

# Ship, guide my landing! A Binocular Extended Kalman Filter for Maritime UAV Landing Operations

Bruno Damas<sup>\*†</sup>, Nuno Pessanha Santos<sup>\*†‡</sup>, and Matilde Correia Vieira<sup>\*</sup>

<sup>\*</sup>Portuguese Navy Research Center (CINAV), Portuguese Naval Academy, Almada, 2810-001, Portugal

<sup>†</sup>Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, 1049-001, Portugal

<sup>‡</sup>Portuguese Military Research Center (CINAMIL), Portuguese Military Academy, Lisbon, 1169-203, Portugal

E-mails: bruno.damas@escolanaval.pt santos.naamp@academiamilitar.pt matilde.de.macedo.vieira@marinha.pt

**Abstract**—Unmanned Aerial Vehicles (UAVs) are continuously being explored as important means of decreasing human intervention and improving system reliability. The more challenging tasks when operating with UAVs are take-off and landing. Most low-size UAVs can have their take-off performed by hand, but the capacity to perform autonomous landing is essential. The autonomous landing environment is usually challenging since we are considering a moving platform, and the system must be able to deal with Global Position System (GPS)-denied environments. The proposed system is based on a ground-based stereoscopic vision system with temporal filtering based on an Extended Kalman Filter (EKF) to track the position of a rotary-wing UAV during landing. The obtained position detects and guides the rotary-wing UAV during landing. The initial results indicate that this setup has the potential to track with low error, demonstrating its suitability for exploration and further improvements.

**Index Terms**—Unmanned Aerial Vehicles, Computer Vision, Motion Estimation, Kalman Filters, Landing Maneuver, Maritime Robotics.

## I. INTRODUCTION

Coastal states are responsible for managing their Exclusive Economic Zone (EEZ), and Portugal is no exception. Portugal has the fifth largest EEZ in Europe [1], making it difficult to conduct Intelligence, Surveillance, and Reconnaissance (ISR) operations [2], [3] in such a vast area.

An Unmanned Aerial Vehicle (UAV) is an aircraft that can operate without a human crew on board. It can perform various operations, including ISR, search and rescue, humanitarian aid, distribution of medical supplies, and emergency support [4]–[7]. In modern military operations, UAVs are mainly used for ISR missions [6], [8], [9] but can also carry explosives and ammunition and identify and neutralize targets [2], [3].

Most accidents involving UAVs are caused by human errors related to operating and maintenance failures [10]. Therefore, automating as many tasks as possible is crucial to increase system reliability. The most challenging operations during UAV deployment are the take-off and the landing, and it is essential to automate them to increase their operational reliability completely [11]–[13]. This becomes even more relevant when dealing with a moving platform such as a ship [14], particularly in an environment prone to Global Positioning System (GPS) jamming [15]. Landing a UAV on

a ship can be difficult due to the limited space available, which limits its payload capacity. Additionally, the maritime environment has unique challenges, particularly when working in a wet and prone environment. It is important to explore alternative methods that do not rely on GPS information to ensure that UAV tracking [16], [17] is performed with low error to perform guidance and control successfully.

Combining a camera with Computer Vision (CV) algorithms can be an effective way to track UAVs while also resisting external jamming. However, the use of a camera also has some limitations. For example, the UAV must be within the camera's line of sight, and certain weather conditions can affect this. Stereoscopic-based vision systems can be used for 3D position estimation and tracking [18]. This well-known and straightforward method is mainly based on the epipolar geometry of the camera images [19].

This work aims to implement a stereoscopic vision system to track a rotary-wing UAV and perform guidance and control during the landing process. The system is pre-calibrated using a known chessboard pattern for calibration [20]. The Efficient Perspective-N-Point (EPnP) [21], [22] method is used with a scaled international landing marker ( $H$ ) to obtain the camera's extrinsic parameters (translation & orientation) and ensure greater accuracy in the obtained results. The 3D tracking result is obtained by temporal filtering the estimation using an Extended Kalman Filter (EKF) [23], [24], resulting in a smoother trajectory with fewer errors. This work is still under development, but promising preliminary results have already been obtained, demonstrating the potential of the implementation.

The article is structured as follows: Section II briefly describes current state-of-the-art implementations. Section III formulates the problem and describes the methodology used. Section IV describes some of the obtained preliminary results. Finally, in Section V, conclusions are provided, and areas for further research are suggested.

## II. RELATED WORK

UAVs can be classified based on several characteristics, such as their weight, type, propulsion, or mission profile [25]. When

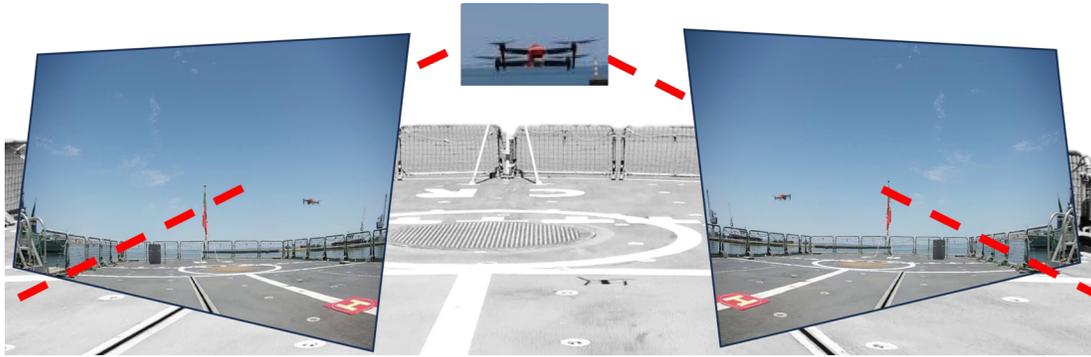


Fig. 1. Stereoscopic camera implementation illustration.

selecting a UAV for a specific task, it is crucial to consider all of these characteristics. If the task is autonomous landing on a moving ship [11], [12], which has limited space available, choosing a UAV with the right characteristics is important.

Some UAV landing systems use a combination of GPS, Inertial Navigation System (INS) data, and CV to perform autonomous landings on ships [26]. However, INS measurements often contain noise [27], and GPS is susceptible to jamming and always has some associated errors [15]. Data fusion [28] brings advantages in most applications. Still, we have to ensure high accuracy in real-time, and the extra effort in processing might not always be significant to the accuracy obtained. There has been a significant interest in using CV in recent years for autonomous landing, with many implementations focused on ground-based or UAV-based applications.

More recently, deep learning techniques have become very popular for real-time, image-based detection and tracking of relevant objects by autonomous unmanned systems [29]. Some ground-based implementations employ the You Only Look Once (YOLO) architecture for UAV detection and use regression [30] or temporal filtering for translation and orientation estimation using a Rao-Blackwellized Particle Filter (PF), an Unscented Bingham-Filter (UBiF) or an Unscented Bingham-Gauss Filter (UBiGaF) [12], [16], [17], [31]. Regarding the implementation of processing on the UAV, small-sized UAVs typically lack significant processing capabilities. Therefore, the method being used must be suitable for real-time implementation. Some UAV-based implementations use cascade models to perform segmentation [32] or use Deep Simple Online and Realtime Tracking (DeepSORT) [33], [34] to detect landing sites.

There are stereoscopic vision systems that use Infra-Red (IR) images [35] or Red, Green, and Blue (RGB) images [36] to estimate the 3D position of UAVs. These systems can combine the estimation with temporal filtering algorithms like the Kalman Filters (KFs) [37] to obtain a smoother trajectory estimation, which reduces the error. From the most popular KF derivation arises the Extend Kalman Filter (EKF) [24], [38] or the Unscented Kalman Filter (UKF) [16]. The standard EKF linearizes non-linear equations using Taylor series expansion [38]. At the same time, UKF avoids this by using a set of

selected points (*sigma points*) to estimate the state distribution [16], [17].

The proposed system is based on a stereoscopic RGB ground-based vision system. This system successfully estimates the trajectory of the UAV using a EKF. The main objective of this implementation was to create a simple and accurate system that can be easily installed at any location upon demand. It is important to consider that the UAV's characteristics may vary depending on the mission, and the system must be flexible enough to accurately estimate the trajectory of various UAV types.

### III. PROBLEM FORMULATION AND METHODOLOGY

For accurate estimation of an approaching UAV's position and velocity, the proposed approach relies on two fixed cameras placed on the ship's deck, pointing towards the stern (Figures 1 and 2). This approach offers several advantages

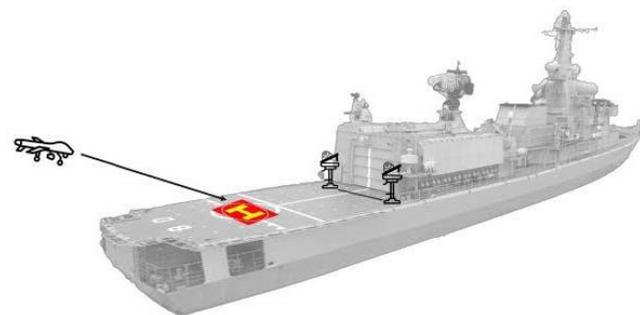


Fig. 2. Used cameras layout on board the ship, including the used international landing marker (*H*).

compared to monocular approaches, even those with high accuracy. The main drawbacks of using a monocular approach are:

- A precise estimation of the UAV size in the image is required;
- Since the UAV attitude and heading can influence its perceived size in the acquired images. A pose estimator is required to estimate the UAV distance to the camera correctly;

- As a consequence, a precise 3D model of the landing UAV may be required, giving away the flexibility needed to deal with unknown UAVs while landing on deck.

The developed system architecture, illustrated in Figure 3, detects the UAV in the image and incorporates that information into a temporal filtering scheme to estimate the UAV's position and linear velocity. Since the obtained measurements are highly noisy, it is essential to implement a temporal filtering step to increase the estimate accuracy. The adopted approach is based on a EKF [23], [24] that can deal with the existing nonlinearities of the adopted observation model.

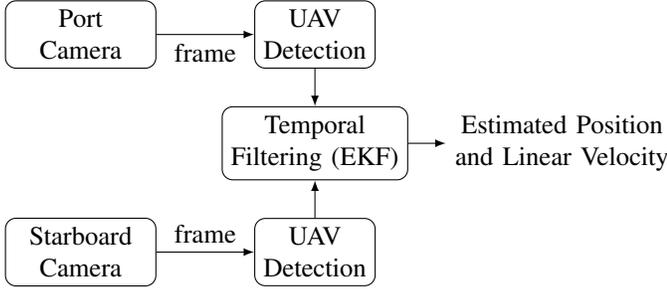


Fig. 3. Simplified implementation schematic.

We first calibrate the fixed cameras on the ship's deck using the international landing marker ( $H$ ) with known geometry and dimensions (Figure 4) to obtain the camera's intrinsic, extrinsic, and radial distortion parameters, according to the standard camera model. This procedure allows us to obtain the position and pose of each camera concerning the coordinate frame defined by the marker. After calibration, the detection bounding boxes corresponding to a UAV, produced by a previously trained state-of-the-art real-time detector such as the YOLO v8 detector, provide a UAV location on each camera image coordinate frame, as illustrated in Figure 1.

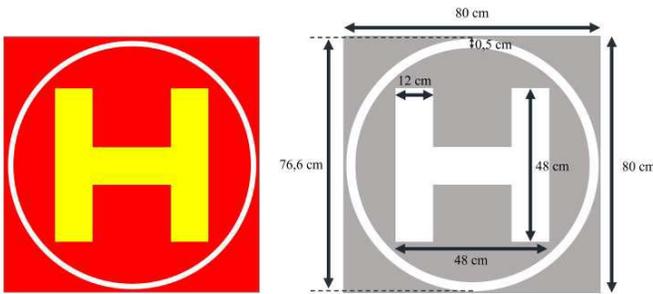


Fig. 4. Used international landing marker ( $H$ ) dimensions.

The EKF [24], [38] proposed in this work addresses the problem of image UAV position and velocity using the image coordinates of the detected UAV on the port and starboard cameras. We consider a constant velocity dynamical model for the UAV that can consider inertial measurements taken from the Inertial Measurement Units (IMUs) installed onboard to tackle the ship's roll and pitch movement caused by sea waves. We also derive the observation model for this filter, which

relates the UAV 3D position in the ship's coordinate frame to the detected bounding boxes in each image, according to the camera's optical models previously obtained in the calibration stage.

#### A. UAV dynamic model

The UAV position, velocity, and acceleration are expressed in a world frame  $\{W\}$  attached to the ship's deck, with the  $x$  axis pointing to the ship's stern, the  $y$  axis pointing starboard, and the  $z$  axis pointing up. In this frame, these quantities are given respectively by vectors  $p = [p_x \ p_y \ p_z]^T$ ,  $v = [v_x \ v_y \ v_z]^T$ , and  $a = [a_x \ a_y \ a_z]^T$ . We consider a standard constant velocity model for the UAV, where, at time  $t_k$ , for  $\Delta T_k = t_k - t_{k-1}$ , we have

$$\begin{aligned} v_k &= v_{k-1} + \Delta T_k a_{k-1}, \\ p_k &= p_{k-1} + \Delta T_k v_{k-1} + \frac{1}{2} \Delta T_k^2 a_{k-1}. \end{aligned}$$

These equations can be written in matrix form as

$$x_k = F_k x_{k-1} + B_k a_{k-1},$$

where the UAV state at time  $k$  is composed of its position and velocity,  $x_k = [p^T \ v^T]^T = [p_x \ p_y \ p_z \ v_x \ v_y \ v_z]^T$ , and the acceleration  $a_k$  is assumed to be unknown, normally distributed with zero mean and diagonal covariance matrix  $\Sigma_a$ ,  $a_k \sim \mathcal{N}(0, \Sigma_a)$ , with  $\Sigma_a = \sigma_a^2 I_{3 \times 3}$ . This results in the following dynamical model for the UAV,

$$x_k = F_k x_{k-1} + w_k, \quad w_k \sim \mathcal{N}(0, Q_k), \quad (1)$$

where  $Q_k = B_k \Sigma_a B_k^T$ . Details on the derivation of  $F_k$  and  $B_k$  can be obtained in [39].

#### B. Camera observation model

The estimation algorithm only has access to the detected UAV position in the images provided by the port (left) and starboard (right) cameras, represented respectively by  ${}^L z = [u_L \ v_L]^T$  and  ${}^R z = [u_R \ v_R]^T$ , where, for each camera,  $u$  and  $v$  are measured in pixels, with  $u$  measured horizontally from the image's top left corner and  $v$  measured downwards from the same corner.

We consider a left and right camera frame attached to each camera's optical center, with  $z$  pointing outwards, denoted respectively by frames  $\{L\}$  and  $\{R\}$ . Given a true UAV position expressed in the ship's coordinate frame  $\{U\}$ ,  $p$ , the corresponding position on an image,  $[u \ v]^T$ , is obtained resorting to the classical image formation model [40]: first  $p$  is transformed to the camera coordinate frame using the camera extrinsic parameters, and then the position  ${}^c p$  in this frame is projected to the image plane using the camera intrinsic model.

1) *Extrinsic parameters:* The transformation from the ship's coordinate frame  $\{W\}$  to a camera coordinate frame  $\{C\}$  is given by the function

$${}^c p = h_{extr}(p) = R_{3 \times 3} p + t_{3 \times 1}, \quad (2)$$

where  $R_{3 \times 3}$  is a  $3 \times 3$  rotation matrix from  $\{W\}$  to  $\{C\}$  and  $t_{3 \times 1}$  is a 3D translation vector corresponding to the

origin of frame  $\{W\}$  expressed in the camera frame  $\{C\}$ . In homogeneous coordinates, this transformation is represented by the extrinsic parameters matrix  ${}^cT_w$ :

$$\begin{bmatrix} {}^c p_x \\ {}^c p_y \\ {}^c p_z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} = {}^c T_w \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix}.$$

2) *Intrinsic parameters*: The distortion free pinhole camera model relates a 3D point expressed in frame  $\{C\}$  to an image location  $[u \ v]^T$  according to the standard relation expressed in homogeneous coordinates:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^c p_x \\ {}^c p_y \\ {}^c p_z \\ 1 \end{bmatrix} = K \begin{bmatrix} {}^c p_x \\ {}^c p_y \\ {}^c p_z \\ 1 \end{bmatrix}.$$

In this intrinsic parameter matrix  $K$ , parameters  $f_x$  and  $f_y$  are the horizontal and vertical focal lengths,  $\gamma$  is the skew coefficient, and  $u_0$  and  $v_0$  represent the optical center in the image. The cameras used in this work have  $f_x = f_y = f$  and  $\gamma = 0$ , and thus the above expression is equivalent to

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = f \begin{bmatrix} \bar{p}_x \\ \bar{p}_y \\ 1 \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \\ 0 \end{bmatrix} \quad \text{with } \bar{p} = \begin{bmatrix} \bar{p}_x \\ \bar{p}_y \\ 1 \end{bmatrix} = \frac{1}{{}^c p_z} \begin{bmatrix} {}^c p_x \\ {}^c p_y \\ 1 \end{bmatrix},$$

as long as  ${}^c p_z \neq 0$ . When  ${}^c p_z = 0$ , the UAV has just crashed into the camera, and the image formation model suddenly turned into an image destruction model. In this situation, estimating the UAV position is no longer useful, and you should probably consider buying a new camera instead. And a new drone.

Also, to obtain accurate results, we must also consider the distortion presented in most real camera lenses [41]: in this work, we only consider radial distortion with coefficients up to the fourth order, and consequently, the relation between a point  ${}^c p$  and the corresponding image projection  $z$  becomes

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = h_{intr}({}^c p) = f(1 + k_1 r^2 + k_2 r^4) \begin{bmatrix} \bar{p}_x \\ \bar{p}_y \\ 1 \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \\ 0 \end{bmatrix}, \quad (3)$$

with

$$r^2 = \bar{p}_x^2 + \bar{p}_y^2 \quad \text{and again} \quad \bar{p} = \begin{bmatrix} \bar{p}_x \\ \bar{p}_y \\ 1 \end{bmatrix} = \frac{1}{{}^c p_z} \begin{bmatrix} {}^c p_x \\ {}^c p_y \\ 1 \end{bmatrix}. \quad (4)$$

The complete image formation (observation) model considered in this paper is given by the composition of the extrinsic and intrinsic camera models, *i.e.*, a UAV position  $p$  expressed in the ship frame  $\{W\}$  generates the corresponding detected positions on the left and right images according to

$$\begin{aligned} L_z &= ({}^L h_{intr} \circ {}^L h_{extr})(p) = {}^L h(p) \\ R_z &= ({}^R h_{intr} \circ {}^R h_{extr})(p) = {}^R h(p), \end{aligned} \quad (5)$$

where functions  $h_{extr}$  and  $h_{intr}$  are given by (2) and (3) for left and right cameras.

The UAV detection on the images, using classical or deep learning methods, inevitably has associated some errors, and it usually does not correspond to the geometric center of the

vehicle associated with position  $p$ : we, therefore, model these detection inaccuracies as additive Gaussian noise,

$$z_k = h(p_k) + v_k, \quad v_k \sim \mathcal{N}(0, R_k), \quad (6)$$

where  $R_k$  is a  $2 \times 2$  diagonal matrix whose elements correspond to the detector variance in the left or right images (in pixels squared), and the function  $h(x_k)$ , accordingly, is equal to  ${}^L h(p_k)$  or  ${}^R h(p_k)$ .

### C. Cameras Calibration: Intrinsic Parameters

Focal length  $f$ , optical center  $[u_0 \ v_0]$ , and radial distortion parameters  $k_1$  and  $k_2$ , the intrinsic parameters, depend only on camera construction, zoom level, and image resolution. To obtain these parameters, we apply the classical Open source Computer Vision (OpenCV) chessboard camera calibration procedure, based on [20], for each *Dahua DH-IPC-HDBW8232EP-Z-SL* camera acquired for this purpose, using a 720p image resolution and a minimum zoom level (Figure 5). After this procedure, the following intrinsic matrices were obtained:

$${}^L K = \begin{bmatrix} 762.7 & 0 & 639.5 \\ 0 & 762.7 & 359.5 \\ 0 & 0 & 1 \end{bmatrix} \quad {}^R K = \begin{bmatrix} 774.4 & 0 & 639.5 \\ 0 & 774.4 & 359.5 \\ 0 & 0 & 1 \end{bmatrix}$$



Fig. 5. *Dahua* camera (left) and chessboard calibration pattern (right).

### D. Cameras Calibration: Extrinsic Parameters

While the intrinsic functions depend only on the camera model used in the image resolution and the chosen field of view, the extrinsic functions are related to the camera's placement on the ship's deck. To obtain the cameras' extrinsic parameters, we placed a landing marker on the ship's deck, visible in both camera images, and we defined the world frame as the marker's center, as depicted in Figure 6.

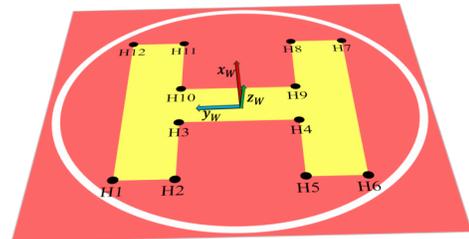


Fig. 6. Illustration of the used international landing marker ( $H$ ).

We also mounted the two cameras on the deck of *NRP Corte-Real*, a Portuguese Navy frigate, pointing to the stern, one placed starboard and the other placed on the port side. We manually identified, for each camera, the  $[uv]$  image locations corresponding to 12 points on this marker whose location in frame  $\{W\}$  is known (Figure 7). After obtaining the correspondence between world frame 3D points and 2D image points, we obtain the following cameras' extrinsic transformation matrices using the OpenCV EPnP algorithm, based on [21]:

$${}^L T_w = \begin{bmatrix} -0.148 & -0.984 & 0.097 & 2.73 \\ 0.182 & 0.069 & -0.980 & 1.52 \\ 0.971 & -0.163 & 0.169 & 3.89 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$${}^R T_w = \begin{bmatrix} 0.292 & -0.956 & 0.024 & -2.52 \\ 0.153 & 0.021 & -0.987 & 1.46 \\ 0.943 & 0.292 & 0.153 & 3.99 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$



Fig. 7. Images used for camera calibration (*top*) and marker detail and annotations (*bottom*).

### E. Extended Kalman Filter

Since the observation model given by (5) is nonlinear, we resort to the standard EKF to recursively estimate the UAV state  $x_k$  using the current observation  $z_k$  and the estimate of the previous state  $x_{k-1}$ . The filter starts with an initial estimate  $\hat{x}_0$  and corresponding uncertainty covariance matrix  $\Sigma_0$ . After that, the state estimate and uncertainty are updated recursively as new observations are acquired using the prediction and update stages of the filter presented next.

1) *Predict Step*: In the prediction stage, the transition model is used to obtain the new estimate and to propagate the uncertainty, given their previous values, using the dynamical model (1):

$$\bar{x}_k = F_k \hat{x}_{k-1}, \quad (7)$$

$$\bar{\Sigma}_k = F_k \Sigma_{k-1} F_k^T + Q_k. \quad (8)$$

2) *Update Step*: In the update step, the observations from the left or right cameras are incorporated using the observation model (6):

$$\hat{x}_k = \bar{x}_k + K_k(z_k - h(\bar{x}_k)), \quad (9)$$

$$\Sigma_k = \bar{\Sigma}_k - K_k H_k \bar{\Sigma}_k, \quad (10)$$

where

$$K_k = \bar{\Sigma}_k H_k^T (H_k \bar{\Sigma}_k H_k^T + R_k)^{-1} \quad (11)$$

is the Kalman gain and  $H_k$  is the Jacobian of the observation  $z$  with respect to the state  $x$ , evaluated at  $\bar{x}_k$ . Since observations do not depend on the UAV velocity  $v$ ,  $H_k$  is given by

$$H_k = \left. \frac{d^L z}{dx} \right|_{x=\bar{x}_k} = [{}^L H_k \quad | \quad 0_{2 \times 3}]$$

for an observation originating on the left camera and

$$H_k = \left. \frac{d^R z}{dx} \right|_{x=\bar{x}_k} = [{}^R H_k \quad | \quad 0_{2 \times 3}]$$

for images from the right camera. Using (5), these Jacobians are equal to

$${}^L H_k = \left. \frac{d^L z}{dp} \right|_{p=\bar{p}_k} = {}^L H_{intr}({}^L \bar{p}) \cdot {}^L H_{extr}(\bar{p}_k) \quad \text{and} \quad (12)$$

$${}^R H_k = \left. \frac{d^R z}{dp} \right|_{p=\bar{p}_k} = {}^R H_{intr}({}^R \bar{p}) \cdot {}^R H_{extr}(\bar{p}_k),$$

where  ${}^L \bar{p} = {}^L h_{extr}(\bar{p}_k)$  and  ${}^R \bar{p} = {}^R h_{extr}(\bar{p}_k)$  are the current estimates for the UAV position on each of the camera frames.

The extrinsic transformation Jacobians are simply given by the corresponding rotation matrices,  ${}^L H_{extr} = {}^L R_{3 \times 3}$  and  ${}^R H_{extr} = {}^R R_{3 \times 3}$ . The Jacobians for the intrinsic transformations, on the other hand, require the derivation of (3–4) with respect to  ${}^c p$ : these are not straightforward calculations and are presented in Appendix A for convenience.

## IV. EXPERIMENTAL RESULTS

We simulate a fixed-wing UAV final approach to the ship to evaluate the proposed approach. This UAV has an approximate width of 1 m and moves with a maximum speed of 20 m/s, reducing this speed in the final stage of the landing approach. Its trajectory is projected to the left and right camera images according to (5), with a sampling rate of  $T_s = 0.1$  s, using actual intrinsic and extrinsic parameters obtained using the procedure detailed in Section III-D, with the cameras placed on the ship's deck (Figure 8).



Fig. 8. Real-data calibration procedure.

### A. EKF vs. Stereo: Low detection noise

In the first experiment, we compare the proposed approach to the classical stereo approach based on binocular geometrical estimation [19]. This latter method requires the simultaneous acquisition of left and right images and successful detection of the UAV in both images to apply the epipolar equations, which is often not a realistic assumption. On the other hand, our approach naturally allows for asynchronous UAV detections from left and right cameras or even sporadic bursts of missing detections in one or both cameras.

We assume the detector has an error with a standard deviation equal to 5% of the detection Bounding Box (BB) width in the images and consider, for now, that both cameras synchronously acquire images with a 100% detection success rate. After a quick trial and error tuning phase, we set a standard deviation 4 m/s for  $\sigma_a$ , the acceleration uncertainty of the dynamical model presented in Section III-A. In Figure 9, we present the estimated trajectories using both methods, together with the ground-truth depiction and the trajectory projection performed by the UAV on the left and right images.

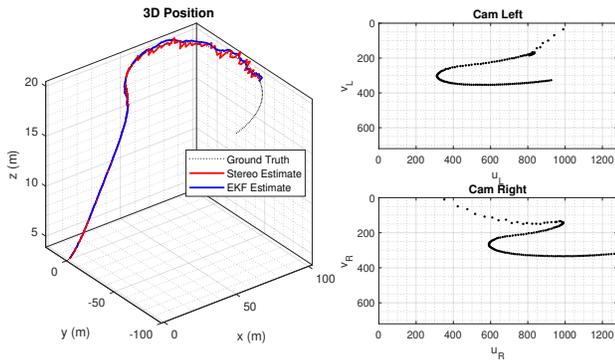


Fig. 9. UAV 3D trajectory (left) and UAV detections along the trajectory in the left camera (right top) and right camera (right bottom).

The results are quite comparable, and the plot of the Root Mean Square Error (RMSE) along the trajectory (Figure 10) confirms this observation.

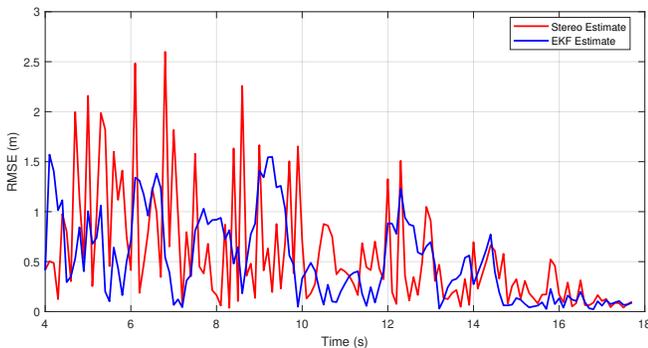


Fig. 10. RMSE along the trajectory.

However, the stereo approach cannot directly estimate the UAV velocity, an important quantity required for autonomous

vehicle control. We can estimate it using the estimated position difference between consecutive samples, which amplifies the noise associated with the position estimation (Figure 11). Our EKF approach, on the other hand, does not seem to suffer from this phenomenon and provides relatively smooth estimated velocity signals that the UAV guidance and control algorithms can employ.

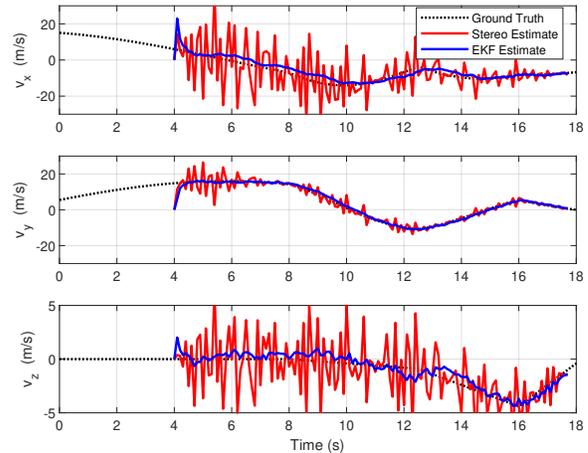


Fig. 11. Obtained velocity estimation.

### B. EKF vs. Stereo: High detection noise

The tracking problem becomes more challenging if we increase the detector standard deviation fourfold to 20% of the detected bounding box width.

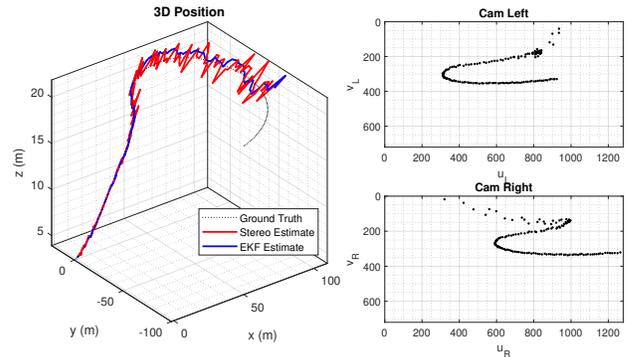


Fig. 12. UAV 3D trajectory with large noise on the detection stage (left) and UAV detections along the trajectory in the left camera (right top) and right camera (right bottom).

As seen in Figure 12, this results, as expected, in a considerable increase in the estimation error, particularly on the depth estimation, which suffers the most from inaccuracies in the detection stage. Nevertheless, the EKF maintains a lower estimation error than the stereo approach (Figure 13). As expected, such an increase in the position estimate noise for the stereo approach results in the corresponding velocity estimation being completely unusable (Figure 14). The EKF

estimate, on the other hand, manages to provide acceptable velocity feedback for the UAV.

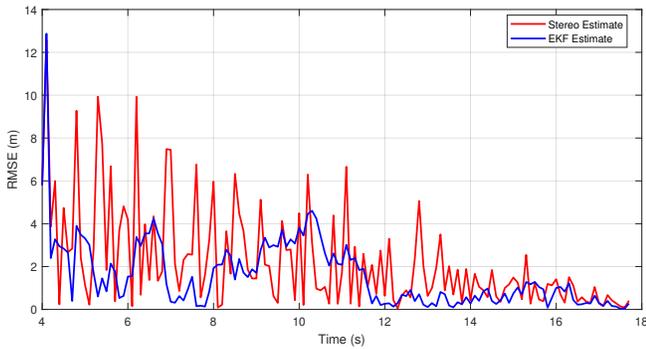


Fig. 13. RMSE along the trajectory with a large noise level.

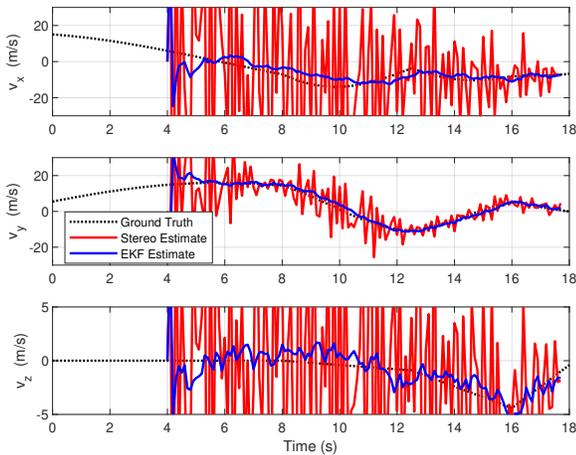


Fig. 14. Velocity estimation, large noise level.

### C. EKF: Low detection failure rate

The classical stereo method for position estimation [19] requires the simultaneous detection of the UAV on both cameras to provide an estimated location. Even if a decent synchronization is achieved between acquired camera images, the UAV localization may become compromised in the presence of medium or high failure rates in the detection stage.

The approach proposed in this paper naturally deals with asynchronous sensor updates, processing each camera detection location as soon as it becomes available. To test the system's capability to deal with missing data from the cameras' detectors, we simulate detection failures in both cameras, with a miss rate of 40% at the beginning of the trajectory, when the UAV is around 100m away from the landing location when its detection bounding box width is only a few pixels wide. This rate drops, as the distance to the cameras decreases, to around 5%, corresponding to the final approach stage, when the UAV is around 10m away from the ship (Figure 15).

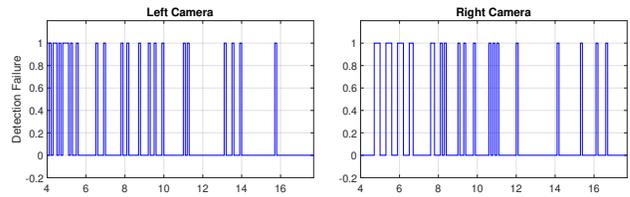


Fig. 15. Camera detection failures along the trajectory: Left camera (*left*) and right camera (*right*).

Nevertheless, in this situation, with an average 18% detection failure in each camera, there are no noticeable differences in the EKF estimated trajectory and velocity profile when compared to the scenario presented in Section IV-A (Figures 16 and 17).

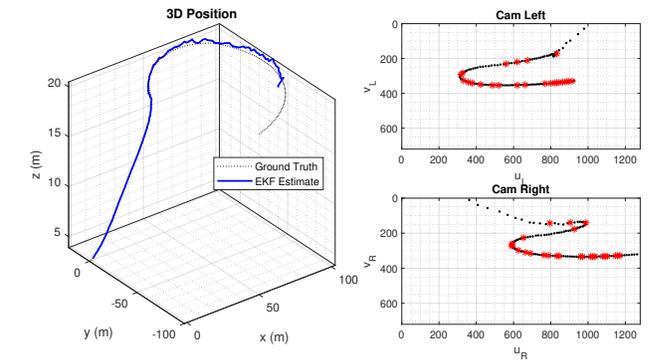


Fig. 16. UAV 3D trajectory (*left*) and UAV detections along the trajectory in left and right cameras (*right*). Missing detections are marked in red.

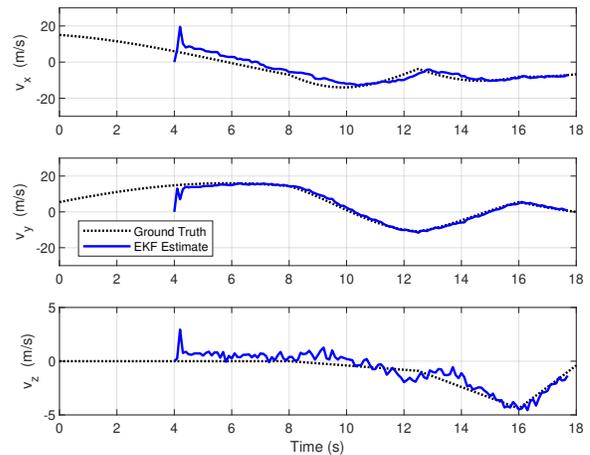


Fig. 17. Velocity estimation with missing data.

### D. EKF: High detection failure rate

In the final experiment, we increased the failure probability for the detector. For the same trajectory, the miss rate now starts at 80%, dropping to 10% when the UAV is a few meters from the ship's deck (Figure 18).

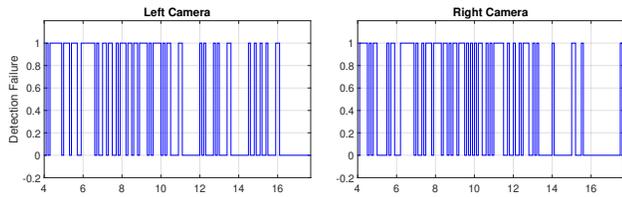


Fig. 18. Camera detection failures along the trajectory with a high miss rate: Left camera (*left*) and right camera (*right*).

During the trajectory, there was no simultaneous detection of the UAV on both cameras on 65 % of the time, and this number was close to 100 % in the first seconds of the trajectory when the UAV was considerable far away. This is clearly unacceptable for the classical stereo estimation approach. However, as can be seen in Figure 19, the estimate provided by the EKF degrades slowly with such an increase in the detector's error rate.

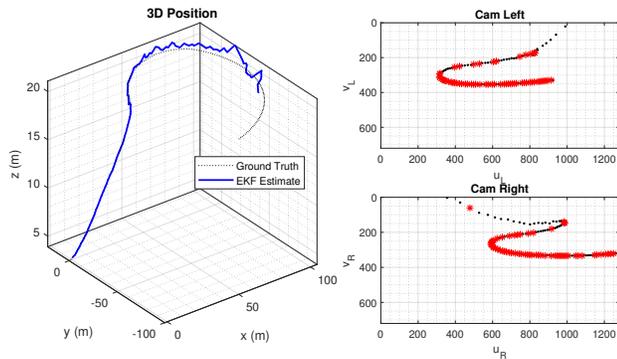


Fig. 19. UAV 3D trajectory with high detection miss rate (*left*) and UAV detections along the trajectory in left and right cameras (*right*). Missing detections are marked in red.

The plot of the RMSE for the low and high detector failure rate in Figure 20 confirms this perception: although the RMSE is slightly larger during the first half of the trajectory, at around 2 m, it quickly drops to less than 0.3 m in the final stage of the UAV approach to the landing spot. This is an acceptable value for a safe landing on the ship.

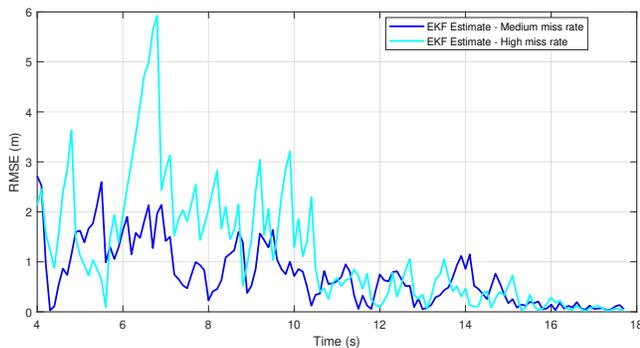


Fig. 20. RMSE along the trajectory.

### E. Real system implementation: What is missing?

Some applications use synthetic data to perform the UAV tracking algorithm development and training [12], [16], [31]. The acquisition and synchronization of real data is a challenging task that limits the existence of real captured publicly available datasets for UAV landing applications with ground truth. Although we do not yet have a complete dataset with ground truth, some efforts have been made towards that objective, developing a proper apparatus to capture real data, as illustrated in Figure 21.



Fig. 21. Dahua camera (*left*), wooden support designed to perform data acquisition (*center*) and GPS (*right*).

Most object detection algorithms in captured frames are based on supervised learning [30]. Gathering data representing the UAV on the frame is essential to properly train the UAV detection algorithm and obtain high detection accuracy in real-world applications. Some efforts toward that goal were already made, as illustrated in Figure 22. An example of estimated UAV BB in a captured frame using Yolo v8 is illustrated in Figure 23.



Fig. 22. Examples of acquired dataset images that can be used for object detection algorithm training.



Fig. 23. Detected BB representing the UAV in a captured frame: Left camera (*left*) and right camera (*right*) - Example.

Since the main objective is to develop a system capable of UAV tracking to guide the landing stage, it is essential to try to capture as much data from real-world scenarios as possible, giving the dataset the needed diversity. We have started capturing some data onboard the Portuguese Navy frigate *NRP Corte-Real*, as illustrated in Figure 24. However, the synchronization process between cameras and sensors is not yet accurate.

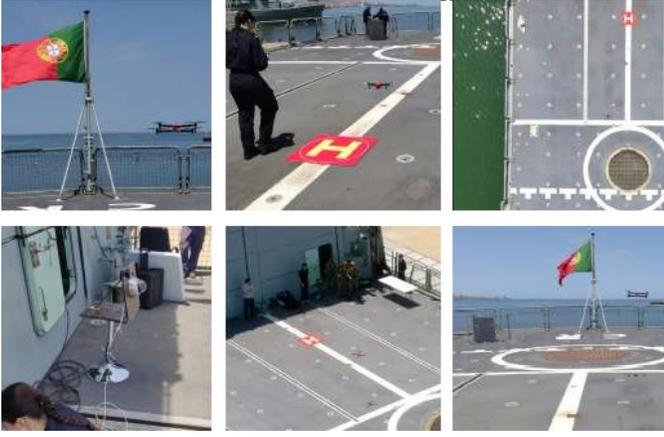


Fig. 24. Real data acquisition tests performed onboard the Portuguese Navy frigate *NRP Corte-Real* - Example.

All the data was acquired to simulate the expected UAV landing trajectory as close to reality as possible, as illustrated in Figure 25. The initial tests were conducted at the dock, but as the system improves and becomes more accurate, gathering data at sea and conducting real sea trials is crucial.



Fig. 25. UAV trajectory during the data acquisition tests performed onboard the Portuguese Navy frigate *NRP Corte-Real* - Example.

Despite demonstrating its suitability for the pretended task, the system is still under development. It needs some improvements to be considered a final product that can be directly applied during real naval operations. As described before, more data will benefit the system's performance by using more accurate data and the lessons learned from the experiential, hands-on approach. After implementing a more robust system, the next step is to incorporate this information into a complete UAV guidance and control system.

## V. CONCLUSIONS

The use of UAVs is increasing, and automating most procedures for successful missions is important. One of the most crucial phases of a mission is landing, and automating it reduces the chances of failure due to human factors. The proposed architecture is designed to estimate the 3D position of the UAV, which is useful for guidance and control. This implementation is still under development, and further testing is needed. However, the results obtained so far have demonstrated the potential of the implementation for real-world applications. Future work will focus on collecting a real dataset to test the developed algorithms and to close the control

loop by incorporating this information into a complete UAV guidance and control implementation.

## APPENDIX A

### DERIVATION OF INTRINSIC TRANSFORMATION JACOBIAN

The Jacobian in (12) for the left camera,

$${}^L H_{intr} = \left. \frac{d^L h_{intr}(p)}{dp} \right|_{p=L\bar{p}} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{bmatrix},$$

can be calculated resorting to (3–4), for  $L\bar{p} = [p_x \ p_y \ p_z]^T$ , according to:

$$\begin{aligned} h_{11} &= \frac{d^L u}{dp_x} = \frac{\partial^L u}{\partial p_x} + \frac{\partial^L u}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_x} \\ h_{12} &= \frac{d^L u}{dp_y} = \frac{\partial^L u}{\partial p_y} + \frac{\partial^L u}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_y} \\ h_{13} &= \frac{d^L u}{dp_z} = \frac{\partial^L u}{\partial p_z} + \frac{\partial^L u}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_z} \\ h_{21} &= \frac{d^L v}{dp_x} = \frac{\partial^L v}{\partial p_x} + \frac{\partial^L v}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_x} \\ h_{22} &= \frac{d^L v}{dp_y} = \frac{\partial^L v}{\partial p_y} + \frac{\partial^L v}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_y} \\ h_{23} &= \frac{d^L v}{dp_z} = \frac{\partial^L v}{\partial p_z} + \frac{\partial^L v}{\partial r^2} \cdot \frac{\partial r^2}{\partial p_z}, \end{aligned}$$

where

$$\begin{aligned} \frac{\partial^L u}{\partial p_x} &= f(1 + k_1 r^2 + k_2 r^4) \frac{1}{p_z}, & \frac{\partial^L u}{\partial p_y} &= 0, \\ \frac{\partial^L v}{\partial p_x} &= \frac{\partial^L u}{\partial p_y} = 0, & \frac{\partial^L v}{\partial p_y} &= \frac{\partial^L u}{\partial p_x}, \\ \frac{\partial^L u}{\partial p_z} &= -\frac{p_x}{p_z} \cdot \frac{\partial^L u}{\partial p_x}, & \frac{\partial^L v}{\partial p_z} &= -\frac{p_y}{p_z} \cdot \frac{\partial^L v}{\partial p_y}, \\ \frac{\partial^L u}{\partial r^2} &= f(k_1 + 2k_2 r^2) \frac{p_x}{p_z}, & \frac{\partial^L v}{\partial r^2} &= f(k_1 + 2k_2 r^2) \frac{p_y}{p_z}, \\ \frac{\partial r^2}{\partial p_x} &= 2 \frac{p_x}{p_z^2}, & \frac{\partial r^2}{\partial p_y} &= 2 \frac{p_y}{p_z^2}, & \frac{\partial r^2}{\partial p_z} &= -2 \frac{r^2}{p_z}. \end{aligned}$$

The same calculations, of course, apply to the right camera using the corresponding intrinsic parameters.

## REFERENCES

- [1] B. Nogueira, A. Torres, N. Moniz, and G. M. Menezes, "Dynamics of fisheries in the azores islands: A network analysis approach," *arXiv preprint arXiv:2309.09378*, 2023.
- [2] A. Konert and T. Balcerzak, "Military autonomous drones (uavs)-from fantasy to reality. legal and ethical implications." *Transportation research procedia*, vol. 59, pp. 292–299, 2021.
- [3] K. Chávez and O. Swed, "Emulating underdogs: Tactical drones in the russia-ukraine war," *Contemporary Security Policy*, vol. 44, no. 4, pp. 592–605, 2023.
- [4] F. Nex, C. Armenakis, M. Cramer, D. A. Cucci, M. Gerke, E. Honkavaara, A. Kukko, C. Persello, and J. Skaloud, "Uav in the advent of the twenties: Where we stand and what is next," *ISPRS journal of photogrammetry and remote sensing*, vol. 184, pp. 215–242, 2022.
- [5] E. Ackerman and E. Strickland, "Medical delivery drones take flight in east africa," *IEEE Spectrum*, vol. 55, no. 1, pp. 34–35, 2018.
- [6] L. Gonçalves and B. Damas, "Automatic detection of rescue targets in maritime search and rescue missions using uavs," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2022.

- [7] M. M. Marques, P. Dias, N. Pessanha Santos, V. Lobo, R. Batista, D. Salgueiro, A. Aguiar, M. Costa, J. E. Da Silva, A. S. Ferreira *et al.*, "Unmanned aircraft systems in maritime operations: Challenges addressed in the scope of the seagull project," in *OCEANS 2015-Genova*. IEEE, 2015, pp. 1–6.
- [8] M. Ribeiro, B. Damas, and A. Bernardino, "Real-time ship segmentation in maritime surveillance videos using automatically annotated synthetic datasets," *Sensors*, vol. 22, no. 21, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/21/8090>
- [9] N. Pessanha Santos, V. B. Rodrigues, A. B. Pinto, and B. Damas, "Automatic detection of civilian and military personnel in reconnaissance missions using a uav," in *2023 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2023, pp. 157–162.
- [10] K. W. Williams Ph D, "A summary of unmanned aircraft accident/incident data: Human factors implications," in *2005 International Symposium on Aviation Psychology*, 2005, p. 824.
- [11] N. Pessanha Santos, V. Lobo, and A. Bernardino, "Autoland project: Fixed-wing uav landing on a fast patrol boat using computer vision," in *OCEANS 2019 MTS/IEEE SEATTLE*. IEEE, 2019, pp. 1–5.
- [12] —, "Fixed-wing unmanned aerial vehicle 3d-model-based tracking for autonomous landing," *Drones*, vol. 7, no. 4, p. 243, 2023.
- [13] N. Pessanha Santos, "Fixed-wing uav pose estimation using a self-organizing map and deep learning," *Robotics*, vol. 13, no. 8, 2024.
- [14] G. Cademartori, L. Oneto, F. Valdenazzi, A. Coraddu, A. Gambino, and D. Anguita, "A review on ship motions and quiescent periods prediction models," *Ocean Engineering*, vol. 280, p. 114822, 2023.
- [15] A. Grant, P. Williams, N. Ward, and S. Basker, "Gps jamming and the impact on maritime navigation," *The Journal of Navigation*, vol. 62, no. 2, pp. 173–187, 2009.
- [16] N. Pessanha Santos, V. Lobo, and A. Bernardino, "Unscented particle filters with refinement steps for uav pose tracking," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 2, p. 52, 2021.
- [17] —, "Directional statistics for 3d model-based UAV tracking," *IEEE Access*, vol. 8, 2020.
- [18] A. Islam, M. Asikuzzaman, M. O. Khyam, M. Noor-A-Rahim, and M. R. Pickering, "Stereo vision-based 3d positioning and tracking," *IEEE Access*, vol. 8, pp. 138 771–138 787, 2020.
- [19] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International journal of computer vision*, vol. 27, 1998.
- [20] —, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [21] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnnp: An accurate o(n) solution to the pnp problem," *International journal of computer vision*, vol. 81, pp. 155–166, 2009.
- [22] F. Morais, T. Ramalho, P. Sinogas, M. M. Marques, N. Pessanha Santos, and V. Lobo, "Trajectory and guidance mode for autonomously landing an uav on a naval platform using a vision approach," in *OCEANS 2015-Genova*. IEEE, 2015, pp. 1–7.
- [23] X. I. Wong and M. Majji, "Extended kalman filter for stereo vision-based localization and mapping applications," *Journal of Dynamic Systems, Measurement, and Control*, vol. 140, no. 3, p. 030908, 2018.
- [24] R. E. Kalman *et al.*, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [25] M. Hassanalian and A. Abdelkefi, "Classifications, applications, and design challenges of drones: A review," *Progress in Aerospace Sciences*, vol. 91, pp. 99–131, 2017.
- [26] G. Xu, Y. Zhang, S. Ji, Y. Cheng, and Y. Tian, "Research on computer vision-based for uav autonomous landing on a ship," *Pattern Recognition Letters*, vol. 30, no. 6, pp. 600–605, 2009.
- [27] H. Naseri and M. Homaeinezhad, "Improving measurement quality of a mems-based gyro-free inertial navigation system," *Sensors and Actuators A: Physical*, vol. 207, pp. 10–19, 2014.
- [28] B. Du, R. Mao, N. Kong, and D. Sun, "Distributed data fusion for on-scene signal sensing with a multi-uav system," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 3, pp. 1330–1341, 2020.
- [29] D. Nunes, J. Fortuna, B. Damas, and R. Ventura, "Real-time vision based obstacle detection in maritime environments," in *2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2022, pp. 243–248.
- [30] A. Barisic, M. Car, and S. Bogdan, "Vision-based system for a real-time detection and following of uav," in *2019 Workshop on Research, Education and Development of Unmanned Aerial Systems*, 2019.
- [31] T. Ferreira, A. Bernardino, and B. Damas, "6d uav pose estimation for ship landing guidance," in *OCEANS 2021: San Diego – Porto*, 2021.
- [32] C. Pires, B. Damas, and A. Bernardino, "An efficient cascaded model for ship segmentation in aerial images," *IEEE Access*, vol. 10, 2022.
- [33] T. D. Trong, Q. T. Hai, M. V. Van, B. N. Thai, T. N. Chi, and T. N. Quang, "Autonomous detection and approach tracking of moving ship on the sea by vtol uav based on deep learning technique through simulated real-time on-air image acquisitions," in *2021 8th NAFOSTED Conference on Information and Computer Science (NICS)*, 2021.
- [34] Y. Li, H. Yuan, Y. Wang, and B. Zhang, "Maritime vessel detection and tracking under uav vision," in *2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT)*. IEEE, 2022, pp. 1–4.
- [35] W. Kong, D. Zhang, X. Wang, Z. Xian, and J. Zhang, "Autonomous landing of an uav with a ground-based actuated infrared stereo vision system," in *2013 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2013, pp. 2963–2970.
- [36] W. Kong, D. Zhou, Y. Zhang, D. Zhang, X. Wang, B. Zhao, C. Yan, L. Shen, and J. Zhang, "A ground-based optical system for autonomous landing of a fixed wing uav," in *2014 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2014, pp. 4797–4804.
- [37] P. S. Maybeck, "The kalman filter: An introduction to concepts," in *Autonomous robot vehicles*. Springer, 1990, pp. 194–204.
- [38] G. Mao, S. Drake, and B. D. Anderson, "Design of an extended kalman filter for uav localization," in *2007 Information, Decision and Control*. IEEE, 2007, pp. 224–229.
- [39] B. Damas, N. Pessanha Santos, and M. Correia Vieira, "A visual tracking system for uav landing on ships," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2024.
- [40] H. Martins, J. R. Birk, and R. B. Kelley, "Camera models based on data from two calibration planes," *Computer Graphics and Image Processing*, vol. 17, no. 2, pp. 173–180, 1981.
- [41] R. Juarez-Salazar, J. Zheng, and V. H. Diaz-Ramirez, "Distorted pinhole camera modeling and calibration," *Applied Optics*, vol. 59, no. 36, 2020.