerou-se esta ferramenta com o objetivo de enriquecer o *dataset*, no que toca a condições de brilho e ângulos de imagem. Neste contexto, a utilização da ferramenta mencionada é recomendável.

TABELA 5.5: Comparação entre com e sem *data augmentation*.

<i>Dataset</i>	VP	FP	FN	P	S	F1
Sem <i>data augmentation</i>	7 172	2 267	16 263	0.7598	0.306	0.4363
Com <i>data augmentation</i>	12 397	1 727	11 038	0.8777	0.5290	0.6601

Na Figura 5.6 é possível observar a diferença nas detecções da imagem. No treino sem *data augmentation*, apenas dois dos seis navios são detetados, Figura 5.6A. Enquanto no treino com *data augmentation*, quatro dos seis navios são detetados, Figura 5.6B.

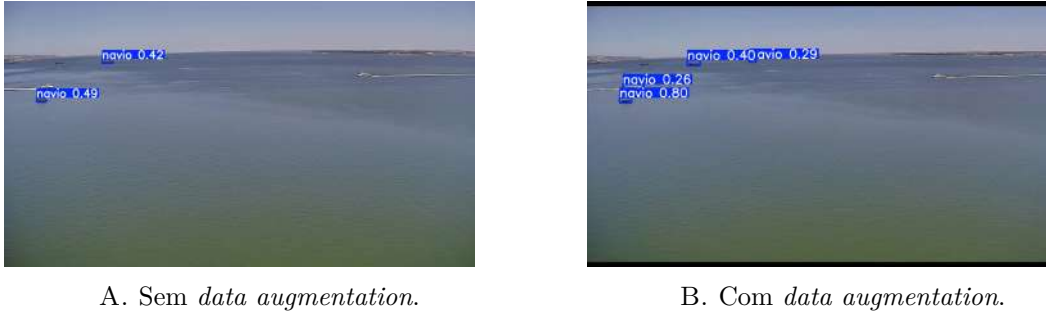


FIGURA 5.6: Imagens da comparação entre sem e com *data augmentation*.

Este estudo demonstra uma melhoria significativa em todos os parâmetros avaliados, nomeadamente um aumento de cerca de 0.23 no *F1 Score*, comprovando o efeito benéfico do *data augmentation*.

A averiguação final foi entre a versão *nano* e *small* do *yolov8*. Estes modelos pré-treinados foram selecionados por serem os que possuem menor poder computacional, o que se revelará importante para operar com UAVs. Esta comparação foi realizada com o *data augmentation* nos *datasets* A + B. Tabela 5.6.

TABELA 5.6: Comparação entre o *yolov8 nano* e *small*.

Modelo	VP	FP	FN	P	S	F1
Yolov8n	12 397	1 727	11 038	0.8777	0.5290	0.6601
Yolov8s	11 759	1 896	11 676	0.8611	0.5018	0.6341

Com base nos resultados obtidos, é possível concluir que, embora a versão *small* apresente uma ligeira superioridade em termos de desempenho de treino, quando utilizada nesta aplicação prática, esta versão obtém resultados inferiores à versão *nano*.

Conforme ilustrado na Figura 5.7, é possível observar a deteção de quatro navios em ambas as imagens. A versão *nano*, representada na Figura 5.7A, apresenta uma confiabilidade ligeiramente superior em todas as deteções em comparação com a versão *small*, Figura 5.7B. No entanto, a confiabilidade é muito semelhante ao longo do vídeo.



FIGURA 5.7: Imagens da comparação entre o *yolov8 nano* e *small*.

Perante isto, o detetor utilizado foi o *yolov8n* pré-treinado com os *datasets* A + B com *data augmentation*. Estas foram as combinações que obtiveram os melhores resultados para a deteção nas imagens capturadas.

## 5.5 Geolocalização

Na presente Secção são apresentados os resultados da geolocalização através da cinemática inversa e do EKF. A apresentação destes resultados ilustra as abordagens que se consideraram mais adequadas por forma a contribuir para a precisão da localização marítima e a sua utilidade em contextos operacionais reais. No entanto, é fundamental salientar que os dados processados apresentaram limitações. No decurso do processo de análise, foram detetados ruídos inerentes aos sensores e às deteções, assim como a sincronização entre os dados do UAV, da câmara e das deteções, contribuindo assim para um aumento da incerteza dos resultados. Por outro lado, o propósito do EKF é mitigar a ocorrência de atrasos e de assincronismos. Contudo, não se verificaram melhorias com a sua utilização. Por fim, será abordado o detetor criado para este efeito.

### 5.5.1 Cinemática inversa

Com o intuito de esclarecer o processo de geolocalização, os cálculos, realizados em MATLAB, foram inicialmente efetuados no sentido mundo  $\rightarrow$  imagem, cinemática direta, permitindo um esclarecimento passo a passo do procedimento. Este processo é validado por meio de testes com pontos possíveis de determinar por cálculo.

A partir desta abordagem, infere-se que a normalização dos pontos na imagem, em que o vetor é dividido pelo seu terceiro elemento,  $\lambda$ , não é aplicável na cinemática inversa (imagem  $\rightarrow$  mundo).

Na cinemática inversa, foi implementada a solução para o referido, apresentada nas Equações 4.14 e 4.16, onde  $\lambda$  surge como uma variável desconhecida. Contudo, este valor não é necessário para os cálculos da geolocalização, uma vez que as posições  $x$  e  $y$  do alvo são obtidas simultaneamente. Este método é verificado aplicando a cinemática direta.

### 5.5.2 Simulação

Recorrendo ao ambiente de programação de sistemas dinâmicos *Simulink* do MATLAB, foi desenvolvido um simulador para validar o processo de geolocalização, conforme ilustrado na Figura 5.8. Este está dividido em seis componentes: ambiente, navio, câmara, UAV, detecção e geolocalização.

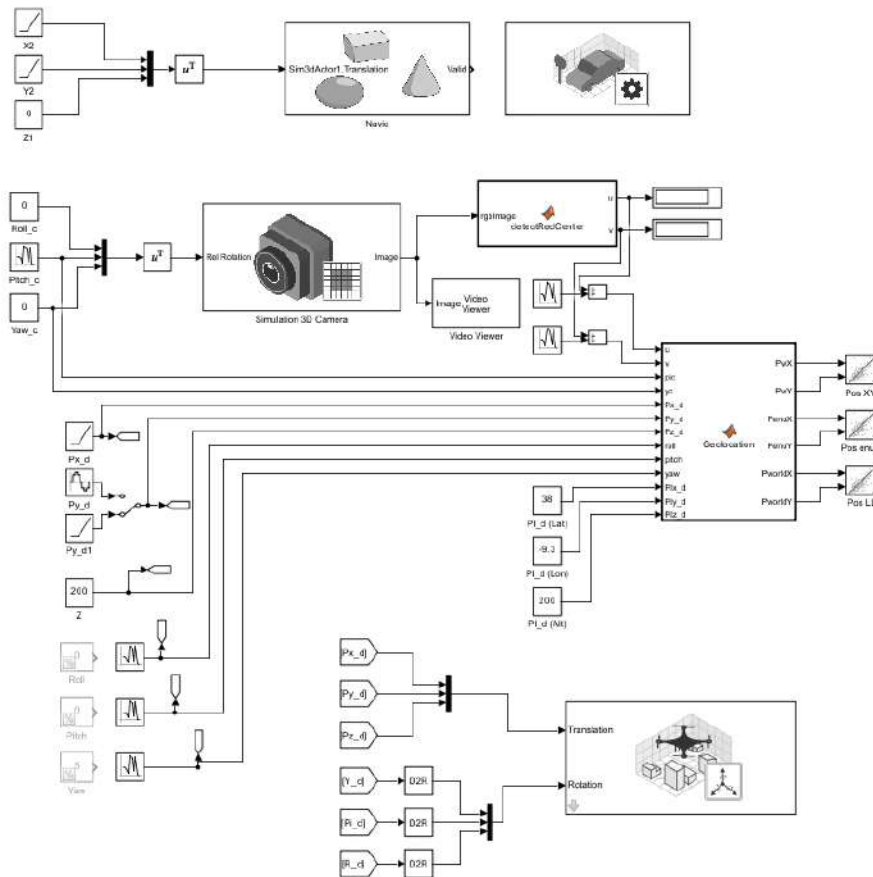
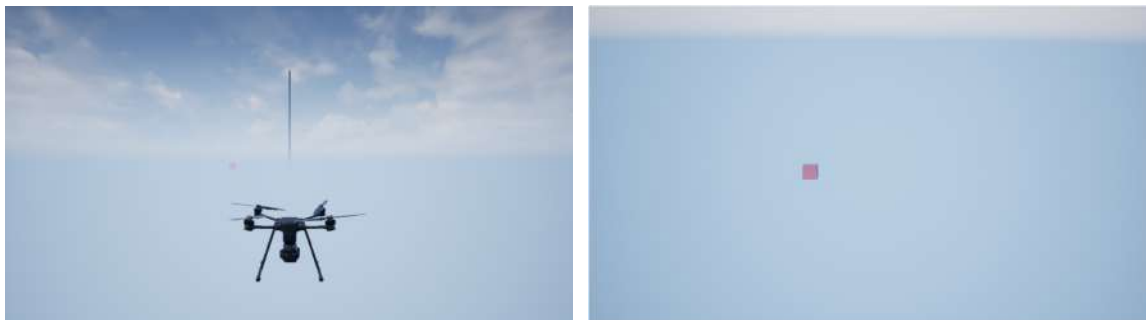


FIGURA 5.8: Modelo *Simulink* da simulação

Em relação ao ambiente de simulação, este utiliza a predefinição do *Simulink*, pois simula a realidade pretendida. Quanto ao UAV, foi selecionado um quadrotor com a câmara acoplada na sua parte inferior. Para os parâmetros da câmara, foram definidos os seguintes valores: distância focal [2872,078, 2871,261], centro ótico [960, 540], tamanho da imagem [1920, 1080] e distorção radial [-0,23813388,

0,14668279]. O navio é representado por um cubo vermelho com dimensões (20, 20, 20), facilitando a sua deteção, pois o detetor identifica a cor vermelha e calcula a média das coordenadas dos píxeis para obter o ponto central, conforme ilustrado na Figura 5.9.



A. Ambiente com UAV e 'navio'.

B. Imagem da câmara com 'navio'.

FIGURA 5.9: Simulação em *Simulink*.

Para esta simulação, foi definida uma duração de 20 segundos com uma taxa de amostragem de 0,1 segundos. O navio inicia-se na posição (1200, -200, 0) m com uma velocidade constante de (15, 10, 0) m/s, deslocando-se até à posição (1500, 0, 0) m. Por outro lado, o UAV inicia-se na posição (0, 0, 200) m com uma velocidade de (10, 15, 2) m/s, terminando na posição (200, 300, 240) m, e mantém um *yaw* fixo de 5 graus. Além disso, a câmara apresenta um *pitch* fixo de 10 graus negativos para capturar as imagens do navio.

Para realizar a geolocalização, foi implementada a metodologia de cinemática inversa, descrita na Subsecção anterior. Para tal, são necessários como dados de entrada os valores mencionados do UAV e da câmara, incluindo posição, *yaw* e *pitch*. Foi também definida uma posição de origem, correspondente ao pontão Norte da Praia da Cova do Vapor (38.6642434°N, -9.2619603°W), de forma a posicionar o navio no rio Tejo e proporcionar uma noção espacial do referencial da simulação.

Na Figura 5.10, observam-se os resultados da cinemática inversa, com a posição estimada sobreposta à posição real, apresentando erros *Root Mean Square Error (RMSE)* e *Mean Absolute Error (MAE)* de 5.01 m e 4.97 m, respetivamente, calculados conforme o Apêndice C.

De igual forma, nesta simulação, foi realizada a aplicação do EKF para estimar a posição do navio, utilizando as observações da câmara e o modelo dinâmico do navio, com o ponto inicial determinado por cinemática inversa. As observações, correspondentes às coordenadas ( $u, v$ ) do cubo vermelho na imagem, incorporam um ruído gaussiano branco com desvio padrão de 30 píxeis, modelado pela matriz

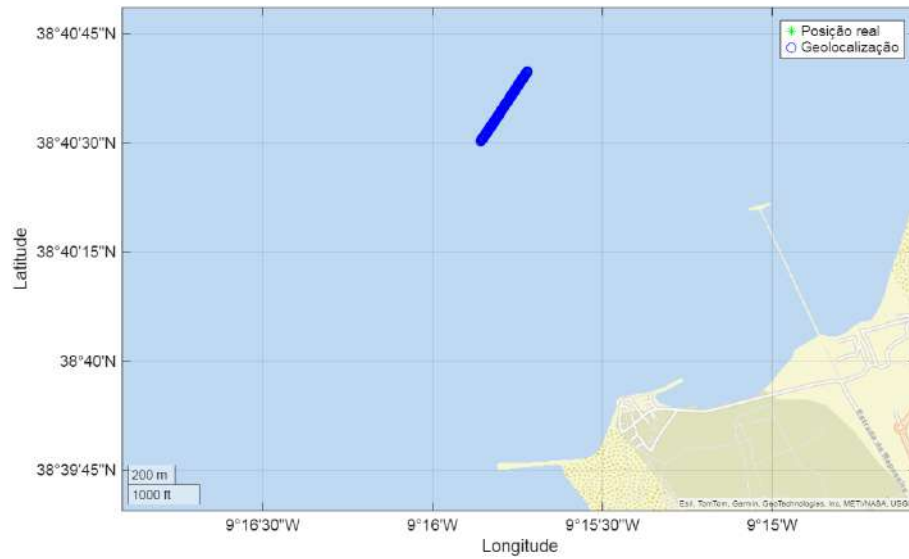


FIGURA 5.10: Simulação da cinemática inversa.

de covariância  $R = 30^2 \cdot \text{eye}(2)$ . O modelo dinâmico assume movimento de velocidade constante, com estado composto por posição  $(x, y)$  e velocidade  $(v_x, v_y)$ , onde a matriz de transição  $F$  integra a velocidade na posição ao longo do intervalo de tempo  $T = 0.1$  segundos. A incerteza do modelo é representada pela matriz de covariância do processo  $Q$ , que incorpora uma variância de aceleração aleatória de  $\sigma_{\text{acc}}^2 = 0.25 \text{ (m/s}^2\text{)}^2$ . Não é esperado observar melhoria com o EKF, pois este considera ruído branco, enquanto a cinemática inversa não incluía ruído. Os resultados obtidos, conforme ilustrado na Figura 5.11, mostram que a estimativa do EKF apresenta ligeiro desvio em relação à posição real, com erros *Root Mean Square Error* (*RMSE*) de 37,26 m e *Mean Absolute Error* (*MAE*) de 37,25 m, valores considerados aceitáveis.

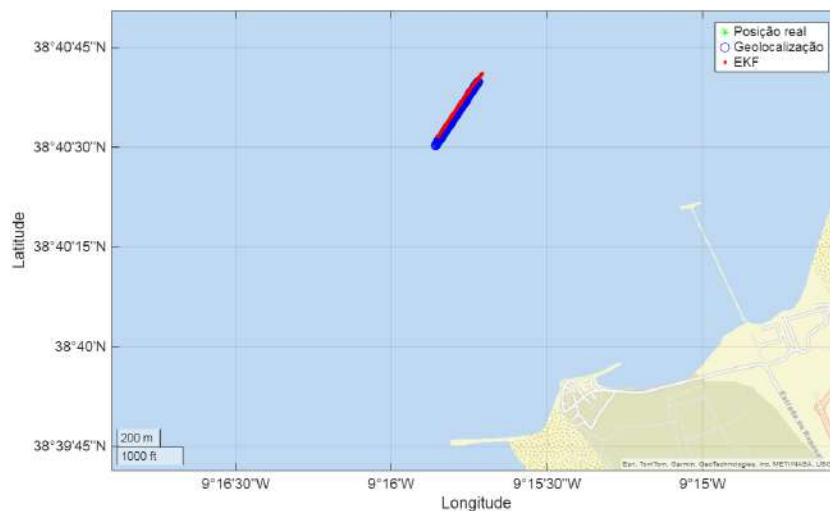


FIGURA 5.11: Simulação da cinemática inversa e EKF.

Nesse sentido, foram introduzidos ruídos na pose do UAV, na câmara e na detecção, com o objetivo de aproximar os dados à realidade. A Figura 5.12 ilustra, dois gráficos de latitude e longitude. A primeira, a Figura 5.12A, mostra uma trajetória nítida do navio, sem ruído. A segunda, a Figura 5.12B, exibe uma dispersão da trajetória estimada, devido a ruídos gaussianos com variâncias de  $0,5 \text{ m}^2$  para a pose do UAV e  $7 \text{ píxeis}^2$  para a detecção.

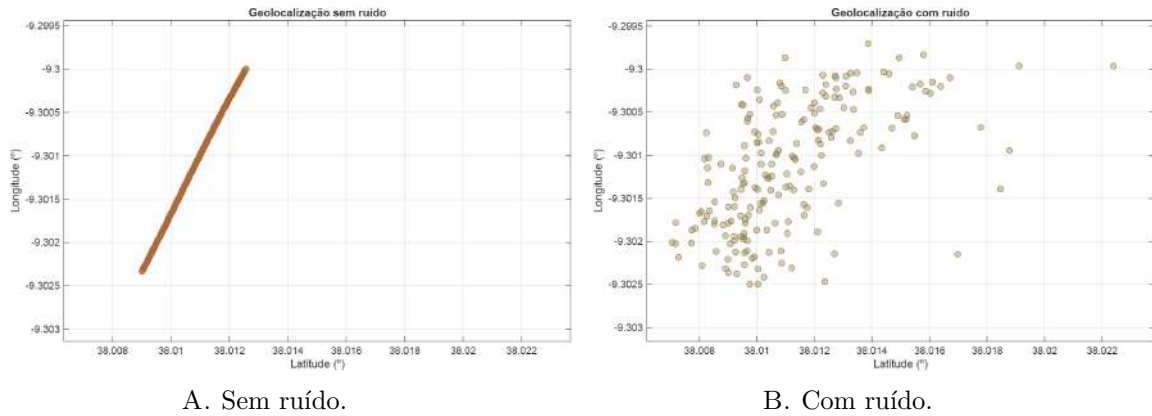


FIGURA 5.12: Resultado da geolocalização em *Simulink*.

Seguindo a simulação anterior, foram introduzidos ruídos adicionais na pose do UAV, na câmara e na detecção, com o objetivo de avaliar a robustez da geolocalização face a condições mais realistas. Os resultados obtidos, conforme ilustrado na Figura 5.13, mostram uma estimativa da cinemática inversa localizada a cerca de  $140\,000 \text{ m}$ , resultando em erros *Root Mean Square Error (RMSE)* de  $9766.24 \text{ m}$  e *Mean Absolute Error (MAE)* de  $967.51 \text{ m}$  em relação à posição real do navio. Com base nisto, foram anulados todos os pontos estimados a mais de  $20 \text{ km}$ , utilizando um limiar ajustável conforme o contexto da simulação.



FIGURA 5.13: Simulação com ruído da cinemática inversa e EKF.

Após a aplicação do limiar de erro à simulação, a cinemática inversa reduz significativamente os seus erros. Conforme apresentado na Tabela 5.7, os erros de geolocalização da cinemática inversa e do EKF aumentam significativamente em relação aos obtidos sem ruído, devido ao ruído intenso introduzido. Na Figura 5.14, ilustra-se o mapa das posições estimadas em comparação com a trajetória real do navio.

TABELA 5.7: Erros de geolocalização da simulação.

Método	RMSE (m)	MAE (m)
Cinamática inversa	617.86	277.98
EKF	1825.85	1641.88

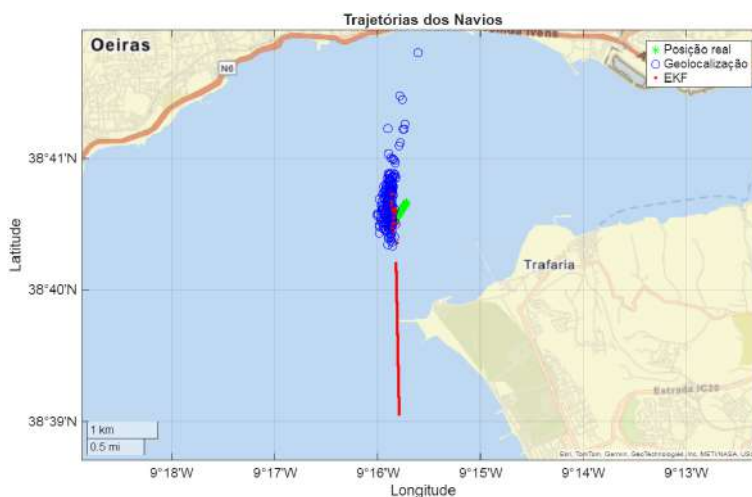


FIGURA 5.14: Simulação com ruído da cinemática inversa vs EKF com limitação do erro.

### 5.5.3 Aplicação a dados reais

Esta secção apresenta a aplicação da cinemática inversa e do EKF a dados reais para validar a geolocalização de navios em cenários reais. De forma a aplicar a cinemática inversa a dados reais, foi utilizada a matriz mencionada na Secção 5.3, contendo todos os dados necessários, exceto os parâmetros intrínsecos da câmara, que devem ser tão rigorosos quanto possível. Os resultados obtidos, utilizando a *ground truth* das deteções para construir a matriz, são apresentados na Figura 5.15.

Conforme apresentado na Figura 5.16, observam-se as posições reais dos navios em análise. Sabe-se, no entanto, que o “Ship 1” não possui dados AIS, mas encontra-se aproximadamente na posição  $38.6654000^{\circ}\text{N}$ ,  $-9.1225000^{\circ}\text{W}$ , determinada por observação visual.

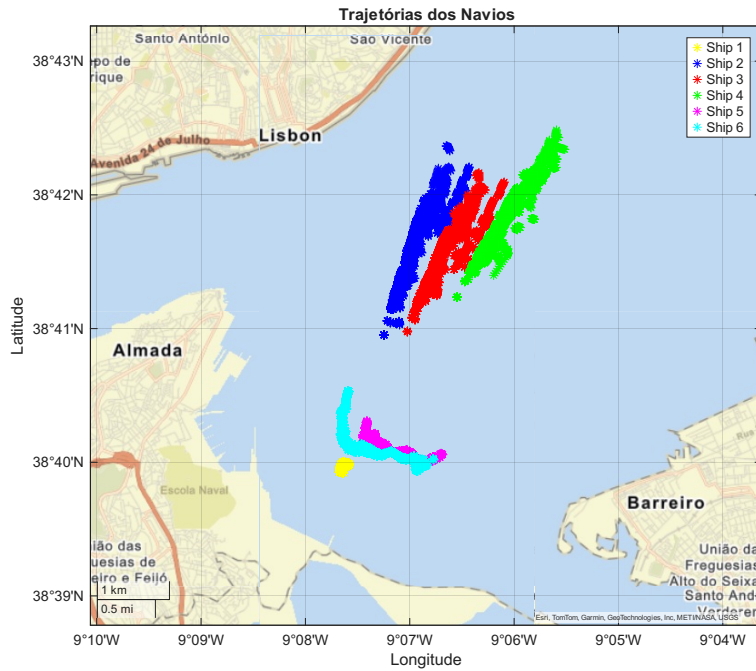


FIGURA 5.15: Geolocalização com a *ground truth* das detecções.

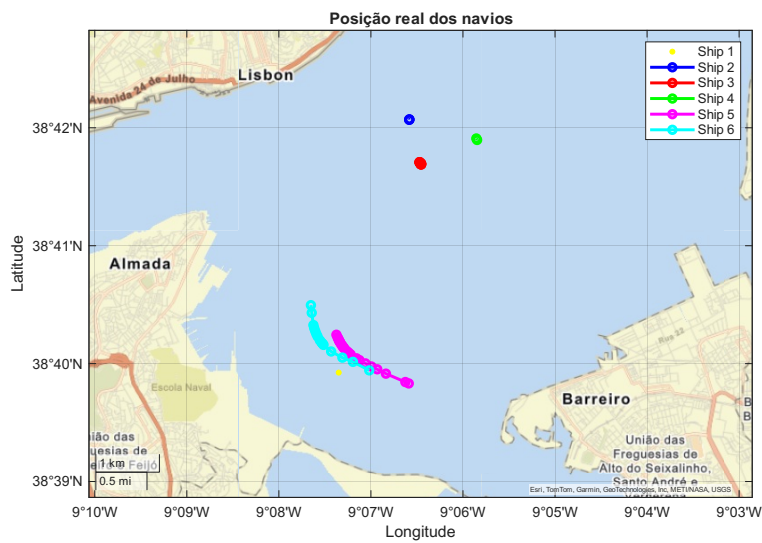


FIGURA 5.16: Posição real dos navios.

Para analisar a exatidão e precisão dos resultados, compararam-se as posições estimadas com as posições reais. Verifica-se que é possível estimar a posição dos navios com alguma exatidão (proximidade ao valor real), mas com baixa precisão (consistência dos resultados), especialmente para os navios 2, 3 e 4, situados a cerca de 5 km, próximos do horizonte da imagem (distância teoricamente infinita), conforme se observa na Figura 5.4. Essa distância é a principal causa, uma vez que, para os navios 1, 5 e 6, localizados a cerca de 1,5 km do UAV, observa-se uma

melhoria significativa na precisão.

Conforme ilustrado na Figura 5.17, a variação dos erros da geolocalização é significativamente maior para os navios mais distantes, assim como a média dos mesmos. Estes resultados são confirmados na Tabela 5.8, que quantifica os erros para os navios 2 a 6, excluindo o navio 1 devido à ausência de dados AIS, confirmando esta tendência.

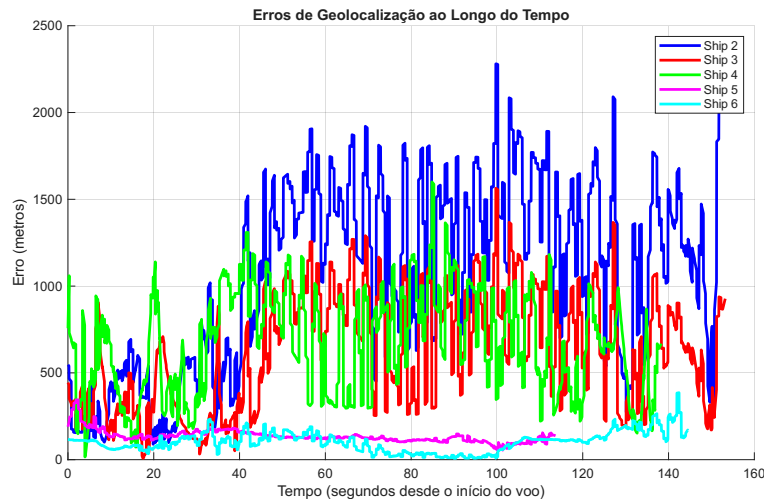


FIGURA 5.17: Erro da geolocalização pelo método da cinemática inversa referente a cada navio.

TABELA 5.8: Erros da geolocalização referente a cada navio.

Ship	RMSE (m)	MAE (m)
Ship 2	1195.39	1066.81
Ship 3	711.34	628.47
Ship 4	735.21	671.24
Ship 5	140.50	135.34
Ship 6	115.74	100.72

Na tentativa de melhorar a geolocalização, foi implementado o EKF em dados reais, utilizando o mesmo modelo dinâmico de velocidade constante utilizado na simulação, esperava-se uma melhoria na geolocalização em relação à cinemática inversa. Contudo, devido a navios fundeados (velocidade nula), o modelo do EKF não processa adequadamente essa dinâmica. Surpreendentemente, também não se observaram melhorias nos navios em movimento, possivelmente devido a ruídos nos dados ou limitações do modelo, ao contrário do esperado. Conforme ilustrado na Figura 5.18, as posições estimadas pelo EKF são praticamente idênticas às da cinemática inversa, sem diferenças significativas.

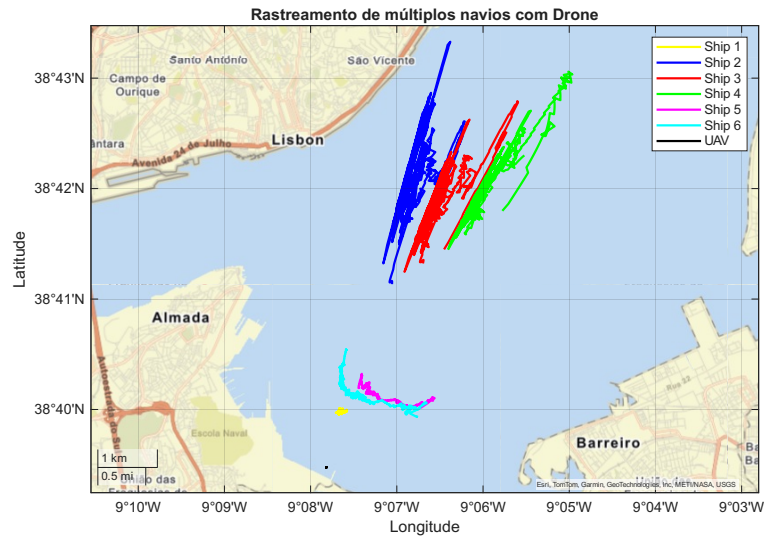


FIGURA 5.18: Geolocalização com EKF.

Estes resultados são confirmados pela variação dos erros ao longo do tempo, conforme ilustrado na Figura 5.19, e pelos erros médios de cada navio, apresentados na Tabela 5.9. Embora se observe uma melhoria significativa para o navio 2, com redução do *RMSE* de 1195,39 m para 730,36 m, aproximando-se dos erros dos navios 3 e 4, para os restantes navios não se constata alterações significativas em relação à cinemática inversa. Além disso, melhorias ligeiras foram observadas no *MAE* dos navios 3 e 4, com reduções para 619,28 m e 640,91 m, respetivamente.

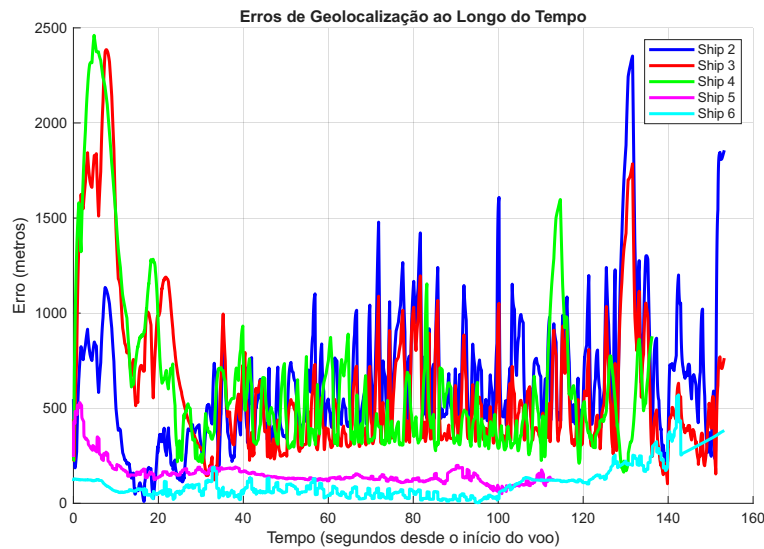


FIGURA 5.19: Erro da geolocalização pelo método do EKF referente a cada navio.

TABELA 5.9: Erros da geolocalização com EKF referente a cada navio.

Ship	RMSE (m)	MAE (m)
Ship 2	730.36	637.50
Ship 3	745.03	619.28
Ship 4	788.60	640.91
Ship 5	168.10	154.94
Ship 6	143.21	110.18

Por fim, foi avaliada a geolocalização utilizando as deteções do modelo treinado, referido na secção anterior, em substituição à *ground truth* das deteções, implementando a cinemática inversa como método de estimação. Para tal, foi gerada uma matriz com as coordenadas de deteção  $(u, v)$  mencionadas. Devido à ausência de seguimento dos navios, não é possível a distinção entre eles, resultando na inclusão de todas as deteções, incluindo falsos positivos. Conforme ilustrado na Figura 5.20, todas as posições são visualizadas sem distinção. Os pontos mais à direita, próximos do horizonte, correspondem a falsos positivos associados a estruturas costeiras no canto superior direito da Figura 5.4. Uma vez que o horizonte representa uma distância teoricamente infinita, essas posições estimadas resultam em valores extremamente longínquos.



FIGURA 5.20: Geolocalização com navios detetados.

Complementando a análise anterior, os resultados do modelo treinado proporcionam um panorama claro da área de interesse, conforme ilustrado na Figura 5.21. A análise temporal das deteções, considerando *frames* consecutivos do vídeo, permitiu identificar a presença dos seis navios distintos, embora menos perceptível na

figura devido à sobreposição de pontos. Além disso, foram detetados dois veleiros, visíveis como pontos mais à esquerda na figura, presentes no início do vídeo.

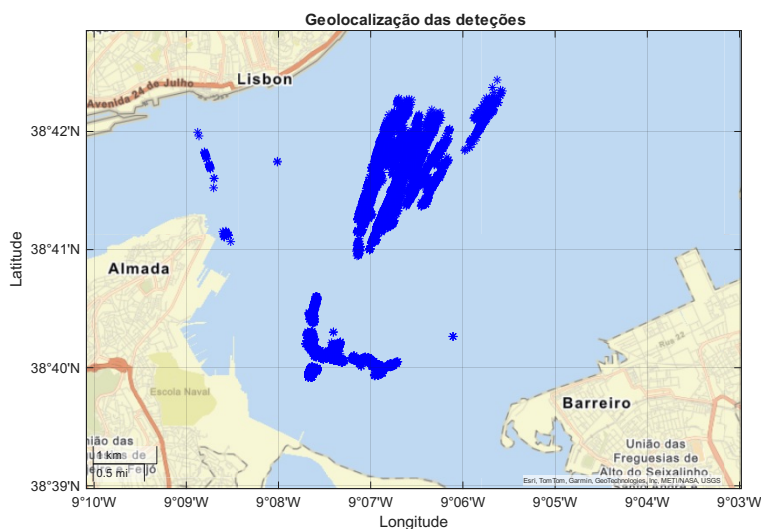


FIGURA 5.21: Geolocalização com navios detetados aproximada à área de interesse.

A aplicação do EKF às deteções do modelo treinado não é viável, devido à ausência de seguimento dos navios e à falta de dados temporais contínuos, necessários para o modelo dinâmico de velocidade constante descrito na Subsecção 5.5.2. Assim, não foi possível realizar estudos com o EKF para estas deteções.

## 5.6 Discussão de resultados e dificuldades encontradas

Na presente secção são abordados os principais resultados, com o propósito de serem explicados e enquadrados no contexto do trabalho. Ademais, são igualmente mencionadas as dificuldades experimentadas no decorrer do trabalho.

### 5.6.1 Calibração da câmara

Para além da calibração da câmara apresentar uma elevada precisão, é igualmente necessário que esta seja exata. No decurso do processo de dissertação, emergiram múltiplos valores resultantes da calibração. A mera substituição do *software* de calibração foi suficiente para a alteração significativa de valores, bem como a adição ou subtração de uma imagem do tabuleiro xadrez. Inicialmente, a geolocalização não se encontrava em conformidade com o esperado, com erros na ordem dos 650 m para navios próximos e 4000 m para os navios distantes. Após verificação, constatou-se

que a origem do erro residia na calibração, constituindo um passo fundamental desta dissertação. A solução encontrada foi a adição de mais imagens para a calibração, o que resultou numa alteração significativa dos valores da calibração, com erros na ordem dos 120 m e 700 m, respetivamente.

### 5.6.2 *Datasets e deteção*

A procura por um *dataset* que correspondesse às condições pretendidas revelou-se um desafio. Em virtude da especificidade requerida, o objetivo consistia em obter imagens de navios capturadas por um veículo aéreo não tripulado (UAV), sem um exagero de *zoom*, com uma variedade de ângulos e brilhos das imagens, bem como de tipologia dos navios.

Num primeiro momento, *dataset* inicial, não se mostrou compatível com a realidade pretendida. De facto, os resultados obtidos no treino com o mesmo conjunto de dados foram manifestamente positivos, uma vez que as imagens de treino e validação eram muito semelhantes. Diante do exposto, tornou-se imperativo integrar outros *datasets*, com o propósito de ampliar a diversidade do conjunto original.

Em conformidade com o anteriormente exposto, optou-se pela utilização das imagens capturadas para a validação dos resultados dos treinos das deteções. Desta forma, assegura-se a eficácia nas imagens pretendidas. De acordo com o esperado, as deteções revelaram-se muito semelhantes à *ground truth*, conforme evidenciado pelos gráficos da geolocalização. O F1 Score de 0.66, alcançado com *datasets* complementares e *data augmentation*, demonstra uma deteção eficaz em cenários reais, superando as limitações iniciais como falsos positivos em condições de brilho elevado ou ângulos variados. Contudo, esta métrica sugere oportunidades de melhoria, num domínio tão vasto e em constante evolução como a visão computacional, impulsionada por avanços em inteligência artificial. A utilização de versões mais recentes e complexas do YOLO poderia otimizar o desempenho em ambientes marítimos dinâmicos, alinhando os resultados com *datasets*, como o MarSyn, que apresentam resultados superiores em condições controladas.

### 5.6.3 Geolocalização

Entre as principais dificuldades identificadas na geolocalização, destaca-se a presença do horizonte na imagem, que corresponde a uma distância teoricamente infinita, resultando num aumento significativo do erro à medida que os navios se

aproximam dessa região, pois um erro mínimo nas detecções pode gerar desvios de dezenas de quilômetros. Além disso, outro desafio significativo reside na posicionamento do centro das BB, que tipicamente se localiza acima do ponto desejado, uma vez que a posição real dos navios corresponde ao centro da área em contacto com a água. Esta discrepância, ilustrada na Figura 5.22, onde os pontos vermelhos representam o centro das detecções e os pontos azuis a posição real desejada, introduz erros na geolocalização, sublinhando a necessidade de ajustar o ponto de referência utilizado nos cálculos.



FIGURA 5.22: Centro das *Bounding Box*.

Como solução imediata para os problemas identificados, o aumento da altitude do UAV ou a inclinação da câmara para baixo podem reduzir o erro, uma vez que tal medida distancia o horizonte dos navios, ampliando o campo de visão útil e diminuindo a influência das medições. Esta abordagem requer, contudo, a realização de testes adicionais para avaliar o impacto na geolocalização.

Outro fator que contribuiu para os erros observados foi a ocorrência de oscilações no UAV, resultantes de imprecisões nos sensores, nomeadamente o GPS e a IMU. Estas variações afetaram a estabilidade dos dados de pose, introduzindo desvios nas estimativas de posição, especialmente em alvos distantes. Para mitigar este efeito, seria necessário proceder à melhoria da calibração dos sensores ou integrar sistemas de estabilização mais robustos.

A sincronização temporal entre os *frames* captados pela câmara e os dados de tempo foi realizada, porém, surgiram desafios durante o processamento, devido ao corte da parte do vídeo pretendida. Este processo potencialmente introduziu assincronismos que comprometeram a precisão da geolocalização, particularmente

quando os dados de pose e as imagens não estavam perfeitamente alinhados no tempo. A combinação destes fatores, com as limitações associadas ao horizonte e ao posicionamento das BB, resultou num erro médio de geolocalização de aproximadamente 120 m para alvos próximos e 800 m para alvos distantes, conforme verificado nos testes reais.

A aplicação do EKF para otimizar a consistência das estimativas não produziu os resultados esperados. Embora se tenha verificado uma redução do ruído sensorial em alvos distantes, o impacto foi limitado, com a persistência de erros elevados (700 m para os navios distantes, conforme os dados experimentais). Esta performance inferior pode estar relacionada com o facto de existirem navios fundeados, uma vez que o EKF foi concebido para modelos de velocidade constante.



# Capítulo 6

## Conclusões e Trabalho Futuro

O capítulo final desta dissertação descreve as conclusões que sintetizam todo o trabalho realizado (Secção 6.1) e explora propostas para possíveis trabalhos futuros (Secção 6.2).

### 6.1 Conclusões

Esta dissertação demonstrou a viabilidade de uma metodologia para a geolocalização de alvos marítimos utilizando uma câmara RGB montada num UAV, alinhando-se com os objetivos propostos de detetar e localizar navios. Através da calibração da câmara, do treino de modelos de deteção baseados em YOLO versão 8 *nano* e da aplicação de transformações de coordenadas, foi possível obter posições geográficas, contribuindo para a vigilância marítima em áreas extensas como a ZEE portuguesa.

Os resultados destacam a eficácia da deteção, com melhorias significativas ao combinar *datasets* e técnicas de *data augmentation*, alcançando um F1 *Score* de 0.66 e reduzindo falsos negativos. A geolocalização pela cinemática inversa revelou precisão razoável para navios a distâncias inferiores a 2 km (erro médio de 120 m), mas demonstrou limitações em alvos distantes, superiores a 5 km (erro médio de 800 m), atribuíveis a ruído proveniente dos sensores, calibração da câmara e aproximação ao horizonte. A integração do EKF mitigou ligeiramente esses erros em alguns casos, melhorando a consistência das trajetórias.

Em termos de contribuições, esta abordagem representa uma solução económica e autónoma para operações de busca e salvamento ou combate à pesca ilegal, complementando os radares costeiros. No entanto, as limitações identificadas, como a dependência de calibrações precisas e a sensibilidade à distância (variações mínimas causam erros elevados), sublinham a necessidade de refinamentos. Os testes

em simulação validaram o modelo teórico, enquanto os dados reais confirmaram a aplicabilidade prática, com erros médios globais de cerca de 500 m. Revelando-se um erro aceitável para determinar o panorama situacional de uma determinada área.

A criação de um *dataset* constitui outra das contribuições, na medida em que foi desenvolvido um conjunto de dados original e específico para a deteção de navios em imagens aéreas captadas por UAV, baseado em recolhas reais efetuadas no Mar da Palha, no estuário do rio Tejo. Este conjunto de dados colmata uma lacuna em recursos para o desenvolvimento de aplicações de deteção e geolocalização marítima, promovendo avanços em visão computacional e vigilância marítima.

Em síntese, os objetivos foram atingidos, provando que UAVs equipados com sensores acessíveis podem fornecer informação situacional crítica. Esta investigação abre caminhos para uma maior autonomia na Marinha Portuguesa, promovendo a integração de tecnologias não tripuladas em missões estratégicas.

## 6.2 Trabalho Futuro

A presente dissertação, que descreve um método de geolocalização de alvos marítimos, contém diversos pontos passíveis de desenvolvimento em trabalhos futuros. O UAV constitui um dos elementos fundamentais da dissertação, uma vez que é nesse que se baseiam todos os dados necessários. Assim, seria vantajoso realizar todo o processamento da geolocalização durante a operação do UAV, enviando para o operador apenas a localização dos alvos. Esta medida resultaria numa diminuição dos erros, visto que a discrepância de sincronismo entre todos os elementos relevantes sofreria uma redução significativa. A integração das posições enviadas pelo UAV com as do AIS permitiria identificar os navios que possuem o AIS desativado de forma propositada. Adicionalmente, seria possível visualizar uma imagem do navio em causa.

# Bibliografia

- Agency, United States. Defense Mapping (1987). *Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems*. Vol. 8350. Defense Mapping Agency.
- Akiyama, Tatsuhiro, Yosuke Kobayashi, Jay Kishigami e Kenji Muto (2018). «CNN-Based Boat Detection Model for Alert System Using Surveillance Video Camera». Em: *IEEE 7th Global Conference on Consumer Electronics (GCCE)*. DOI: 10.1109/GCCE.2018.8574704.
- Antonini, Amado, Winter Guerra, Varun Murali, Thomas Sayre-McCord e Sertac Karaman (set. de 2020). «The Blackbird UAV dataset». Em: *International Journal of Robotics Research* 39 (10-11), pp. 1346–1364. ISSN: 17413176. DOI: 10.1177/0278364920908331.
- Babinec, Adam e Jiří Apeltauer (out. de 2016). «On accuracy of position estimation from aerial imagery captured by low-flying UAVs». Em: *International Journal of Transportation Science and Technology* 5 (3), pp. 152–166. ISSN: 20460449. DOI: 10.1016/j.ijtst.2017.02.002.
- Badrudin, Arif, Siswo Hadi Sumantri, Rudy Agus Gemilang Gultom, I Nengah, Putra Apriyanto, Hendy Risdianto Wijaya e Indraajani Sutedja (2023). «Ship Trajectory Prediction for Anomaly Detection Using AIS Data and Artificial Intelligence: a Systematic Literature Review». Em: *Journal of Theoretical and Applied Information Technology* 31 (16). ISSN: 1817-3195. URL: [www.jatit.org](http://www.jatit.org).
- Cacace, Jonathan, Bartomeu Rubi, Javier Rodriguez e Julian Cayero (2024). «Precise Multi-Target Detection and Geolocalization using Unmanned Aerial Vehicles Supporting Surveillance Operations». Em: *IEEE International Conference on Automation Science and Engineering*. IEEE Computer Society, pp. 1777–1782. ISBN: 9798350358513. DOI: 10.1109/CASE59546.2024.10711373.
- Carrillo-Perez, Borja, Sarah Barnes e Maurice Stephan (abr. de 2022). «Ship Segmentation and Georeferencing from Static Oblique View Images». Em: *Sensors* 22 (7). ISSN: 14248220. DOI: 10.3390/s22072713.

- Chen, Haifeng, Jinlei Zhuang, Bingyou Liu, Lichao Wang e Luxian Zhang (2023). *Camera calibration method based on circular array calibration board*. DOI: 10.1080/21642583.2023.2233562.
- CVAT.ai Corporation (2023). *CVAT: Computer Vision Annotation Tool*. Versão v2.4.8. DOI: 10.5281/zenodo.8070041. URL: <https://github.com/opencv/cvat>.
- Damas, Bruno, Nuno Pessanha Santos e Matilde Correia Vieira (2024). «Ship, Guide My Landing! A Binocular Extended Kalman Filter for Maritime UAV Landing Operations». Em: *Oceans Conference Record (IEEE)*. Institute of Electrical e Electronics Engineers Inc. ISBN: 9798331540081. DOI: 10.1109/OCEANS55160.2024.10753980.
- Direção-Geral de Recursos Naturais, Segurança e Serviços Marítimos (2025). *Zonas Marítimas sob Jurisdição e ou Soberania Nacional*. Acessado em 30 de maio de 2025. URL: <https://www.dgrm.pt/am-ec-zonas-maritimas-sob-jurisdicao-ou-soberania-nacional>.
- Eling, C, L Klingbeil e H Kuhlmann (2014). «A Precise Direct Georeferencing System for UAVs». Em: *Proceedings of the Workshop on UAV-basaed Remote Sensing Methods for Monitoring Vegetation. Kölner Geographische Arbeiten*. DOI: 10.5880/TR32DB.KGA94.6.
- Farahnakian, Fahimeh, Mohammad Hashem Haghbayan, Jonne Poikonen, Markus Laurinen, Paavo Nevalainen e Jukka Heikkonen (jul. de 2018). «Object Detection Based on Multi-sensor Proposal Fusion in Maritime Environment». Em: *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*. Institute of Electrical e Electronics Engineers Inc., pp. 971–976. ISBN: 9781538668047. DOI: 10.1109/ICMLA.2018.00158.
- Fonder, Michaël e Marc Van Droogenbroeck (jun. de 2019). *Mid-Air: A multi-modal dataset for extremely low altitude drone flights*. URL: <http://midair.ulg.ac.be>.
- Geers, Lisa Yvette (2023). *MSc thesis in Geomatics Localising objects with drones: A case study on the localisation of fisher boats in restricted areas*.
- Gibbins, Danny, Philip Roberts e Leszek Swierkowski (dez. de 2004). «A Video Geolocation and Image Enhancement Tool for Small Unmanned Air Vehicles (UAVs)». Em: *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*. IEEE. ISBN: 0-7803-8894-1. DOI: 10.1109/ISSNIP.2004.1417506.

- Girshick, Ross, Jeff Donahue, Trevor Darrell e Jitendra Malik (jan. de 2016). «Region-Based Convolutional Networks for Accurate Object Detection and Segmentation». Em: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (1), pp. 142–158. ISSN: 01628828. DOI: 10.1109/TPAMI.2015.2437384.
- Hata, Kenji e Silvio Savarese (2017). *CS231A Course Notes 1: Camera Models*.
- He, Fangning, Tian Zhou, Weifeng Xiong, Seyyed Meghdad Hasheminasab e Ayman Habib (dez. de 2018). «Automated aerial triangulation for UAV-based mapping». Em: *Remote Sensing* 10 (12). ISSN: 20724292. DOI: 10.3390/rs10121952.
- Heikkinen, Markku (1982). «Geschlossene Formeln Zur Berechnung Räumlicher Geodätischer Koordinaten Aus Rechtwinkligen Koordinaten». Em: ISSN: 0340-4560. URL: <http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL82X0285851>.
- Helgesen, Håkon Hagen, Frederik Stendahl Leira, Torleiv H. Bryne, Sigurd M. Albrektsen e Tor Arne Johansen (ago. de 2019). «Real-time georeferencing of thermal images using small fixed-wing UAVs in maritime environments». Em: *ISPRS Journal of Photogrammetry and Remote Sensing* 154, pp. 84–97. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2019.05.009.
- Jiang, Peiyuan, Daji Ergu, Fangyao Liu, Ying Cai e Bo Ma (2021). «A Review of Yolo Algorithm Developments». Em: *Procedia Computer Science*. Vol. 199. Elsevier B.V., pp. 1066–1073. DOI: 10.1016/j.procs.2022.01.135.
- Jiao, Licheng, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng e Rong Qu (2019). «A survey of deep learning-based object detection». Em: *IEEE Access* 7, pp. 128837–128868. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2939201.
- Johnston, Michael G (2006). *Ground Object Geo-location Using UAV Video Camera*.
- Kalman, Rudolph Emil (1960). «A new approach to linear filtering and prediction problems». Em: *Journal of Basic Engineering*. DOI: <https://doi.org/10.1115/1.3662552>.
- Krizhevsky, Alex, Ilya Sutskever e Geoffrey E Hinton (2012). «ImageNet Classification with Deep Convolutional Neural Networks». Em: *Advances in neural information processing systems* 25.
- Kurekin, Andrey A, Benjamin R Loveday, Oliver Clements, Graham D Quartly, Peter I Miller, George Wiafe e Kwame Adu Agyekum (2019). «Operational monitoring of illegal fishing in Ghana through exploitation of satellite earth observation and AIS data». Em: *Remote Sensing* 11.3, p. 293.

- L. e M.A. Sahawneh Jarrah (2008). «Development and calibration of low cost MEMS IMU for UAV applications». Em: *2008 5th International Symposium on Mechatronics and Its Applications*. Ed. por 2008 5th International Symposium on Mechatronics e Its Applications. DOI: 10.1109/ISMA.2008.4648819.
- Li, Shenglin e Hwan Sik Yoon (abr. de 2023). «Vehicle Localization in 3D World Coordinates Using Single Camera at Traffic Intersection». Em: *Sensors* 23 (7). ISSN: 14248220. DOI: 10.3390/s23073661.
- Liu, Li, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu e Matti Pietikäinen (fev. de 2020). «Deep Learning for Generic Object Detection: A Survey». Em: *International Journal of Computer Vision* 128 (2), pp. 261–318. ISSN: 15731405. DOI: 10.1007/s11263-019-01247-4.
- Liu, Ryan Wen, Weiqiao Yuan, Xinqiang Chen e Yuxu Lu (set. de 2021). «An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system». Em: *Ocean Engineering* 235. ISSN: 00298018. DOI: 10.1016/j.oceaneng.2021.109435.
- Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu e Alexander C Berg (2016). «Ssd: Single shot multibox detector». Em: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, pp. 21–37.
- Lyu, Mingyang, Yibo Zhao, Chao Huang e Hailong Huang (jul. de 2023). *Unmanned Aerial Vehicles for Search and Rescue: A Survey*. DOI: 10.3390/rs15133266.
- Makrigiorgis, Rafael, Panayiotis Kolios e Christos Kyrkou (set. de 2022). *Aerial Vessels Detection Dataset*. Versão 1.0. Zenodo. DOI: 10.5281/zenodo.7076145. URL: <https://doi.org/10.5281/zenodo.7076145>.
- Marine ship 2, Marine (out. de 2023). *ship aerials 2 Dataset*. <https://universe.roboflow.com/marine-ship-2/ship-aerials-2>. Open Source Dataset. visited on 2025-07-30. URL: <https://universe.roboflow.com/marine-ship-2/ship-aerials-2>.
- NATO Shipping Centre (2021). *AIS (Automatic Identification System) Overview*. Acesso em: 31 jan. 2025. URL: <https://shipping.nato.int/nsc/operations/news/2021/ais-automatic-identification-system-overview>.
- OpenCV (2018). *OpenCV: Open Source Computer Vision Library*. URL: <https://github.com/opencv/opencv/blob/4.x/doc/pattern.png>.

- Pfeifer, Norbert, Philipp Glira e Christian Briese (2012). «Direct Georeferencing with on board Navigation Components of Light Weight UAV Platform». Em: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. URL: <http://chdk.wikia.com/wiki/CHDK>.
- Ribeiro, Miguel, Bruno Damas e Alexandre Bernardino (nov. de 2022). «Real-Time Ship Segmentation in Maritime Surveillance Videos Using Automatically Annotated Synthetic Datasets». Em: *Sensors* 22 (21). ISSN: 14248220. DOI: 10.3390/s22218090.
- Ribeiro, Ricardo, Gonçalo Cruz, Jorge Matos e Alexandre Bernardino (set. de 2019). «A Data Set for Airborne Maritime Surveillance Environments». Em: *IEEE Transactions on Circuits and Systems for Video Technology* 29 (9), pp. 2720–2732. ISSN: 15582205. DOI: 10.1109/TCSVT.2017.2775524.
- Rubí, Bartomeu, Jonathan Cacace, Javier Rodriguez, Rafael Company, Mark Tanner, Roberto Arzo e Julián Cayero (2024). «VESSELIing: A Large UAV-Based Vessel Image Dataset for Port Surveillance». Em: *2024 International Conference on Unmanned Aircraft Systems, ICUAS 2024*. Institute of Electrical e Electronics Engineers Inc., pp. 76–83. ISBN: 9798350357882. DOI: 10.1109/ICUAS60882.2024.10556944.
- Scanu, David (mar. de 2024). *Aerial views of boats Dataset*. <https://universe.roboflow.com/david-scanu-mfhtq/aerial-views-of-boats>. Open Source Dataset. visited on 2025-07-30. URL: <https://universe.roboflow.com/david-scanu-mfhtq/aerial-views-of-boats>.
- Sévigny, P, D Kirkland, X Li e B Balaji (2021). *Unmanned Aircraft (UA) Telemetry Data for Track Modelling and Classification*. NATO.
- SIYI Technology (jul. de 2024). *SIYI ZR10 2K 4MP QHD 30X Hybrid 10X Optical Zoom Gimbal Camera*. <https://shop.siyi.biz/products/siyi-zr10-gimbal-camera>. Acedido em 31 de junho de 2025.
- Ultralytics (2025). *YOLO Data Augmentation Guide*. Web page. URL: <https://docs.ultralytics.com/pt/guides/yolo-data-augmentation/>.
- Vermeille, H. (set. de 2004). «Computing geodetic coordinates from geocentric coordinates». Em: *Journal of Geodesy* 78 (1-2), pp. 94–95. ISSN: 09497714. DOI: 10.1007/s00190-004-0375-4.
- Welch, Greg e Gary Bishop (1995). *An Introduction to the Kalman Filter*. URL: <http://www.cs.unc.edu/~gb>.
- Xiong, K., H. Y. Zhang e C. W. Chan (fev. de 2006). «Performance evaluation of UKF-based nonlinear filtering». Em: *Automatica* 42 (2), pp. 261–270. ISSN: 00051098. DOI: 10.1016/j.automatica.2005.10.004.

- Yang, Ning, Wei Feng Tian, Zhi Hua Jin e Chuan Bin Zhang (2005). «Particle filter for sensor fusion in a land vehicle navigation system». Em: *Measurement Science and Technology* 16 (3), pp. 677–681. ISSN: 09570233. DOI: 10.1088/0957-0233/16/3/008.
- Yang, Xi, Xin Zhang, Nannan Wang e Xinbo Gao (2022). «A Robust One-Stage Detector for Multiscale Ship Detection with Complex Background in Massive SAR Images». Em: *IEEE Transactions on Geoscience and Remote Sensing* 60. ISSN: 15580644. DOI: 10.1109/TGRS.2021.3128060.
- Zhang, Zhengyou (2000). *A Flexible New Technique for Camera Calibration*. Rel. téc.
- Zhao, Dewei, Faming Shao, Li Yang, Xiannan Luo, Qiang Liu, Heng Zhang e Zihan Zhang (dez. de 2023). «Object Detection Based on an Improved YOLOv7 Model for Unmanned Aerial-Vehicle Patrol Tasks in Controlled Areas». Em: *Electronics (Switzerland)* 12 (23). ISSN: 20799292. DOI: 10.3390/electronics12234887.
- Zhao, Jiansen, Yanjun Chen, Zhenzhen Zhou, Jingying Zhao, Shengzheng Wang e Xinqiang Chen (2023). «Multiship Speed Measurement Method Based on Machine Vision and Drone Images». Em: *IEEE Transactions on Instrumentation and Measurement* 72. ISSN: 15579662. DOI: 10.1109/TIM.2023.3272406.
- Zhao, Zhong Qiu, Peng Zheng, Shou Tao Xu e Xindong Wu (nov. de 2019). *Object Detection with Deep Learning: A Review*. DOI: 10.1109/TNNLS.2018.2876865.
- Zou, Zhengxia, Keyan Chen, Zhenwei Shi, Yuhong Guo e Jieping Ye (mai. de 2019). «Object Detection in 20 Years: A Survey». Em: *Proceedings off the IEEE* 111, pp. 257–276. DOI: 10.1109/JPROC.2023.3238524. URL: <http://arxiv.org/abs/1905.05055>.

## Apêndice A - Transformação de coordenadas espaciais

Transformação de WGS-84 para ECEF e posteriormente para ENU.

```
1 function [e, n, u] = WGS84_2_ENU(lat_d, lon_d, h_d, lat_ref, lon_ref)
2 h=h_d-h_d;
3
4 a=6378137.0;
5 f= 1/298.257223563;
6
7 RN=a/(stringstylesqrt(1-(fstringstyle*(2-f)stringstyle*sind(lat_d)^2)));
8
9 x= (RN+h)stringstyle*cosd(lat_d)stringstyle*cosd(lon_d);
10
11 y= (RN+h)stringstyle*cosd(lat_d)stringstyle*sind(lon_d);
12
13 z=((1-f)^2stringstyle*RN+h)stringstyle*sind(lat_d);
14
15 %% ENU
16 Rx=[1 0 0;
17     0 cosd(90-lat_ref) -sind(90-lat_ref);
18     0 sind(90-lat_ref) cosd(90-lat_ref)];
19
20 Rz=[cosd(lon_ref+90) -sind(lon_ref+90) 0;
21     sind(lon_ref+90) cosd(lon_ref+90) 0;
22     0 0 1];
23
24 enu = [x y z]stringstyle*Rzstringstyle*Rx;
25
26 e= enu(1,1);
27 n= enu(1,2);
28 u= enu(1,3);
```

Transformação de ENU para ECEF e, por último, para WGS-84.

```
1 function [lat, lon, h] = ENU_2_WGS84Exact(e, n, u, lat_d, lon_d)
2 %%ENU
3 Rx=[1 0 0;
4     0 cosd(lat_d-90) -sind(lat_d-90);
5     0 sind(lat_d-90) cosd(lat_d-90)];
6
7 Rz=[cosd(-lon_d-90) -sind(-lon_d-90) 0;
```

```

8     sind(-lon_d-90) cosd(-lon_d-90) 0;
9     0 0 1];
10
11 xyz = [e n u]*Rx*Rz;
12
13 x = xyz(1,1);
14 y = xyz(1,2);
15 z = xyz(1,3);
16
17 %%
18 a= 6378137.0;
19 f= 1/ 298.257223563;
20 b= a*(1-f);
21 e2= 1-(b^2/a^2);
22 e12= (a^2/b^2)-1;
23 r= sqrt(x^2+y^2);
24 %beta= atan((a*z)/(b*p));
25
26 F=54*b^2*z^2;
27
28 G=r^2+(1-e2)*z^2-e2*(a^2-b^2);
29
30 c=e2^2*F*r^2/G^3;
31
32 s=(1+c+sqrt(c^2+2*c))^(1/3);
33
34 P=F/(3*(s+(1/s)+1)^2*G^2);
35
36 Q=sqrt(1+2*e2^2*P);
37
38 r0=-((P*e2*r)/(1+Q))+sqrt((a^2/2*(1+1/Q))-((P*(1-e2)*z^2)/(Q*(1+Q)))-(P
    *r^2/2));
39
40 V=sqrt((r-e2*r0)^2+(1-e2)*z^2);
41
42 z0=(b^2*z)/(a*V);
43
44 lat=atand((z+e12*z0)/r);
45
46 if y>=0
47     lon=((pi/2)-2*atan(x/(sqrt(x^2+y^2)+y)))*(180/pi);
48 else
49     lon=(-(pi/2)+2*atan(x/(sqrt(x^2+y^2)-y)))*(180/pi);
50 end
51

```

```
52 %h=p*cosd(lat)+z*sind(lat)-(a^2/RN);  
53 h=0;
```



## Apêndice B - Determinação das matrizes F, Q, h e H

```
1 function [F, Q] = calcular_F_e_Q(T, sigma_sqr)
2     F = stringstyleeye(4);
3     F(1,3) = T; F(2,4) = T;
4     Q = [0.25stringstyle*T^4stringstyle*stringstyleeye(2), 0.5stringstyle*T
5         ^3stringstyle*stringstyleeye(2);
6         0.5stringstyle*T^3stringstyle*stringstyleeye(2), T^2stringstyle*
7         stringstyleeye(2)] stringstyle* sigma_sqr;
8 end
9
10 function [h, Hfinal] = calcular_h_e_H(Pw, paramExtr, paramIntr,
11     coefDistCam)
12     Pc = paramExtr stringstyle* [Pw; 1];
13     x = Pc(1); y = Pc(2); z = Pc(3);
14
15     K1 = coefDistCam(1);
16     K2 = coefDistCam(2);
17     f = paramIntr(1,1);
18     u0 = paramIntr(1,3);
19     v0 = paramIntr(2,3);
20     R = (x/z)^2 + (y/z)^2;
21     A = f/z stringstyle* (1 + K1stringstyle*R + K2stringstyle*R^2);
22     B = f/z stringstyle* (K1 + 2stringstyle*K2stringstyle*R^2);
23     u = A stringstyle* x + u0;
24     v = A stringstyle* y + v0;
25     h = [u v]';
26
27     % Calculo do Jacobiano H:
28     dudx = A; dudy = 0; dudz = -A stringstyle* x/z;
29     dvdx = 0; dvdy = A; dvdz = -A stringstyle* y/z;
30
31     dRdx = 2 stringstyle* x / z^2;
32     dRdy = 2 stringstyle* y / z^2;
33     dRdz = -2 stringstyle* R / z;
34
35     dudR = B stringstyle* x;
36     dvdR = B stringstyle* y;
37
38     Hint = [dudx+dudRstringstyle*dRdx, dudy+dudRstringstyle*dRdy, dudz+
39     dudRstringstyle*dRdz; stringstyle...
40     dvdx+dvdRstringstyle*dRdx, dvdy+dvdRstringstyle*dRdy, dvdz+
41     dvdRstringstyle*dRdz];
```

```
37
38     Hext = paramExtr(:, 1:3);
39     H = Hint stringstyle* Hext;
40
41     % Mapeia para o espaço do estado [x y vx vy]
42     Hfinal = stringstylezeros(2,4);
43     Hfinal(:,1:2) = H(:,1:2);
44 end
```

## Apêndice C - Definição de RMSE e MAE

A métrica do erro *Root Mean Squared Error (RMSE)*, em português, corresponde ao Erro Quadrático Médio (EQM) e avalia a raiz quadrada da média dos quadrados das diferenças entre os valores reais (observados) e os valores previstos (estimativas). Esta fórmula é apresentada na Equação C.1:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (C.1)$$

A métrica do erro *Mean Absolute Error (MAE)*, em português, designa-se por Erro Médio Absoluto e determina a média das diferenças absolutas entre os valores reais (observados) e os valores previstos (estimativas). Esta expressão é dada pela Equação C.2:

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (C.2)$$