Journal Title
XX(X):1–20
©The Author(s) 2023
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

A Bearing-Angle Approach for Unknown Target Motion Analysis Based on Visual Measurements

Zian Ning[1,2] and Yin Zhang[2] and Jianan Li[2] and Zhang Chen[3], and Shiyu Zhao[4,2]

Abstract

Vision-based estimation of the motion of a moving target is usually formulated as a *bearing-only* estimation problem where the visual measurement is modeled as a bearing vector. Although the bearing-only approach has been studied for decades, a *fundamental limitation* of this approach is that it requires extra lateral motion of the observer to enhance the target's observability. Unfortunately, the extra lateral motion conflicts with the desired motion of the observer in many tasks. It is well-known that, once a target has been detected in an image, a bounding box that surrounds the target can be obtained. Surprisingly, this common visual measurement especially its size information has not been well explored up to now. In this paper, we propose a new *bearing-angle* approach to estimate the motion of a target by modeling its image bounding box as bearing-angle measurements. Both theoretical analysis and experimental results show that this approach can significantly enhance the observability *without* relying on additional lateral motion of the observer. The benefit of the bearing-angle approach comes with no additional cost because a bounding box is a standard output of object detection algorithms. The approach simply exploits the information that has not been fully exploited in the past. No additional sensing devices or special detection algorithms are required.

Keywords

Bearing-only target motion estimation, Pseudo-linear Kalman filter, Observability enhancement

1 Introduction

This paper studies the problem of estimating the motion of a moving target object using a moving monocular camera. The target's geometric information such as its physical size is *unknown* in advance. This problem is important in many fields (Qiu et al. 2019; Griffin and Corso 2021; Tekin et al. 2018). Our present work is particularly motivated by the task of aerial target pursuit, where a micro aerial vehicle (MAV) uses its onboard camera to detect, localize, and then pursue another flying MAV. The task of aerial target pursuit, originally motivated by the interesting bird-catching-bird behaviors in nature (Brighton and Taylor 2019), potentially provides an effective approach to the defense of misused MAV (Rothe et al. 2019; Dressel and Kochenderfer 2019; Vrba and Saska 2020).

When a target has been detected in an image by a vision detection algorithm, we usually obtain a *bounding box* that surrounds the target's image (see Fig. 1). The bounding box carries two types of useful information that can be used to estimate the target's motion.

The first type of useful information is the *center point* of the bounding box. The pixel coordinate of the center point can be used to calculate the spatial *bearing vector* pointing from the camera to the target based on the pin-hole camera model (Ma et al. 2012). Using the bearing vector to estimate the target's motion is referred to as *bearing-only* target motion estimation (Fogel and Gavish 1988; He et al. 2019; Li et al. 2022). As a problem that has been studied for more than 40 years, bearing-only target motion estimation was originally studied to estimate the motion of ships on the ocean surface (Hoelzer et al. 1978), and regained increasing research attention in recent years in vision-based target estimation tasks (Ponda et al. 2009; Anjaly and Ratnoo 2018; He et al. 2019).

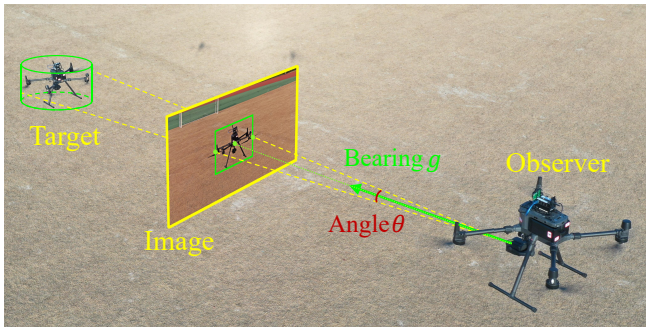


Figure 1. An observer MAV observes a target MAV with a monocular camera. The bearing g and angle θ can be obtained from the bounding box that surrounds the target in the image.

[1] Department of Computer Science & Technology at Zhejiang University, Hangzhou, China
[2] School of Engineering at Westlake University, Hangzhou, China
[3] Department of Automation at Tsinghua University, Beijing, China
[4] Research Center for Industries of the Future, Westlake University, Hangzhou, China

Corresponding author:
Shiyu Zhao, School of Engineering, Westlake University, No.600 Dunyu Road, HangZhou 310024, China.
Email: zhaoshiyu@westlake.edu.cn

Bearing-only target motion estimation requires an *observability condition*: The observer must have higher-order motion than the target and, more importantly, the higher-order motion must contain components that are orthogonal to the target's bearing vector (Fogel and Gavish 1988). Motivated by this observability condition, enormous works have studied how an observer should move to enhance the observability (Hammel et al. 1989; Sabet et al. 2016; Anjaly and Ratnoo 2018; He et al. 2019). For instance, in our recent work (Li et al. 2022), we proposed a helical guidance law so that a MAV moves along a helical curve to optimize the observability in the 3D space.

A *limitation* of the observability condition of the classic bearing-only approach is that the observer must move in the lateral directions that are orthogonal to the bearing vector of the target. Such additional lateral motion is usually unfavorable because it may conflict with the desired motion of the observer in many tasks. For example, in an aerial target pursuit task, the pursuer is desired to approach the target as fast as possible and then keep stationary relative to the target. Then, the additional lateral motion would conflict with the desired motion. It is, therefore, important to study other ways that can enhance the observability while avoiding unfavorable lateral motion.

The second type of useful information of a bounding box is its *size* (either width or height). The size of a bounding box is jointly determined by several factors such as the target's distance, the target's physical size, and the orientation of the camera. The target's physical size is usually unknown in many tasks, especially in those antagonistic ones such as aerial pursuit of misused MAVs. As a result, the size of the bounding box cannot directly infer the target's distance. Nevertheless, it carries valuable information for localizing the target.

Surprisingly, the size information of the bounding box has not been well explored so far. The work that is closely relevant to ours is the state-of-the-art one in (Griffin and Corso 2021), where the size of a bounding box is used to localize unknown target objects. Although the approach in (Griffin and Corso 2021) is inspiring, it relies on two assumptions: The target objects are stationary and the camera can only translate without rotating. It is still an open problem how to estimate a target's motion when the two assumptions are not valid. Moreover, the theoretical role of the size of a bounding box in target motion estimation has not been fully understood so far. Although the work in (Vrba and Saska 2020) also utilizes the size of the bounding box to estimate the target's position, it is assumed that the target's physical size is known in advance.

Estimating the motion of moving objects is also a fundamental problem in dynamic SLAM. For example, the works in (Yang and Scherer 2019; Qiu et al. 2019) firstly estimate the camera's pose and secondly estimate the target object's pose subject to a scale factor, and finally estimate the scale factor from multi-view measurements. To estimate the target object's pose subject to a scale factor, (Yang and Scherer 2019) and (Qiu et al. 2019) rely on detecting, respectively, a 3D bounding box and sufficient feature points inside the 2D bounding box. Different from (Yang and Scherer 2019; Qiu et al. 2019), our proposed approach merely utilizes a 2D image bounding box without further extracting feature points or a 3D bounding box inside the 2D bounding box. As a result, one benefit is that this approach is more computationally efficient. Moreover, this approach can handle the challenging small-target case where the target object is far and hence its image is small. In this case, it would be unreliable to extract sufficient stable features or conduct 3D detection.

The aforementioned approaches in (Griffin and Corso 2021; Yang and Scherer 2019; Qiu et al. 2019) are all based on multiple views. It is also possible to estimate the target's depth from a single view/image (Tekin et al. 2018; Vrba and Saska 2020). The single-view approach however requires prior information of the objects. Moreover, it would be unable to successfully localize target objects with different sizes but similar appearances. In this paper, we focus on the multi-view case.

In this paper, we propose a novel *bearing-angle* target motion estimation approach that models a bounding box as bearing-angle measurements. This approach can enhance the observability by fully exploiting the information in a bounding box rather than relying on the additional lateral motion of the observer. The benefit of the proposed bearing-angle approach comes with no additional cost since the bounding box is a standard output of object detection algorithms. The approach simply exploits the angle information that has not been fully exploited in the past. No additional sensing devices or special detection algorithms are required.

The technical novelties of this approach are threefold.

1) The proposed approach does not directly use the size of a bounding box because the size is variant to the orientation of the camera. That is, even if the target's relative position is unchanged, the size of the bounding box still varies when the camera rotates. Motivated by this problem, we convert the size of the bounding box to an angle subtended by the target (see Fig. 1). The merit of using the angle measurement is that it is *invariant* to the camera's orientation change (see Fig. 2) and hence can greatly facilitate the estimator design. In this way, the assumption in (Griffin and Corso 2021) that the camera can only translate but not rotate can be avoided.

2) Although the bearing-angle approach incorporates an additional angle measurement, it is nontrivial to see how to properly use this measurement because the angle does not directly infer the target's distance given that the target's size is unknown. We notice that the angle is a joint nonlinear function of the target's physical size and relative distance. Hence, the state vector, which only consists of the target's position and velocity in the conventional bearing-only approach, is augmented by the unknown target's physical size. Since the bearing and angle measurements are all nonlinear functions of the target's state, we establish a pseudo-linear Kalman filter to properly utilize the measurements to enhance estimation stability. Both simulation and real-world experiments verify the effectiveness of the proposed estimator.

3) Although an additional angle measurement is used, an additional unknown, the target's physical size, is also introduced into the estimator. It is, therefore, nontrivial to see how the additional angle measurement can help improve the observability. Motivated by this problem, we prove the

necessary and sufficient observability condition for bearing-angle target motion estimation. In particular, we show that the target's motion can be recovered if and only if the observer has a higher-order motion than the target. Different from the bearing-only case, the higher-order motion is *not* required to be in the lateral directions that are orthogonal to the bearing vector. This is an important enhancement of the observability. As we show in various experiments, the bearing-angle approach can successfully recover the target's motion in many scenarios where the bearing-only approach fails.

2 Related Work

2.1 Algorithms for bearing-only target motion estimation

Bearing-only target motion analysis aims to estimate the target's motion states, such as position and velocity, using bearing measurement only. It was originally motivated by ship localization and tracking in the ocean (Hoelzer et al. 1978). With the rapid development of small-scale mobile robots equipped with cameras, the bearing-only approach regained increasing attention in recent years (Ponda et al. 2009; Anjaly and Ratnoo 2018; He et al. 2019).

Kalman filter-based estimators are widely used in the bearing-only target motion. One challenge of applying the Kalman filter to the bearing-only estimation is the nonlinearity of the bearing measurement. The conventional extended Kalman filter (EKF) exhibits divergence problems when applied to bearing-only target motion estimation (Aidala 1979; Lin et al. 2002). Several methods have been proposed to solve this problem. They can be divided into two types. The first type is the modified polar EKF, which was first proposed in (Hoelzer et al. 1978). In this approach, three observable quantities are separated from the unobservable ones to prevent divergence. The work in (Stallard 1991) extends this approach to the case of spherical coordinates to track targets in 3D space. The second type is the pseudo-linear Kalman filter, which is first proposed in (Lingren and Gong 1978) to solve the instability problem by transforming the nonlinear measurement equation into a pseudo-linear one. However, this transformation makes the noise become non-Gaussian and highly correlated to the measurement matrix and then causes estimation bias. Nevertheless, the work in (Aidala and Nardone 1982) theoretically proves that the velocity estimation has no bias, and the position estimation bias can be removed by the observer's maneuvers.

Recently, other estimation algorithms based on advanced but more complex filters have been proposed. The work in (Farina 1999) uses the maximum likelihood (MLE) algorithm to estimate the target's motion using bearing-only measurements. The comparison with the Cramer-Rao lower bound indicates that the MLE-based estimator is effective against measurement errors. The work in (Doğançay 2005) proposes a constrained total least-squares algorithm, which can improve the estimation accuracy when the error of bearing measurement is large. Three different algorithms are used and compared in (Lin et al. 2002). The results show that the EKF, the pseudo-linear filter, and the particle filter have similar performances in most situations, while the EKF loses track when the initial estimate error is large.

Another type of approach, called bearing-only trajectory triangulation (Avidan and Shashua 2000), estimates the target's position from the perspective of trajectory fitting. It reconstructs the trajectory by intersecting parametric trajectory to a series of sight rays obtained from bearing measurement. Once the trajectory is successfully fitted, the target's position at each time instant can be estimated by the intersection of the bearing and the trajectory. The trajectory fitting relies on the assumption of the trajectory's shape. However, in many applications, the target's trajectory is complex and unknown in advance. Many consecutive studies aim to relax this assumption in various ways based on hypersurfaces (Kaminski and Teicher 2004), parametric temporal polynomials (Yu et al. 2009), or compact basis vectors (Park et al. 2015).

2.2 Observability analysis of bearing-only target motion estimation

Observability is a fundamental problem in bearing-only target motion estimation. Early works mainly focus on whether the system is observable or not. For example, the work in (Lingren and Gong 1978) uses the rank of observation matrix to determine the observability. The work in (Fogel and Gavish 1988) extends the observability criterion in (Nardone and Aidala 1981) to the Nth-order target dynamics and inspires us for the observability analysis in Section 6. All these conditions indicate that the observer must have extra high-order motion in the lateral direction. The observability condition can be significantly relaxed in our approach.

Unlike the works on determining whether the system is observable or not, some studies focus on quantifying the observability degree. The work in (Hammel et al. 1989) first introduces the Fisher information matrix (FIM) into the observability analysis. The works in (Sabet et al. 2016) and (Anjaly and Ratnoo 2018) use FIM-based objective functions to maximize observability. We also use the FIM in our former work (Li et al. 2022) to optimize the 3D helical guidance law for better observability. Another method called the geometric method uses the geometric relationship between the target and the observer in two consecutive time instants to derive the measure of observability (He et al. 2019; Woffinden and Geller 2009), and the results are consistent with those derived using FIM. Compared to the bearing-only approach, the observability degree of our bearing-angle method is sufficient to estimate the target's motion in many common scenarios such as tracking and guidance (see experiment results in Figs. 4c and 13).

3 Problem Formulation

Consider a target object moving in the 3D space. Its position and velocity at time t_k are denoted as $p_T(t_k) \in \mathbb{R}^3$ and $v_T(t_k) \in \mathbb{R}^3$, respectively. Suppose there is an observer carrying a monocular camera to observe the target. The position of the observer is denoted as $p_o(t_k) \in \mathbb{R}^3$. Here, we assume that the observer/camera's pose including its position and orientation can be obtained in other ways. For example, it can be measured directly by RTK GPS (Li et al. 2022) or

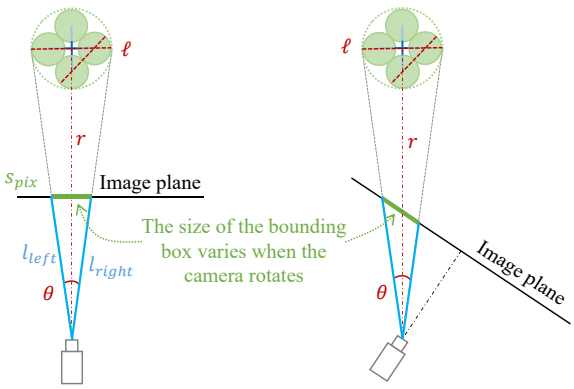


Figure 2. The size of the bounding box varies when the camera rotates. By contrast, the angle subtended by the target object is invariant to the camera's orientation change.

estimated by visual inertial odometry (Qiu et al. 2019). In the rest of the paper, the dependence of a variable on t_k is dropped when the context is clear.

If the target object can be detected by a vision algorithm, we can obtain a bounding box surrounding the target object in the image. Two types of information carried by the bounding box can be used to estimate the motion of the target.

First, the center point of the bounding box can be used to calculate the *bearing* vector of the target. In particular, denote $g \in \mathbb{R}^3$ as the unit bearing vector pointing from p_o to p_T. Suppose $P_{\text{cam}} \in \mathbb{R}^{3\times 3}$ is the intrinsic parameter matrix of the camera (Ma et al. 2012, Section 4), and $R_{\text{c}}^{\text{w}} \in \mathbb{R}^{3\times 3}$ is the rotation from the camera frame to the world frame. Then, the bearing vector g can be calculated as

$$g = \frac{R_{\text{c}}^{\text{w}} P_{\text{cam}}^{-1} q_{\text{pix}}}{\| R_{\text{c}}^{\text{w}} P_{\text{cam}}^{-1} q_{\text{pix}} \|},$$

where $q_{\text{pix}} = [x_{\text{pix}}, y_{\text{pix}}, 1]^{\text{T}} \in \mathbb{R}^3$. Here, $(x_{\text{pix}}, y_{\text{pix}})$ is the pixel coordinate of the center point of the bounding box.

Second, the size of the bounding box can be used to calculate the *angle* subtended by the target in the camera's field of view. The reason that we convert the bounding box's size to the angle is that the angle is invariant to the camera's orientation change (see Fig. 2). In particular, let s_{pix} denote the size of the bounding box. It can be either the width or the height. Let $\theta \in (0, \pi/2)$ be the angle. According to the pinhole camera model (Ma et al. 2012, Section 4) and the law of cosine (see Fig. 2), the angle can be calculated as

$$\theta = \arccos\left(\frac{l_{\text{left}}^2 + l_{\text{right}}^2 - s_{\text{pix}}^2}{2 l_{\text{left}} l_{\text{right}}} \right),$$

where $l_{\text{left}} = \sqrt{(f/\alpha)^2 + (\delta x - s_{\text{pix}}/2)^2 + \delta y^2} \in \mathbb{R}$ and $l_{\text{right}} = \sqrt{(f/\alpha)^2 + (\delta x + s_{\text{pix}}/2)^2 + \delta y^2} \in \mathbb{R}$ are the distances in pixel from the camera center to the middle points of the left and right sides of the bounding box, respectively (Fig. 1). Moreover, f and α denote the camera's focal length and single pixel size, respectively. i_{width} and i_{height} represent the width and the height of the whole image in pixels, respectively. $\delta x = \| x_{\text{pix}} - i_{\text{width}}/2 \| \in \mathbb{R}$ and $\delta y = \| y_{\text{pix}} - i_{\text{height}}/2 \| \in \mathbb{R}$ are the distances between the center of the bounding box and the center of the image.

Our goal is to estimate the target's position and velocity, p_T and v_T, based on the noisy measurements of the bearing vector g and the angle θ together with the observer's own position p_o. To achieve this goal, we propose a new bearing-angle target motion estimator (Fig. 3). The estimator is introduced in detail in Section 4. The observability of this estimator is analyzed based on Kalman's observability criterion in Section 5. We further prove a necessary and sufficient observability condition of the observer in Section 6. Numerical simulation results are given in Section 7. More realistic AirSim simulation results are given in Section 8. Finally, real-world experiments are given in Section 9.

4 Bearing-Angle Target Motion Estimator

This section designs a bearing-angle target motion estimator based on the framework of pseudo-linear Kalman filtering. The key here is to establish appropriate measurement and state transition equations.

4.1 States transition equation

The state vector of the target is designed as

$$x = \begin{bmatrix} p_T \\ v_T \\ \ell \end{bmatrix} \in \mathbb{R}^7,$$

where p_T and v_T are target's global position and velocity, respectively. Here, $\ell > 0$ is a scalar that represents the physical size of the target object in the dimension that is orthogonal to the bearing vector (Fig. 2). In this paper, ℓ is assumed to be constant or varying slowly, which means that the physical size of the target object should be approximately invariant from different viewing angles. Here, ℓ corresponds to θ, which further corresponds to either the width or height of the bounding box. Whether ℓ should correspond to the width or height depends on in which dimension the physical size of the target object is invariant when viewed from different angles. More explanation is given in Section 4.2.

Different from the bearing-only case where the state merely consists of the position and velocity, the state here is augmented by the target's physical size. This is due to the fact that the angle measurement is a function of the target's physical size, which should be estimated as well. One may wonder whether the state vector can also incorporate the target's acceleration. To estimate high-order motion (e.g., acceleration) of the target, the observer must have higher-order motion (e.g., nonzero jerk) according to the observability condition presented in Section 6. Otherwise, the estimation would diverge. Therefore, it is preferred to exclude the acceleration and merely estimate the position and velocity.

If no information of the target's motion is available, it is common to model the target's motion as a discrete-time noise-driven double integrator:

$$x(t_{k+1}) = F x(t_k) + q(t_k), \tag{1}$$

where

$$F = \begin{bmatrix} I_{3\times 3} & \delta t I_{3\times 3} & 0_{3\times 1} \\ 0_{3\times 3} & I_{3\times 3} & 0_{3\times 1} \\ 0_{1\times 3} & 0_{1\times 3} & 1 \end{bmatrix} \in \mathbb{R}^{7\times 7}, \tag{2}$$

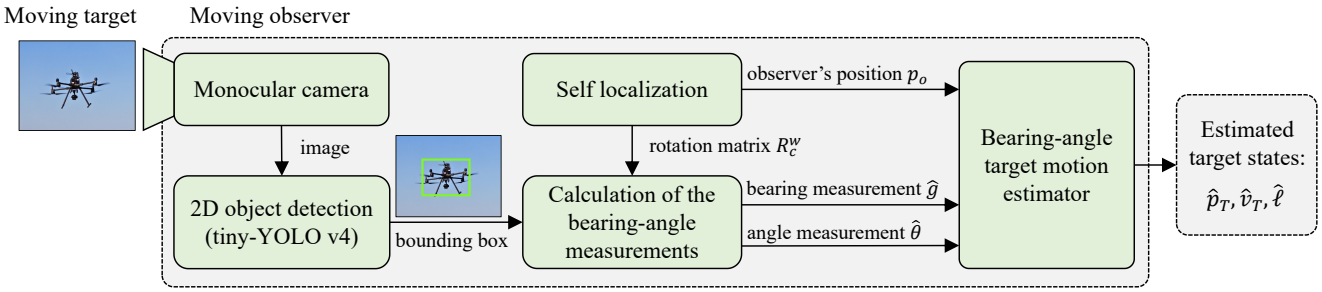


Figure 3. The architecture of the proposed approach. All the simulation and real-world experiments in this paper follow this architecture.

with δt as the sampling time, and I and 0 as the identity and zero matrices, respectively. Here, $q \in \mathbb{R}^7$ is a zero-mean process noise satisfying $q \sim \mathcal{N}(0, \Sigma_q)$, where the covariance matrix is

$$\Sigma_q = \mathrm{diag}(0, 0, 0, \sigma_v^2, \sigma_v^2, \sigma_v^2, \sigma_\ell^2) \in \mathbb{R}^{7\times 7}. \tag{3}$$

Here, $\sigma_v \in \mathbb{R}$ and $\sigma_\ell \in \mathbb{R}$ are the standard deviations of the target's velocity and size, respectively. When the target's shape is irregular, ℓ may vary when viewed from different angles. By letting $\sigma_\ell \neq 0$, we can handle the case where ℓ varies slowly. The dynamic modeling of ℓ is discussed in the following subsection.

4.2 Dynamic modeling of target's physical size

Since the target's physical size ℓ is a state variable to be estimated, it is important to discuss its dynamic model. In fact, the dynamic model of ℓ in (1) assumes that ℓ varies slowly. We next justify this modeling and provide more discussion.

First of all, ℓ corresponds to the physical size of the target object in the dimension that is orthogonal to the bearing vector. Its dynamics can be categorized into three cases.

1) ℓ is invariant. In theory, when ℓ is invariant, a change of θ implies a change of r. As a result, the measurement of θ can help improve the system's observability, as proven in Section 6. An ideal case where ℓ is invariant is that the target object is a sphere or cylinder so that ℓ corresponds to its diameter (Vrba and Saska 2020). In practice, the target object does not have to be the ideal case. For example, consider an autonomous driving scenario where a focal vehicle uses a camera to localize its surrounding vehicles in the 2D plane. Although the physical size of a surrounding vehicle changes greatly when viewed from behind or side, the height of the vehicle is *invariant* from different side-view angles. In this case, ℓ corresponds to the height of the vehicle, and we need to use the height of the image bounding box to calculate θ.

2) ℓ varies slowly. If there does not exist any dimension in which the physical size of the target remains invariant, ℓ may vary slowly when the target is viewed from different angles. For example, in the tasks of aerial target pursuit, if the target is a quadcopter or hexacopter, then ℓ is approximately equal to the wheelbase but may vary slightly when viewed from different angles since the MAV is not a perfect cylinder. In this case, ℓ corresponds to the wheelbase of the MAV, and we need to use the width of the image bounding box to calculate θ.

If ℓ varies slowly, it can still be treated as invariant within short time intervals. As long as the observability condition (Section 6) is satisfied, the motion of the target as well as ℓ can be successfully estimated. This fact is supported by the experimental results in Section 8.4. It is however worth nothing that the performance of the proposed bearing-angle approach would degenerate to the conventional bearing-only one because the additional information brought by θ is used to estimate the time-varying ℓ rather than helping improve the system's observability.

3) ℓ varies rapidly. If ℓ varies rapidly due to certain reasons, it would be difficult to distinguish whether the change of θ is caused by the change of ℓ or the change of r. For example, when a MAV is used to track a ground vehicle, ℓ in any dimension may vary rapidly when the relative motion between the MAV and the ground vehicle is highly dynamic. In such scenarios, the additional information brought by θ is no longer sufficient to estimate the rapidly varying ℓ in this case. Additional visual information such as a 3D bounding box that indicates the target's 3D attitude is required. This is an important topic for future research but out of the scope of the present paper.

4.3 Nonlinear measurement equations

The bearing vector g and the subtended angle θ are both nonlinear functions of the target's position. In particular,

$$g = \frac{p_T - p_o}{r}, \tag{4a}$$

$$\theta = 2\arctan\left(\frac{\ell}{2r}\right) \approx \frac{\ell}{r}, \tag{4b}$$

where

$$r = \|p_T - p_o\|$$

is the distance between the target and the observer. It is notable that there is an approximation in (4b). This approximation is accurate. Specifically, when $r > 3\ell$, which is common in practice, it can be verified that the approximation error is less than 0.08%. The approximation error further decreases as r increases.

In practice, measurements always contain noises. First, denote $\hat{g} \in \mathbb{R}^3$ as the noise-corrupted bearing measurement. Then, we have

$$\hat{g} = R(\eta, \epsilon)\, g, \tag{5}$$

where $R(\eta, \epsilon) \in \mathbb{R}^{3\times 3}$ is a rotation matrix that perturbs g. Here, $\eta \in \mathbb{R}^3$ is a unit vector representing a random rotation axis, and $\epsilon \in \mathbb{R}$ is a random rotation angle. This rotation matrix would rotate the vector g by an angle ϵ around the

axis η. The productive noise in (5) can be transformed into an additive one:

$$\hat{g} = g + \mu, \tag{6}$$

where $\mu = (R(\eta, \epsilon) - I_{3\times3})g \in \mathbb{R}^3$ is the measurement noise of the bearing vector. The covariance of μ is derived in our previous work (Li et al. 2022). Since the covariance is complex and involves unknown true values, we can approximately treat it as a Gaussian noise: $\mu \sim \mathcal{N}(0, \sigma_\mu^2 I_{3\times3})$ (Li et al. 2022).

Substituting (4a) into (6) gives the *nonlinear bearing measurement equation:*

$$\hat{g} = \frac{p_T - p_o}{r} + \mu. \tag{7}$$

Second, denote $\hat{\theta} \in \mathbb{R}$ as the noise-corrupted measurement of the subtended angle. Then, we have

$$\hat{\theta} = \theta + w, \tag{8}$$

where $w \sim \mathcal{N}(0, \sigma_w^2)$ is the measurement noise. Substituting (4b) into (8) yields the *nonlinear angle measurement equation:*

$$\hat{\theta} = \frac{\ell}{r} + w. \tag{9}$$

4.4 Pseudo-linear measurement equations

The measurement equations (7) and (9) are nonlinear in the target's state. In the following, we convert the two equations to be pseudo-linear and then apply pseudo-linear Kalman filtering to achieve better estimation stability (Lin et al. 2002).

First, to convert the 3D bearing measurement to pseudo-linear, we introduce a useful orthogonal projection matrix:

$$P_{\hat{g}} \doteq I_{3\times3} - \hat{g}\hat{g}^{\mathrm{T}} \in \mathbb{R}^{3\times3}.$$

This matrix plays an important role in the analysis of bearing-related estimation and control problems (Zhao and Zelazo 2019). It has an important property:

$$P_{\hat{g}}\hat{g} = 0_{3\times1}.$$

As a result, multiplying $rP_{\hat{g}}$ on both side of (7) yields

$$0_{3\times1} = P_{\hat{g}}(p_T - p_o) + rP_{\hat{g}}\mu$$

and consequently

$$P_{\hat{g}}p_o = P_{\hat{g}}p_T + rP_{\hat{g}}\mu.$$

Rewriting this equation in terms of the target's state variables yields the *pseudo-linear bearing measurement equation:*

$$P_{\hat{g}}p_o = \begin{bmatrix} P_{\hat{g}} & 0_{3\times4} \end{bmatrix} \begin{bmatrix} p_T \\ v_T \\ \ell \end{bmatrix} + rP_{\hat{g}}\mu. \tag{10}$$

Here, $P_{\hat{g}}p_o$ on the left-hand side is the new measurement, which is pseudo-linear in the target's state variables. The reason that it is called "pseudo" is because the measurements also appear on the right-hand side of the equation, especially in the measurement matrix.

Second, we convert the nonlinear angle measurement in (9) to be pseudo-linear. To that end, multiplying $r\hat{g}$ on both side of (9) yields

$$\hat{\theta}r\hat{g} = \ell\hat{g} + wr\hat{g}. \tag{11}$$

It follows from (7) that $r\hat{g} = p_T - p_o + r\mu$, substituting which into the left-hand side of (11) gives

$$\hat{\theta}(p_T - p_o + r\mu) = \ell\hat{g} + wr\hat{g}.$$

Reorganizing the above equation gives

$$\hat{\theta}p_o = \hat{\theta}p_T - \ell\hat{g} + r(\hat{\theta}\mu - w\hat{g}).$$

Rewriting this equation in terms of the target's state variables yields the *pseudo-linear angle measurement equation:*

$$\hat{\theta}p_o = \begin{bmatrix} \hat{\theta}I_{3\times3} & 0_{3\times3} & -\hat{g} \end{bmatrix} \begin{bmatrix} p_T \\ v_T \\ \ell \end{bmatrix} + r(\hat{\theta}\mu - w\hat{g}), \tag{12}$$

where $\hat{\theta}p_o$ is the new measurement that is pseudo-linear in the target's state variables.

4.5 Bearing-angle estimation algorithm

Combining (10) and (12) gives the compact form of the measurement equation:

$$z = Hx + \nu, \tag{13}$$

where

$$z = \begin{bmatrix} P_{\hat{g}}p_o \\ \hat{\theta}p_o \end{bmatrix} \in \mathbb{R}^6, \tag{14a}$$

$$H = \begin{bmatrix} P_{\hat{g}} & 0_{3\times3} & 0_{3\times1} \\ \hat{\theta}I_{3\times3} & 0_{3\times3} & -\hat{g} \end{bmatrix} \in \mathbb{R}^{6\times7}, \tag{14b}$$

$$\nu = \begin{bmatrix} rP_{\hat{g}}\mu \\ r(\hat{\theta}\mu - w\hat{g}) \end{bmatrix} \in \mathbb{R}^6. \tag{14c}$$

Here, ν can be rewritten as a matrix form

$$\nu = E\begin{bmatrix} \mu \\ w \end{bmatrix},$$

where

$$E = r\begin{bmatrix} P_{\hat{g}} & 0_{3\times1} \\ \hat{\theta}I_{3\times3} & -\hat{g} \end{bmatrix} \in \mathbb{R}^{6\times4}. \tag{15}$$

As a result, ν can be approximately treated as a linear transformation of Gaussian noises. Its covariance matrix can be calculated as

$$\Sigma_\nu = E\begin{bmatrix} \sigma_\mu^2 I_{3\times3} & 0_{3\times1} \\ 0_{1\times3} & \sigma_w^2 \end{bmatrix} E^{\mathrm{T}} \in \mathbb{R}^{6\times6}.$$

Although the quantities in E such as $\hat{g}$ and $\hat{\theta}$ contain measurement noises, it is a common practice to treat them as deterministic quantities. Otherwise, if, for example, $\hat{g}$ is split to $\hat{g} = g + \mu$ and we consider the noise separately,

the expression of ν would be a complex function of the true values and the noises. Since the true values are unknown, the covariance cannot be calculated. Moreover, r in (15) is the true target range, which is unknown. We can use the estimated value $\hat{r} = \|\hat{p}_T - p_o\|$ to replace it in implementation. Here, $\hat{p}_T \in \mathbb{R}^3$ is the estimated value of the target's position. This technique has been used in bearing-only target estimation (He et al. 2018; Li et al. 2022).

With the state transition equation (1) and the measurement equation (13), the bearing-angle estimator can be realized by the Kalman filter. For a quick reference, we list the steps below. The prediction steps are

$$\hat{x}^-(t_k) = F\hat{x}(t_{k-1}),$$
$$P^-(t_k) = FP(t_{k-1})F^{\mathrm{T}} + \Sigma_q,$$

where $\hat{x}^-(t_k) \in \mathbb{R}^7$ and $P^-(t_k) \in \mathbb{R}^{7\times 7}$ are the prior estimated state and covariance matrix, respectively. The correction steps are

$$K(t_k) = P^-(t_k)H^{\mathrm{T}}(t_k)\left[H(t_k)P^-(t_k)H^{\mathrm{T}}(t_k) + \Sigma_\nu\right]^\dagger,$$
$$\hat{x}(t_k) = \hat{x}^-(t_k) + K(t_k)\left[z(t_k) - H(t_k)\hat{x}^-(t_k)\right],$$
$$P(t_k) = \left[I_{7\times 7} - K(t_k)H(t_k)\right]P^-(t_k),$$

where $K(t_k) \in \mathbb{R}^{7\times 6}$ is the Kalman gain matrix, $\hat{x}(t_k)$ and $P(t_k)$ are posterior estimated state and covariance matrix, and symbol $\dagger$ denotes the pseudoinverse. The usage of pseudoinverse in the Kalman filter is a common practice to prevent the situation that $H(t_k)P^-(t_k)H^{\mathrm{T}}(t_k) + \Sigma_\nu$ is rank deficient (Yoshikawa 1972; Kulikov and Kulikova 2018).

5 Observability Analysis by Kalman's Criterion

Although an additional angle measurement is adopted in the bearing-angle estimator, it is nontrivial to see whether this additional measurement can improve the system's observability because an additional unknown variable, the target's physical size, is also required to estimate. It is therefore necessary to study the observability conditions under which the target's motion can be successfully estimated.

In this and the next sections, we present two methods to analyze the observability conditions. The first method, as presented in this section, relies on Kalman's observability criterion, which is to check the rank of the observability matrix of a linear system. The second method, as presented in the next section, relies on solving a set of linear equations. Both methods have been adopted in the literature to analyze the observability of estimators (Zhao et al. 2015; Fogel and Gavish 1988). For the bearing-angle estimator, the first method considers the specific dynamics of the filter but is not able to handle the case when the target's motion has a higher order. The second method can handle the high-order motion of the target but does not consider the dynamics of the filter. We will show that the conclusions given by the two methods are consistent. In both of the methods, we consider the case where ℓ is invariant.

5.1 The observability matrix

Consider a time horizon of $k \geq 3$ consecutive steps. The observability matrix of the system of (14b) and (2) can be calculated as

$$Q = \begin{bmatrix} H(t_1) \\ H(t_2)F \\ H(t_3)F^2 \\ \cdots \\ H(t_k)F^{k-1} \end{bmatrix} \in \mathbb{R}^{6k\times 7}. \tag{16}$$

Substituting the expressions of F and H in (2) and (14b) into (16) yields

$$Q = \left[\begin{array}{ccc} P_g(t_1) & 0_{3\times 3} & 0_{3\times 1} \\ \theta(t_1)I_{3\times 3} & 0_{3\times 3} & -g(t_1) \\ \hdashline P_g(t_2) & \delta t P_g(t_2) & 0_{3\times 1} \\ \theta(t_2)I_{3\times 3} & \delta t\theta(t_2)I_{3\times 3} & -g(t_2) \\ \hdashline \vdots & \vdots & \vdots \\ \hdashline P_g(t_k) & (k-1)\delta t P_g(t_k) & 0_{3\times 1} \\ \theta(t_k)I_{3\times 3} & (k-1)\delta t\theta(t_k)I_{3\times 3} & -g(t_k) \end{array}\right].$$

Note that the noises in the bearing and angle measurements are neglected when we analyze the fundamental observability property. After a series of elementary row transformations in Q, we can obtain

$$Q \to \left[\begin{array}{ccc} I_{3\times 3} & 0_{3\times 3} & -g(t_1)/\theta(t_1) \\ 0_{3\times 3} & I_{3\times 3} & -\delta v(t_2)/\ell \\ \vdots & \vdots & \vdots \\ 0_{3\times 3} & I_{3\times 3} & -\delta v(t_k)/\ell \\ \hdashline 0_{3k\times 3} & 0_{3k\times 3} & 0_{3k\times 1} \end{array}\right], \tag{17}$$

where

$$\delta v(t_k) \doteq v_T(t_k) - v_o(t_k)$$

is the relative velocity.

In the following two subsections, we analyze the rank of the observability matrix in two scenarios where the observer moves with zero and nonzero acceleration, respectively. In the two scenarios, the target is always assumed to move with a constant velocity:

$$v_T(t_k) = v_T^{\text{const}}.$$

5.2 Case 1: the observer's velocity is constant

Denoted $v_o \in \mathbb{R}^3$ as the velocity of the observer. Consider the case where the observer has a constant velocity $v_o^{\text{case1}}(t_i) = v_o^{\text{const}}$ for any $i \in \{1, \ldots, k\}$. Then, the relative velocity is also constant:

$$\delta v^{\text{case1}}(t_i) = v_T^{\text{const}} - v_o^{\text{const}} = \delta v^{\text{const}}. \tag{18}$$

Substituting (18) into (17) and conducting elementary row transformation yields

$$Q^{\text{case1}} \to \left[\begin{array}{cc:c} I_{3\times 3} & 0_{3\times 3} & -g(t_1)/\theta(t_1) \\ 0_{3\times 3} & I_{3\times 3} & -\delta v^{\text{const}}/\ell \\ \hdashline 0_{6(k-1)\times 3} & 0_{6(k-1)\times 3} & 0_{6(k-1)\times 1} \end{array}\right]. \tag{19}$$

Since the upper 6×7 block of (19) has full row rank and the lower block is zero, the rank of Q^{case1} is

$$\text{rank}\left(Q^{\text{case1}}\right) = 6.$$

Since the number of states is seven and the rank is six, we know there is *one unobservable mode*. To identify this unobservable mode, we calculate the unobservable subspace, which is the null space of Q:

$$\text{Null}\left(Q^{\text{case1}}\right) = \text{span}\left\{\begin{bmatrix} g(t_1)/\theta(t_1) \\ \delta v^{\text{const}}/\ell \\ 1 \end{bmatrix}\right\}. \tag{20}$$

According to (20), the unobservable mode is

$$x^T \begin{bmatrix} g(t_1)/\theta(t_1) \\ \delta v^{\text{const}}/\ell \\ 1 \end{bmatrix} = p_T^{\text{T}} \frac{g(t_1)}{\theta(t_1)} + v_T^{\text{T}} \frac{\delta v^{\text{const}}}{\ell} + \ell. \tag{21}$$

Although there is only one unobservable mode, this mode given in (21) involves all the states including the target's position, velocity, and physical size. It suggests that the estimation of the three quantities is coupled. In conclusion, we know that, if the target moves with a constant velocity, its states are unobservable when the observer moves with a constant velocity.

5.3 Case 2: the observer's velocity is time-varying

We now consider the case where the observer has nonzero acceleration so that its velocity is time-varying across the time horizon from t_1 to t_k.

Denote $a_o(t_i) \in \mathbb{R}$ as the observer's acceleration, which can be approximated as

$$\begin{aligned} a_o(t_i) &\approx \frac{v_o(t_i) - v_o(t_{i-1})}{\delta t} \\ &= -\frac{[v_T^{\text{const}} - v_o(t_i)] - [v_T^{\text{const}} - v_o(t_{i-1})]}{\delta t} \\ &= -\frac{\delta v(t_i) - \delta v(t_{i-1})}{\delta t}. \end{aligned} \tag{22}$$

Substituting (22) into (17) and performing elementary row transformation yields

$$Q^{\text{case2}} \to \left[\begin{array}{ccc} I_{3\times3} & 0_{3\times3} & -g(t_1)/\theta(t_1) \\ 0_{3\times3} & I_{3\times3} & -\delta v(t_2)/\ell \\ 0_{3\times3} & 0_{3\times3} & \delta t a_o(t_3)/\ell \\ \hdashline \vdots & \vdots & \vdots \\ 0_{3\times3} & 0_{3\times3} & \delta t a_o(t_k)/\ell \\ 0_{3k\times3} & 0_{3k\times3} & 0_{3k\times1} \end{array}\right]. \tag{23}$$

The upper 6×7 block in (23) has full column rank. Therefore, if $a_o(t_i) \neq 0$ for any $i \geq 3$, then

$$\text{rank}\left(Q^{\text{case2}}\right) = 7,$$

Which is the same as the number of estimated states. Therefore, the target's state is observable when the observer moves with nonzero acceleration.

5.4 Summary of this section

From the above analysis, we know that when the target has a constant velocity, its states including its position, velocity, and physical size are observable if and only if the observer has non-zero accelerations.

The critical difference of this condition from the bearing-only case is that the target's states are still observable *even if the observer moves along the bearing vector* towards or backward the target. By contrast, for a bearing-only estimator, moving along the bearing vector is insufficient to recover the target's motion. Therefore, the additional lateral motion of the observer required in the bearing-only case is *not* required in the bearing-angle case anymore, which provides better flexibility for designing the observer's motion.

6 Observability Analysis by Solving Linear Equations

This section extends the observability condition obtained in the last section to more general cases where the target's velocity does not have to be constant.

6.1 Problem formulation

The observability problem that we aim to solve is to determine whether $p_T(t)$ can be recovered from $p_o(t)$ and $g(t), \theta(t)$.

Suppose the target's motion can be described by an nth-order polynomial during a time interval:

$$p_T(t) = b_0 + b_1 t + \cdots + b_n t^n, \tag{24}$$

where $b_0, b_1, \cdots, b_n \in \mathbb{R}^3$ are unknown constant vectors. If we can determine the values of $\{b_i\}_{i=0}^n$, then we can determine the target's motion and hence it is observable. Although polynomials cannot represent all trajectories, they can effectively approximate a majority of them according to the method of Taylor expansion. This is especially true if we consider a short time horizon. This kind of technique has been adopted in the observability analysis of bearing-only target motion estimation tasks (Nardone and Aidala 1981; Lee et al. 2010).

Suppose the observer's motion is described by

$$p_o(t) = c_0 + c_1 t + \cdots + c_n t^n + h(t),$$

where $c_0, c_1, \cdots, c_n \in \mathbb{R}^3$ are constant parameters, and

$$h(t) = d_1 t^{n+1} + d_2 t^{n+2} + \cdots \tag{25}$$

represents *higher-order* motion with $d_1, d_2, \cdots \in \mathbb{R}^3$. It can be verified that the derivatives of $h(t)$ satisfy $h^{(i)}(0) = 0_{3\times1}$ for $i = 0, 1, \cdots, n$. Let $s(t) \in \mathbb{R}^3$ be the relative motion between the target and the observer:

$$\begin{aligned} s(t) &\doteq p_T(t) - p_o(t) \\ &\doteq s_0 + s_1 t + \cdots + s_n t^n + h(t), \end{aligned} \tag{26}$$

where $s_i = d_i - c_i \in \mathbb{R}^3$ for $i = 0, 1, \cdots, n$.

If we can determine $\{s_i\}_{i=0}^n$, then $s(t)$ and hence $p_T(t)$ can be determined. Therefore, we next study under what conditions $\{s_i\}_{i=0}^n$ can be uniquely determined. Since $p_T(t) - p_o(t) = g(t)r(t)$ according to (4a) and $r(t) = \ell/\theta(t)$ according to (4b), we have

$$s(t) = p_T(t) - p_o(t) = g(t)r(t) = \frac{g(t)}{\theta(t)}\ell.$$

Substituting the above equation into (26) yields

$$s_0 + s_1 t + \cdots + s_n t^n + h(t) = \frac{g(t)}{\theta(t)}\ell. \tag{27}$$

Here, $s_0, \cdots, s_n, \ell$ are unknowns to be determined and $g(t), \theta(t), h(t)$ are known. Equation (27) can be reorganized to a linear equation:

$$A(t)X = h(t), \tag{28}$$

where

$$X = \left[s_0^{\mathrm{T}}, s_1^{\mathrm{T}}, \cdots, s_n^{\mathrm{T}}, \ell\right]^{\mathrm{T}} \in \mathbb{R}^{3n+4},$$

and

$$A(t) = \left[I_{3\times3}, tI_{3\times3}, \cdots, t^n I_{3\times3}, \rho(t)\right] \in \mathbb{R}^{3\times(3n+4)}, \tag{29}$$

where

$$\rho(t) \doteq -\frac{g(t)}{\theta(t)} \in \mathbb{R}^3. \tag{30}$$

Therefore, the problem that we aim to solve becomes determining whether X can be uniquely solved from (28).

6.2 Necessary and sufficient observability condition

We next present a necessary and sufficient condition under which the solution X of (28) is unique.

Theorem 1. (Necessary and sufficient observability condition). *The target's motion $p_T(t)$ can be uniquely determined by the observer's motion $p_o(t)$, the bearing $g(t)$, and the angle $\theta(t)$ if and only if*

$$h(t) \neq 0_{3\times1},$$

which means that the order of the observer's motion must be greater than the target.

Proof 6.1. *Since the row number of $A(t)$ is less than its column number, (28) is an under-determined system whose solution cannot be uniquely determined. However, in the continuous time domain, we can use additional higher derivatives of this equation to uniquely determine X.*

In particular, taking the ith-order derivative on both sides of (28) gives $A^{(i)}(t)X = h^{(i)}(t)$. Consider any integer N satisfying $N \geq n+1$. Combining the equations with $i \in \{0, 1, \ldots, N\}$ gives

$$\bar{A}(t)X = \bar{h}(t), \tag{31}$$

where

$$\bar{A}(t) = \begin{bmatrix} A(t) \\ A^{'}(t) \\ \vdots \\ A^{(N)}(t) \end{bmatrix}, \qquad \bar{h}(t) \begin{bmatrix} h(t) \\ h^{'}(t) \\ \vdots \\ h^{(N)}(t) \end{bmatrix}. \tag{32}$$

Here, $\bar{A}(t) \in \mathbb{R}^{(3N+3)\times(3n+4)}$ and $\bar{h}(t) \in \mathbb{R}^{3N+3}$. Since $N \geq n+1$, $\bar{A}(t)$ is a tall matrix and (31) is an over-determined system.

We next examine when $\bar{A}(t)$ has full column rank. Substituting (29) into $\bar{A}(t)$ yields

$$\bar{A}(t) = \left[\begin{array}{cccc|c} I_{3\times3} & tI_{3\times3} & \cdots & t^n I_{3\times3} & \rho(t) \\ 0_{3\times3} & I_{3\times3} & \cdots & nt^{n-1}I_{3\times3} & \rho^{'}(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{3\times3} & 0_{3\times3} & \cdots & n!I_{3\times3} & \rho^{(n)}(t) \\ \hline 0_{3\times3} & 0_{3\times3} & \cdots & 0_{3\times3} & \rho^{(n+1)}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{3\times3} & 0_{3\times3} & \cdots & 0_{3\times3} & \rho^{(N)}(t) \end{array}\right].$$

Since the top-left block of $\bar{A}(t)$ is a full-rank square matrix, $\bar{A}(t)$ has full column rank if and only if there exists $i \in \{n+1, \ldots, N\}$ such that

$$\rho^{(i)}(t) \neq 0_{3\times1}. \tag{33}$$

Since $\rho(t) = -g(t)/\theta(t)$ as shown in (30) and $g(t)/\theta(t) = (s_0 + s_1 t + \cdots + s_n t^n + h(t))/\ell$ as shown in (27), we can rewrite (33) to

$$-\frac{1}{\ell}(s_0 + s_1 t + \cdots + s_n t^n + h(t))^{(i)} \neq 0_{3\times1}. \tag{34}$$

Since $i \geq n+1$, (34) is equivalent to

$$h^{(i)}(t) \neq 0_{3\times1}. \tag{35}$$

According to the definition of $h(t)$ in (25), the condition in (35) is equivalent to

$$h(t) \neq 0_{3\times1}.$$

The proof is complete.

Some important remarks about Theorem 1 are given below.

1) The necessary and sufficient condition suggested by Theorem 1 is that the observer should have higher-order motion than the target. For example, when the target is stationary, the observer should move with a nonzero velocity. When the target moves with a constant velocity, the observer should move with a nonzero acceleration.

2) The necessary and sufficient condition given by Theorem 1 has a *key difference* from the bearing-only case that the higher-order motion in the bearing-angle case is *not* required to be orthogonal to the bearing vector, making the bearing-angle approach more flexible than the bearing-only one. For example, the bearing-angle approach can estimate the target's motion even if the observer simply moves along the bearing vector.

3) In the special case where the target moves with a constant velocity, the condition in Theorem 1 is consistent with the one obtained in Section 5. Although the condition in Theorem 1 allows more general target motion, the analysis in Section 5 is still meaningful since it is directly related to the dynamic model used in the pseudo-linear Kalman filter.

4) In practice, we would not estimate the target's motion by using the method of solving an equation like (31). That is because such a method involves calculating high-order derivatives, which are challenging to obtain accurately in practice. The role of this equation is to provide a fundamental perspective on whether there is sufficient information to uniquely recover the target's motion.

$$\tilde{A} \rightarrow \left[\begin{array}{cccccc:c} I & t_1 I & \cdots & t_1^{n-1} I & t_1^n I & & \rho(t_1) \\ 0 & I & \cdots & \Delta(t_2^{n-1}, t_1^{n-1}) I & \Delta(t_2^n, t_1^n) I & & \Delta(\rho(t_2), \rho(t_1)) \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (n-1)! I & \Delta^{n-1}(t_n^n, \cdots, t_1^n) & & \Delta^{n-1}(\rho(t_n), \cdots, \rho(t_1)) \\ 0 & 0 & \cdots & 0 & n! I & & \Delta^{n}(\rho(t_{n+1}), \cdots, \rho(t_1)) \\ \hdashline 0 & 0 & \cdots & 0 & 0 & & \Delta^{n+1}(\rho(t_{n+2}), \cdots, \rho(t_1)) \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & & \Delta^{N-1}(\rho(t_N), \cdots, \rho(t_1)) \end{array}\right] \tag{36}$$

6.3 Number of observations required

It is of practical importance to study how many discrete observations are required to recover the target's motion. Although Theorem 1 gives an observability condition, it does not answer this question because it is based on the continuous time domain. We next answer this question by exploring multiple discrete time steps.

Theorem 2. (Number of discrete observations). *If the observer's motion satisfies the observability condition in Theorem 1, it is necessary and sufficient to use at least $n+2$ observations to recover the target's motion. Here, n is the order of the target's polynomial motion as shown in (24).*

Proof 6.2. *Consider $t_1, \ldots, t_N$ time instances. Each time instance corresponds to an equation like (28): $A(t_i)X = h(t_i)$ for $i = 1, \ldots, N$. Combining these equations gives*

$$\tilde{A}X = \tilde{h}, \tag{37}$$

where

$$\tilde{A} = \begin{bmatrix} A(t_1) \\ \vdots \\ A(t_N) \end{bmatrix}, \qquad \tilde{h} \begin{bmatrix} h(t_1) \\ \vdots \\ h(t_N) \end{bmatrix}. \tag{38}$$

Here, $\tilde{A} \in \mathbb{R}^{(3N)\times(3n+4)}$ and $\tilde{h} \in \mathbb{R}^{3N}$.

(Necessity) *Since $X \in \mathbb{R}^{3n+4}$, we need at least $N \geq n+2$ observations so that $\tilde{A}$ is a tall matrix and hence (37) is an over-determined system.*

(Sufficiency) *Suppose we have $N \geq n+2$ discrete observations. Substituting (29) into (38) yields*

$$\tilde{A} = \begin{bmatrix} I_{3\times3} & t_1 I_{3\times3} & \cdots & t_1^n I_{3\times3} & \rho(t_1) \\ I_{3\times3} & t_2 I_{3\times3} & \cdots & t_2^n I_{3\times3} & \rho(t_2) \\ \vdots & \vdots & & \vdots & \vdots \\ I_{3\times3} & t_{n+1} I_{3\times3} & \cdots & t_{n+1}^n I_{3\times3} & \rho(t_{n+1}) \\ I_{3\times3} & t_{n+2} I_{3\times3} & \cdots & t_{n+2}^n I_{3\times3} & \rho(t_{n+2}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_{3\times3} & t_N I_{3\times3} & \cdots & t_N^n I_{3\times3} & \rho(t_N) \end{bmatrix}.$$

Starting from the last line in $\tilde{A}$, subtract the previous line from each subsequent line, and repeat this process. Finally, we can obtain (36) (the equation is too long and located at the top of another page). Here, Δ^n represents the nth-order time difference (Milne-Thomson 2000). For example, $\Delta(a_2, a_1) = (a_2 - a_1)/\delta t$, $\Delta^2(a_3, a_2, a_1) = \Delta(\Delta(a_3, a_2), \Delta(a_2, a_1)) = [(a_3 - a_2)/\delta t - (a_2 - a_1)/\delta t]/\delta t$. When δt is sufficiently small, the time difference is an approximation of the derivative. When the observability condition in Theorem 1 is satisfied, there exists $i \geq n+1$ such that $\rho^{(i)}(t) \neq 0$ as shown in (33). As a result, there exists $i \geq n+1$ such that

$$\Delta^i(\rho(t_{i+1}), \cdots, \rho(t_1)) \neq 0.$$

The above implication is valid because Δ^i is an approximation of the ith-order derivative when δt is sufficiently small. Then, $\tilde{A}$ in (36) has full column rank and hence (37) has a unique solution.

Theorem 2 suggests that when the target is stationary and hence $n = 0$, at least two discrete observations are sufficient to localize the target. This is true even if the two observations are acquired when the observer moves along the bearing vector. When the target moves with a constant velocity and hence $n = 1$, at least three discrete observations are sufficient to localize the target, which is consistent with the results in Section 5.

7 Numerical Simulation Results

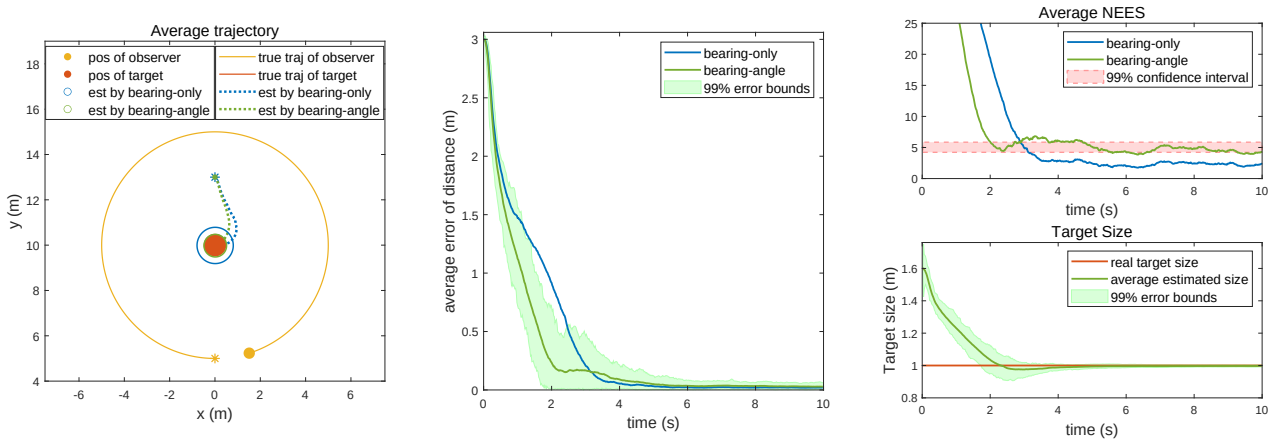
This section presents a set of numerical simulation results to demonstrate the effectiveness of the proposed bearing-angle approach.

The values of the parameters in two estimators are selected as $\sigma_v = 10^{-3}$, $\sigma_l = 10^{-4}$, $\sigma_\mu = 0.01$, and $\sigma_w = 0.01$. The selection of these values is inspired by the measurement noises obtained in the AirSim simulation and real-world experiments as shown later. The initial covariance matrix of the estimated states is set to $P(t_0) = 0.1I$. The target is a circle whose diameter is $\ell = 1$. The update rate of the system is 50 Hz. In addition, we use the same parameter values across all the simulation examples to verify the robustness of the algorithm. Better performances can be achieved if the parameters are well-tuned for specific scenarios. We perform $N_x = 100$ Monte Carlo simulations for each scenario.

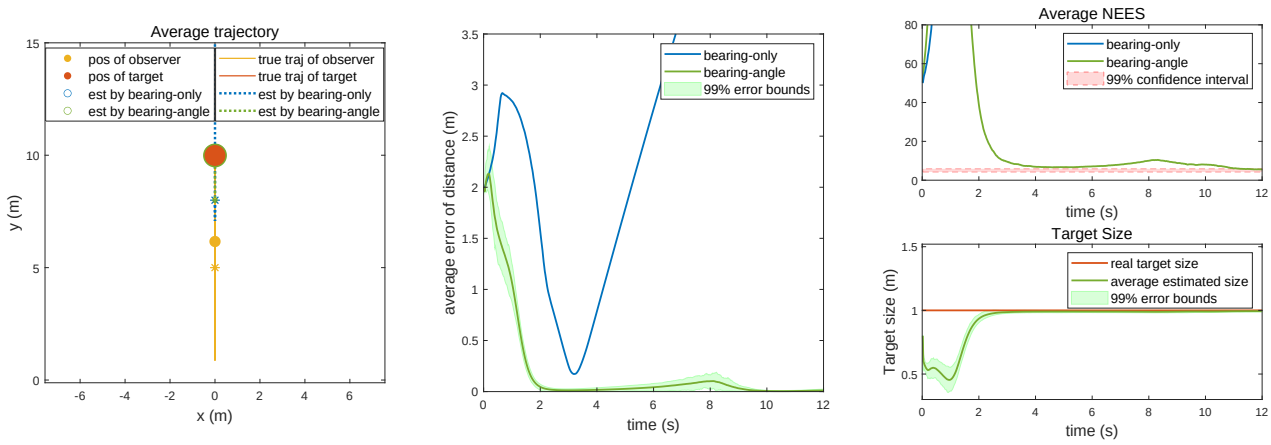
We use the normalized-estimation error squared (NEES) (Bar-Shalom and Li 1998) to analyze the consistency of the estimation algorithms. In particular, the value of the average NEES is

$$\bar{\epsilon}_{\text{NEES}} = \frac{1}{N_x} \sum_{i=1}^{N_x} (x - \hat{x}_i)^{\mathrm{T}} P_i^{-1} (x - \hat{x}_i), \tag{39}$$

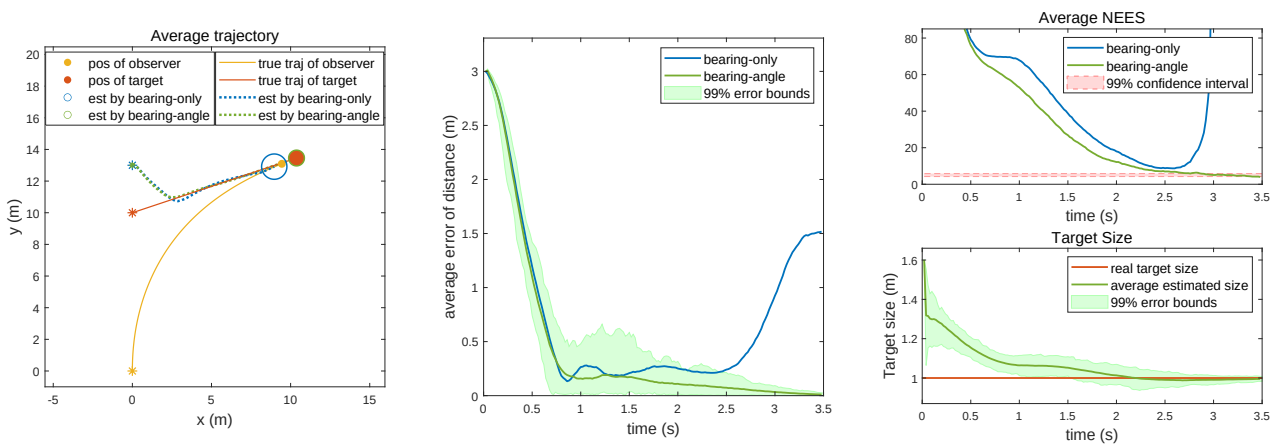
where $\hat{x}_i$ is the estimated states in the ith simulation, and P_i is the covariance matrix obtained from the estimator in the ith simulation.



(a) Scenario 1: Circular motion around the target. Both the bearing-only and bearing-angle approaches work well, but the bearing-angle one converges faster.



(b) Scenario 2: Straight motion towards and backwards the target. The bearing-only approach fails, but the bearing-angle approach works effectively.



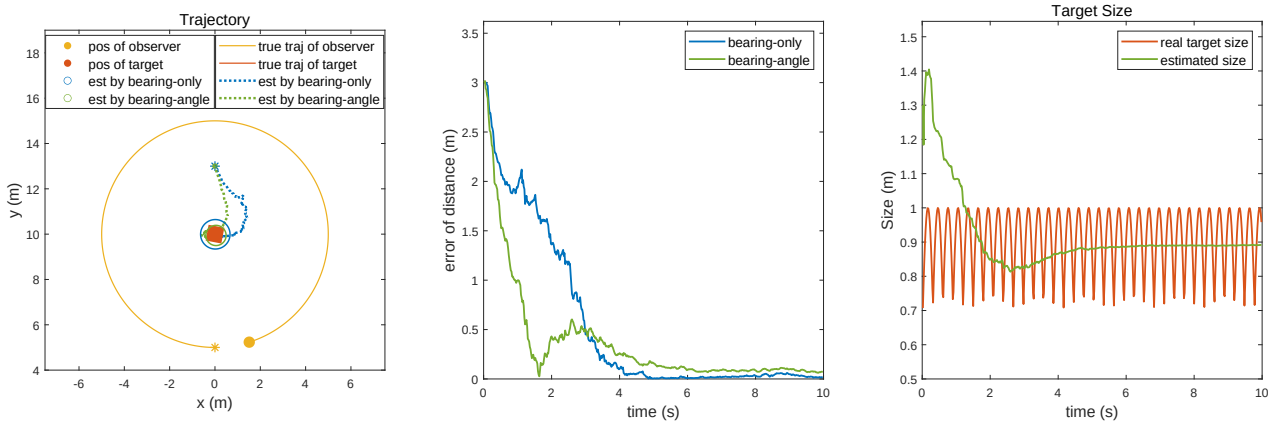
(c) Scenario 3: Approaching the target by a guidance law. The bearing-only approach works unstably, but the bearing-angle approach works effectively.

Figure 4. Numerical simulation results based on 100 Monte Carlo runs in three scenarios.

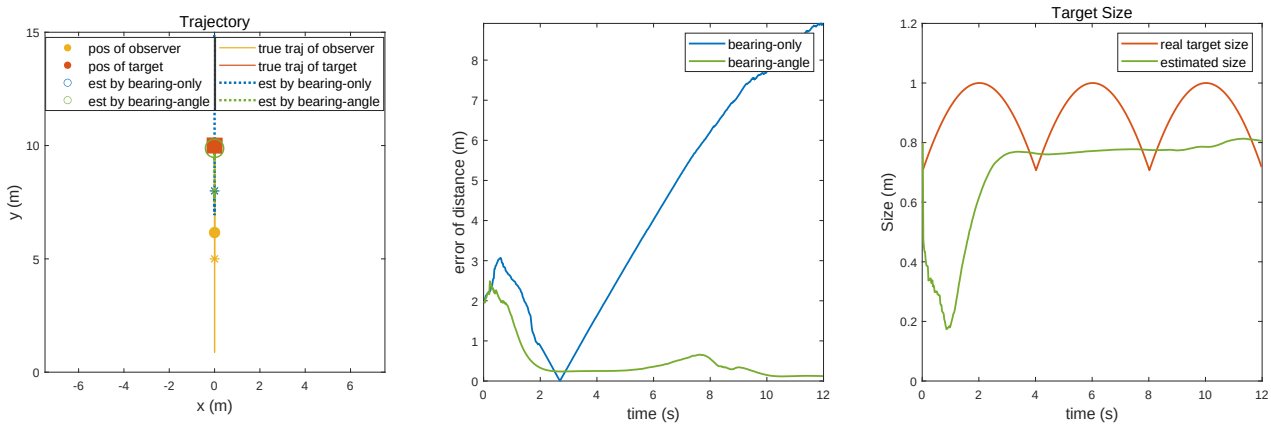
Finally, image acquisition and visual detection are not considered in these numerical simulation scenarios. They will be considered in Section 8 and Section 9.

7.1 Scenario 1: Circular motion around the target

In the first scenario, the target is stationary and located at $p_T = [0, 10]^{\mathrm{T}}$. The observer moves on a circle centered at the target with the speed of 3 m/s (see Fig. 4a). The radius of the circle is 5 m. The initial estimates are $\hat{p}_o(t_0) = [0, 13]^{\mathrm{T}}$, $\hat{v}_o(t_0) = [0, 0]^{\mathrm{T}}$, $\hat{\ell}(t_0) = 1.6$. During this process, the bearing vector varies while the angle subtended by the target remains constant. The angle measurement varies slightly due to the measurement noise. This scenario is favorable to the conventional bearing-only approach because its observability condition that the target should be viewed from different angles is well satisfied (Li et al. 2022).



(a) The observer moves around the square-shaped target. The target spins rapidly at 2π rad/s.



(b) The observer moves along the bearing vector. The target's spinning speed is $\pi/8$ rad/s.

Figure 5. Numerical simulation results for time-varying ℓ.

Fig. 4a shows the estimation results by the two approaches of bearing-only and bearing-angle. As can be seen, both algorithms perform well. The convergence of the bearing-angle approach is faster than the bearing-only one, as shown in the middle and right subfigures of Fig. 4a, due to the additional angle measurement. The bearing-angle approach can successfully estimate the size of the target as shown in the right subfigure of Fig. 4a.

7.2 Scenario 2: Straight motion towards and backwards the target repeatedly

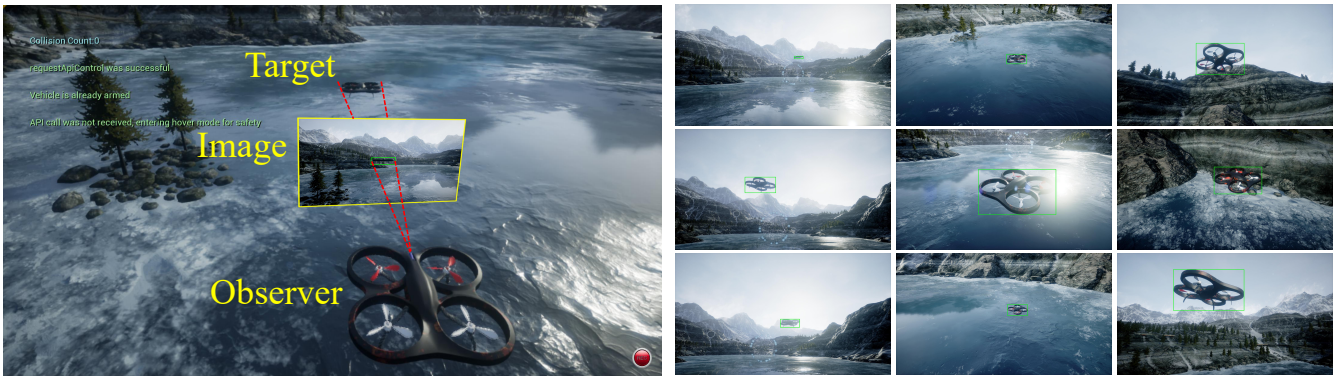
In the second scenario, the target is also stationary but the observer moves along a straight line towards and backwards the target repeatedly (Fig. 4b). During this process, the bearing vector remains constant while the angle varies. This scenario is most challenging for the bearing-only approach because its observability condition is not fulfilled.

In this simulation scenario, the target is stationary and located at $p_T(t_0) = [0, 10]^{\mathrm{T}}$. The observer moves along a straight line towards and backwards the target with a constant acceleration of -2 m/s^2. The initial conditions are $v_o(t_0) = [0, 4]^{\mathrm{T}}$ and $p_o(t_0) = [0, 5]^{\mathrm{T}}$. The initial estimates are $\hat{p}_o(t_0) = [0, 8]^{\mathrm{T}}$, $\hat{v}_o(t_0) = [0, 0]^{\mathrm{T}}$, $\hat{\ell}(t_0) = 0.8$. In this scenario, the true bearing of the target relative to the observer remains unchanged though the bearing measurement may vary slightly due to the measurement noise.

Fig. 4b shows the estimation results of the bearing-only and bearing-angle approaches. As can be seen, the bearing-only approach diverges since its observability condition is not satisfied. By contrast, the proposed bearing-angle approach converges, and is able to localize the target and estimate its size, which demonstrates the strong observability of the bearing-angle approach. One may notice that the estimated size and the NEES value get worse first before converging. This is because the noise level of the angle is set to be constant. Since the angle is small in the beginning, the noise-angle ratio is large, causing a relatively large NEES value.

7.3 Scenario 3: Approaching the target by a guidance law

The third scenario is more complex than the first two. Here, the target moves with a constant velocity where the observer is controlled by a proportional navigation guidance (PNG) law to approach the target (Fig. 4c). During this process, both the bearing and angle vary. This scenario is also challenging for the bearing-only approach because its observability is weak due to the fact that the lateral motion of the observer is small. Many researchers have studied how to add extra control commands to the PNG to enhance the observability based on the bearing-only approach (Song and Um 1996; Seo and Tahk 2015; Lee et al. 2015).



(a) An AirSim simulation experimental scenario.

(b) Samples of the dataset collected automatically in AirSim.

Figure 6. The setup of the AirSim simulation experiments.

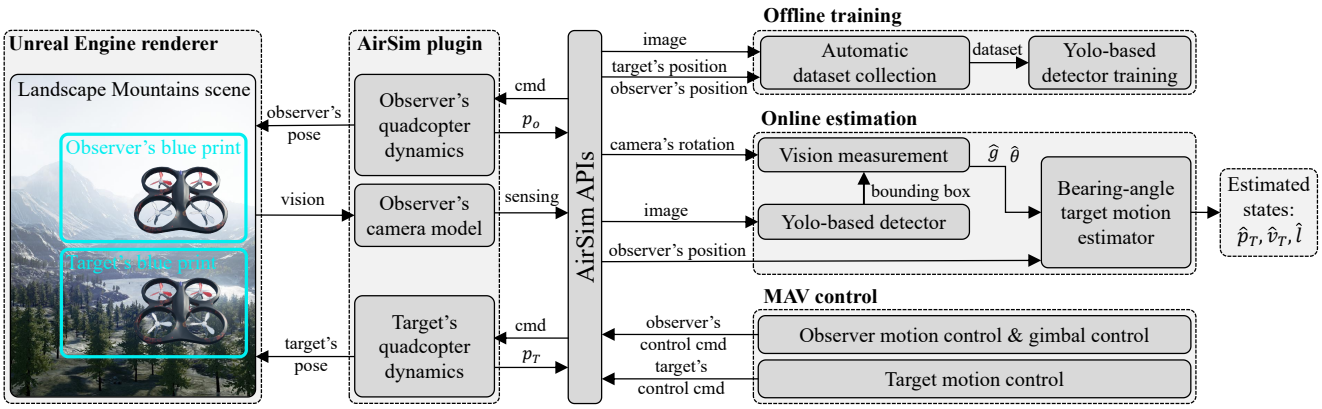


Figure 7. The software architecture of the AirSim simulation system. AirSim is a plugin for Unreal Engine. Three programmed modules (Offline training, Online estimation, and MAV control) communicate with the AirSim plugin through APIs.

In this simulation scenario, the target moves along a straight line with a constant velocity $v_T = [1/\sqrt{2}, 1/\sqrt{2}]^{\mathrm{T}}$. The observer's velocity magnitude is constantly 3 m/s while the velocity direction is controlled by a PNG law. The navigation gain of the PNG law is selected as one. The initial estimates of the target's states are the same as Scenario 1. The simulation stops just before the observer collides with the target.

Fig. 4c shows the estimation results by the bearing-only and bearing-angle approaches. As can be seen, the bearing-angle algorithm successfully converges before the collision occurs, but the bearing-only algorithm fails to estimate the target's states due to its weak observability. This simulation example demonstrates that the bearing-angle algorithm can be used directly in the guidance scenario without extra maneuvers required by the bearing-only approach (Song and Um 1996; Seo and Tahk 2015; Lee et al. 2015).

7.4 Simulation results for time-varying ℓ

Although ℓ is assumed to be invariant, it is meaningful to challenge the proposed bearing-angle approach by considering time-varying ℓ. We will see through simulation examples that the bearing-angle approach is still effective when ℓ varies slowly. It becomes unstable when ℓ varies rapidly since the assumption of invariant ℓ is severely invalid.

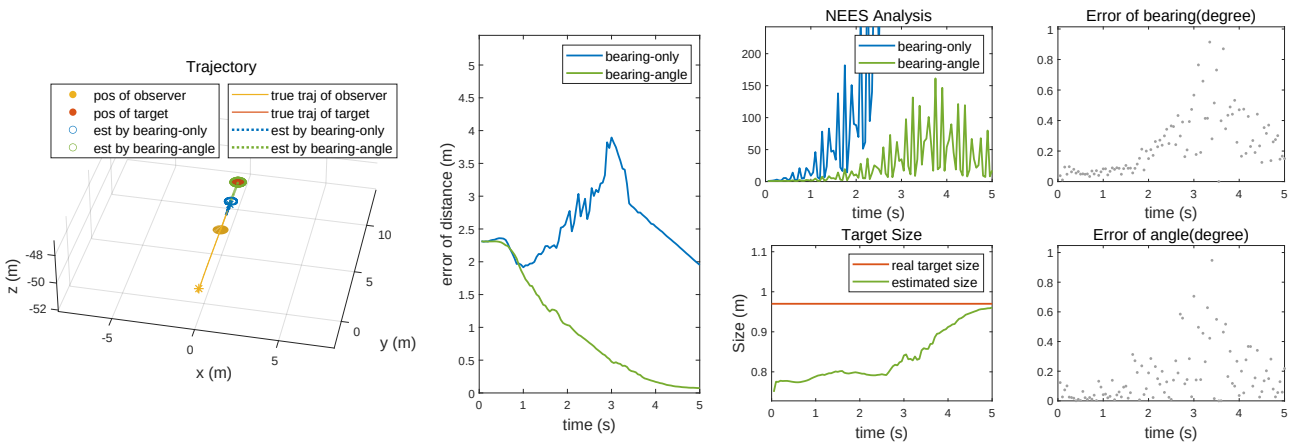
Suppose that the target object has a square shape. Then, ℓ varies when the object is observed from different viewing angles or the object spins. Fig. 5a shows a scenario where the observer moves around the target, whose spinning speed is 2π rad/s. The red curve in the right subfigure represents the true value of ℓ, which varies rapidly. As can be seen, the bearing-angle algorithm works effectively though there is a small estimation bias. Fig. 5b shows a scenario where the observer moves along the bearing vector. The spinning speed of the target object is $\pi/8$ rad/s. As can be seen, the bearing-only approach diverges due to the lack of observability. The bearing-angle algorithm can still converge since ℓ varies slowly. When we further increase the spinning speed of the target, the bearing-angle algorithm will also diverge because the algorithm cannot distinguish whether the change of θ is caused by the change of ℓ or the change of r.

8 AirSim Simulation Results

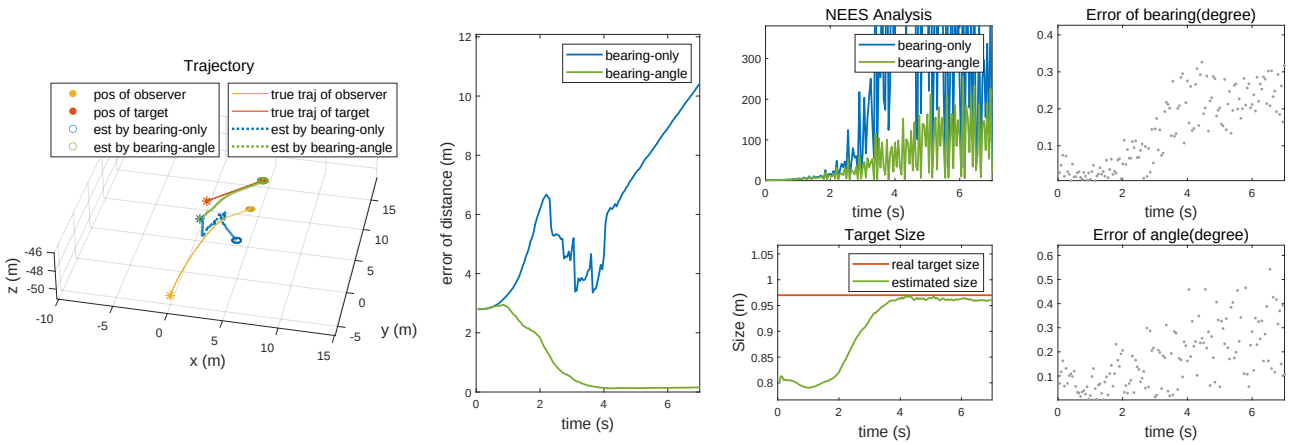
In this section, we show simulation results under a more realistic setup. In particular, the simulation is based on AirSim, a simulator that can provide high-quality visual simulation (Shah et al. 2017). Nonlinear MAV dynamics and control are also considered.

8.1 Simulation setup

Fig. 6a shows an AirSim simulation scenario. As can be seen, there are two flying quadcopter MAVs. The observer MAV can capture images of the target MAV using its simulated onboard camera. A simple gimbal camera controller is



(a) The target MAV hovers stationarily, while the observer MAV approaches the target MAV under the control of (40).



(b) The target MAV moves with a constant velocity, while the observer MAV follows the target MAV under the control of (40).

Figure 8. AirSim simulation results in the approaching and following scenarios.

implemented so that the target MAV is always located inside the field of view of the camera. The visual environment used in the simulation is called Landscape Mountains, which includes realistic mountains, lakes, trees, and roads. Other environments can also be used if needed.

The bearing and angle measurements are obtained from the bounding boxes generated by a Yolo-based detection algorithm. A tiny-YOLO v4 network (Bochkovskiy et al. 2020) is trained to detect the target MAV in the images. Although the visual detector can be replaced by other state-of-the-art ones, the tiny-YOLO v4 network is already sufficient to verify our proposed approach. The architecture of the entire simulation system is shown in Fig. 7. The system consists of the modules of automatic image dataset collection, Yolo-based target detection, gimbal camera control, nonlinear quadcopter dynamics, and quadcopter flight control. The quadcopter dynamics and flight control used in the simulation are similar to (Meier et al. 2011; Shah et al. 2017) and omitted here due to space limitation. The quadcopter's physical size varies slightly when viewed from different directions, although it is assumed to be invariant. All of these factors make the Airsim simulation more realistic and challenging.

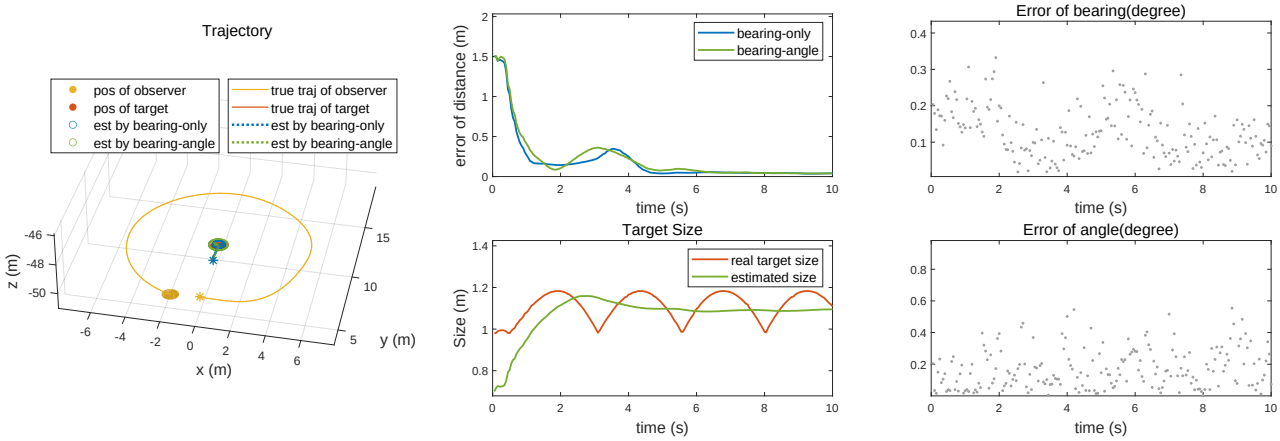
8.2 Automatic dataset collection

To train the Yolo-based detector, we developed a module to automatically collect an image dataset. This module has some advantages. First, it is efficient. More than ten thousand labeled images can be collected automatically in 24 hours. Second, it is flexible. It can acquire images with random target's positions, random target's attitudes, random camera's view angles, and random background scenes. These images are beneficial to achieve a good generalization ability of the detector. Third, the image labels are of high quality. Since the ground truth of the target's image is known in the simulation, the generated bounding box is tight. The collected dataset contains 17,000 labeled images (see Fig. 6b). The resolution of the images is 1536×864 pixels. The simulation system was deployed on a Dell Precision 7920 Tower Workstation with two NVIDIA Quadro GV100 graphic cards. Since the dataset is sufficient and high-quality, the detection can achieve the accuracy of mAP=99.5%.

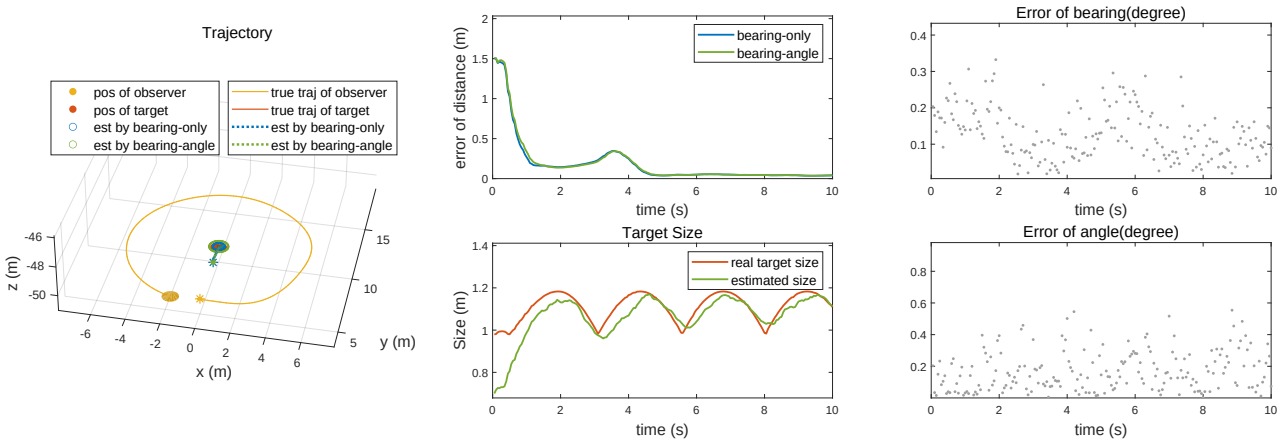
8.3 Scenario 1: Approaching and following the target

We first consider the scenarios where the observer MAV approaches or follows a target MAV. These scenarios widely exist in practical applications such as aerial target pursuit.

We show two simulation examples in Fig. 8a and Fig. 8b, respectively. In both examples, the observer is controlled by



(a) Estimation results when $\sigma_l = 10^{-4}$ and the other parameters are the same as those in Section 8.3.



(b) Estimation results when $\sigma_l = 0.01$ and the other parameters are the same as those in Section 8.3.

Figure 9. AirSim simulation results in the circular motion scenario where ℓ varies.

a controller so that it can approach the target and maintain a desired separation. In particular, the controller is

$$v_o^{\text{cmd}}(t) = v_T(t) + k^{\text{track}} \frac{r^2(t) - r_d^2}{r^2(t)} g(t), \tag{40}$$

where $v_o^{\text{cmd}}(t)$ is the velocity command of the observer MAV, $k^{\text{track}} = 3$ is the control gain, and $r_d = 3$ is the desired separation. The magnitude of the observer's velocity is bounded from above by 3 m/s. It should be noted that (40) relies on the true position and velocity of the target MAV in the simulation. Therefore, the data is collected first and then processed offline so that we can compare the performances of the bearing-only and bearing-angle approaches.

In the first example, the target MAV hovers constantly at $p_T(t_0) = [0, 10, 10]^{\text{T}}$. The observer MAV moves along a straight line toward the target with a decreasing velocity command. Since the bearing of the target MAV remains the same, this example is challenging for the bearing-only approach. As shown in Fig. 8a, the bearing-only approach fails to converge while the bearing-angle approach can successfully estimate the target's motion.

In the second example, the target MAV moves with a constant velocity of $v_T = [1/\sqrt{2}, 1/\sqrt{2}, 0]^{\text{T}}$. The trajectory of the observer MAV under the control of (40) is still close to (though not strictly) a straight line. As a result, the observability is weak by the bearing-only approach. As shown in Fig. 8b, the bearing-angle approach successfully converges while the bearing-only one fails. It is notable that ℓ is invariant in the first example and varies slowly in the second example.

It is worth mentioning that the detection results used in the estimation algorithms are obtained from the Yolo-based estimator. The ground truth obtained from AirSim is only used to calculate the errors of measurements, as shown in the right figures of Figs. 8a and 8b. It is not surprising that the measurement noises are not strictly Gaussian since the 2D bounding box is generated by a deep learning vision algorithm. It is noticed that the noises are inversely correlated to the observer-target range. This is reasonable because, when the target is close to the camera and hence its image is large, the center point and the size of the bounding box usually vary for a few pixels.

The NEES values are also shown in Fig. 8. As can be seen, the NEES value of the bearing-only approach diverges. The NEES value of the bearing-angle approach oscillates and converges slowly. The reasons are analyzed as follows. Compared to the Matlab-based numerical simulation, the visual measurements here are generated by deep learning algorithms, and the measurement noises are non-Gaussian. The non-Gaussian noises propagate into P in (39) since the calculation of P relies on noisy measurements. The noises may also cause an estimation bias that can further aggravate the NEES error. Moreover, although the system is observable in the two simulation examples, the observability

is relatively weak compared to the case where the observer moves surrounding the target. As a result, the matrix P may not be able to perfectly describe the estimation accuracy. These elements may jointly cause the convergence behavior of the NEES values shown in Fig. 8.

8.4 Scenario 2: Circular motion and varying ℓ

We next examine a case where ℓ is time-varying. In particular, suppose a target quadcopter MAV hovers constantly at $p_T(t_0) = [0, 10, 10]^{\mathrm{T}}$. The observer MAV moves on a circle centered at the target (Fig. 9a). Since the target quadcopter MAV has a square shape from the top view, its size ℓ is time-varying when viewed from side angles (see the red curves in the middle subfigure of Fig. 9a).

We show two simulation examples in Fig. 9a and Fig. 9b, respectively. The two simulation examples share the same measurement data but different values of σ_ℓ. Moreover, the other parameters are the same as those in Section 8.3.

In the first simulation example, σ_ℓ is set to be a small value: $\sigma_\ell = 10^{-4}$. Its interpretation is that ℓ is treated as invariant during the process. In this case, the performance of the bearing-angle approach is almost the same as the bearing-only one as shown in Fig. 9a. Since ℓ is treated to be invariant, the estimated value $\hat{\ell}$ converges to a constant which is the mean value of the time-varying ℓ.

In the second simulation example, the value of σ_ℓ is larger than the first example: $\sigma_\ell = 0.01$. Its interpretation is that ℓ is believed to be time-varying during the process. In this case, the performance of the bearing-angle approach is still almost the same as the bearing-only one. Moreover, since σ_ℓ is large, the bearing-angle approach can successfully estimate the true time-varying value of ℓ.

In summary, in the case where ℓ varies slowly, the bearing-angle approach would degenerate to the bearing-only one. The fundamental reason is that the extra information embedded in the angle measurement is used to estimate the time-varying ℓ rather than improving the observability of the target's motion.

9 Real-World Experimental Results

In this section, two sets of real-world experiments are presented to further verify the effectiveness of the approach. The first is based on a hand-held camera and a ground robot. The second is based on two quadcopter MAVs. The second experimental scenario is motivated by aerial target pursuit tasks.

9.1 Experiment 1: Hand-held camera

The experimental setup is shown in Fig. 10a. The observer is a hand-held camera (Hik Vision DS-E14S) connected to a laptop. The camera's intrinsic parameters are calibrated beforehand. The robot built on Mecanum wheels can move in any direction on the ground under velocity control. The ground truth of the states of the camera and the robot are provided by a Vicon indoor motion capture system. The key experimental specifications are listed in Table 1.

A dataset of 5,514 images was collected and used to train a tiny-YOLO v4 network to detect the target robot (see Fig. 10b). The detection precision of the trained network is mAP=99.8%. In the experiment, the target robot is commanded to move with a constant velocity. In the meantime, a person holding the camera moves along some trajectories. Two different cases are studied. In both of the cases, the target robot moves with a constant velocity $v_T = [-0.1, 0.1, 0]^{\mathrm{T}}$. The noises of the measurements are calculated based on the ground truth provided by the Vicon system. The noises are shown in the right subfigures of Fig. 11a and Fig. 11b.

In the first case, the camera is held about 1.5 meters above the ground and moves around the target robot. In this case, the bearing vector varies sufficiently and hence the observability conditions for the bearing-only and bearing-angle approaches are both well satisfied. As shown in Fig. 11a, both approaches perform well in this case while the bearing-angle approach performs slightly better than the bearing-only one.

In the second case, the camera moves along the trajectory of the robot by getting close or far from it periodically. In this case, the angle varies significantly, but the bearing does not. Without surprise, the bearing-only approach performs poorly in this case due to weak observability (Fig. 11b). By contrast, the bearing-angle approach can perform stably due to its enhanced observability.

Table 1. Key specifications of the *indoor* hardware system.

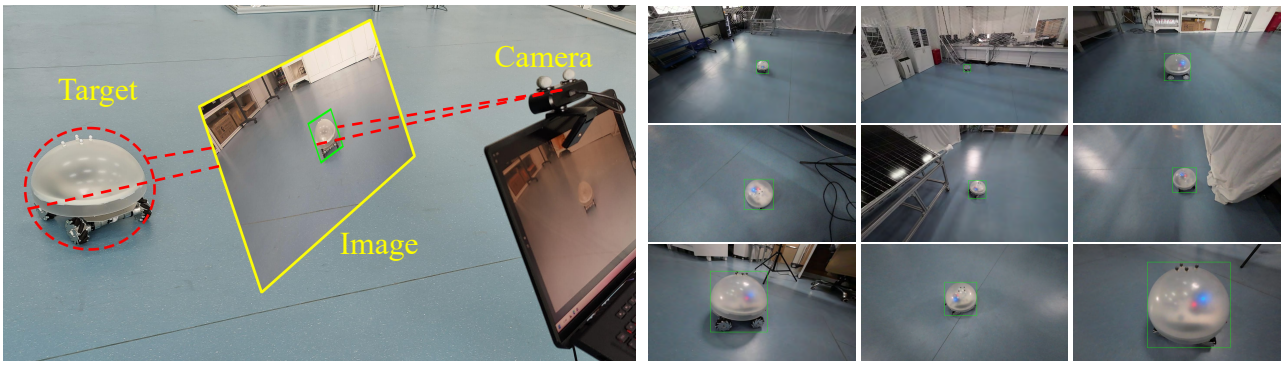
	Parameter	Value	Unit
Camera	Resolution	640× 480	pixel
	Max frequency	30	fps
Robot	Max speed	1	m/s
	Diameter size	295	mm
Vicon	Localization accuracy	1	mm
	Max frequency	100	Hz

Table 2. Key specifications of the *outdoor* hardware system.

	Parameter	Value	Unit
M300 quadcopter	Diagonal size	895	mm
	Total mass	7.4	kg
	Max pitch/roll	30	degree
	Max flight time	30	minutes
RTK	Accuracy	1	cm
	Max frequency	10	Hz
H20 gimbal & camera	Resolution	1920×1080	pixel
	Frequency	15	Hz
	Max angular rate	180	deg/s

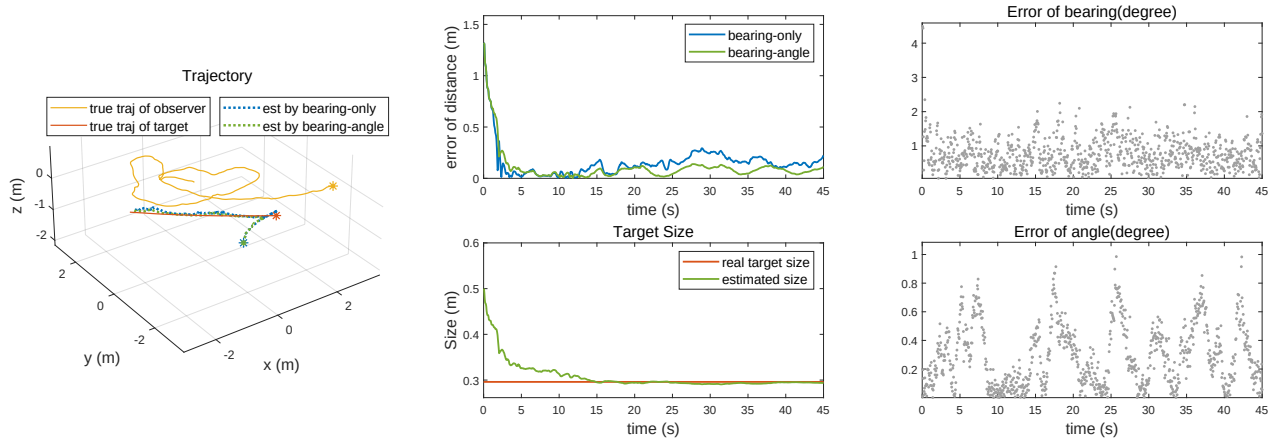
9.2 Experiment 2: MAV-following-MAV

Two MAV platforms were developed based on DJI M300 quadcopters (Fig. 12a). The MAV platforms are equipped with RTK GPS modules for accurate self-localization, an H20 camera for visual detection, a Manifold 2G onboard computer for onboard flight control, and a Zigbee module for wireless communication. Some key specifications of the MAV platforms are listed in Table 2. The structure of the hardware perception and communication system is illustrated in Fig. 12c. The target MAV is also equipped with an RTK GPS module, whose measurements are used as the ground truth to calculate the noises of the visual

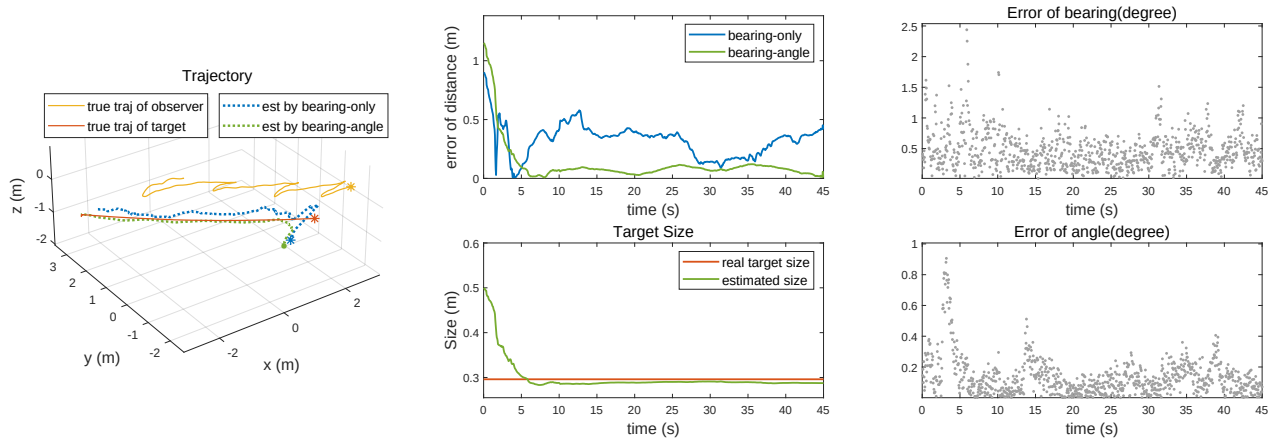


(a) Experimental setup **(b)** Samples in the dataset

Figure 10. The setup of the experiments based on a hand-held camera.



(a) Case 1: The observer moves around the target. Both the bearing-only and bearing-angle approaches work well.



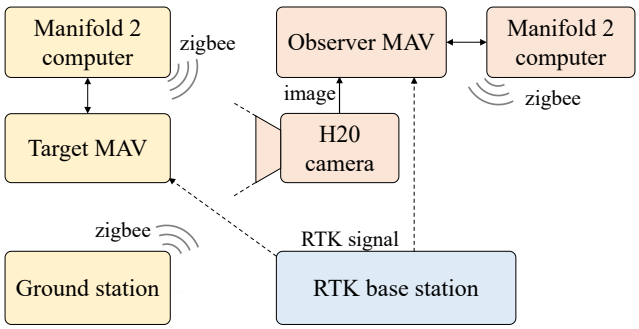
(b) Case 2: The observer moves close or far from the target periodically. The bearing-angle approach performs effectively, but the bearing-only approach works unstably.

Figure 11. Experimental results based on a hand-held camera.

measurements. The noises are shown in the right subfigure of Fig. 13.

The experiment consists of two stages: data acquisition and offline data processing. In the data acquisition stage, the target MAV is commanded to fly with a constant velocity, and the observer MAV is automatically controlled to follow the target MAV to maintain a constant distance from the target. More specifically, the procedure of the flight experiment is as follows. Initially, the observer MAV is placed about 20 meters behind the target MAV on the ground. Then, the two MAVs take off and fly to the same specified height automatically upon a takeoff command sent from the ground control station. After they have reached the desired height, all deployed algorithms are activated. Then, the target MAV moves with a constant velocity of $v_T = [1/\sqrt{2}, 1/\sqrt{2}, 0]^{\mathrm{T}}$. The observer MAV approaches the target by the controller in (40). It takes the observer MAV about eight seconds to reach the desired relative distance. Then, the two MAVs fly with the same velocity and remain relatively stationary for

(a) Hardware platforms **(b)** Samples of the images in the dataset



(c) System architecture

Figure 12. The setup of the MAV-following-MAV experiment.

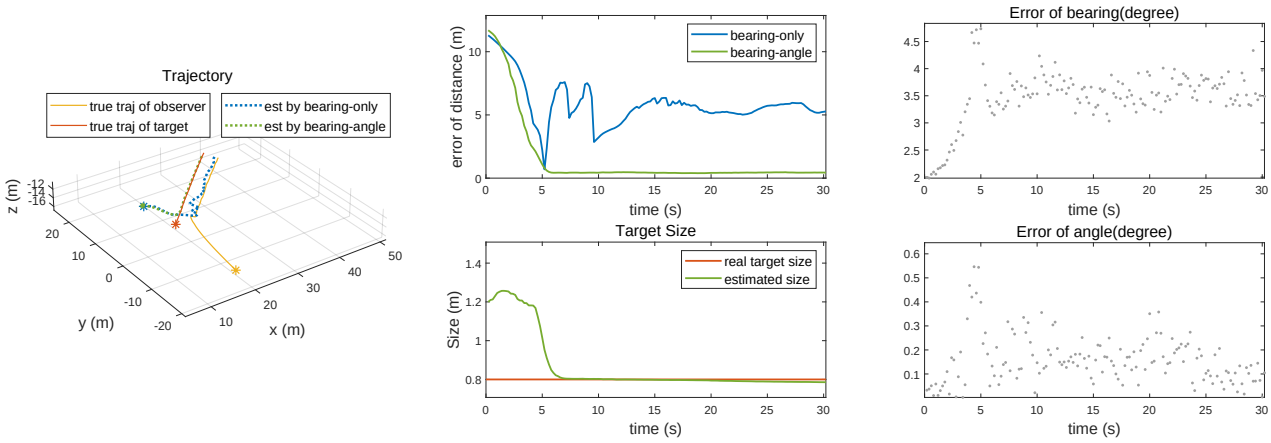


Figure 13. The results of the MAV-following-MAV experiments. The bearing-angle approach performs effectively, but the bearing-only approach works unstably.

another 20 seconds. Finally, the ground station sends a stop command and the two MAVs return and land automatically.

During the flight, the gimbal camera of the observer MAV is automatically controlled so that the target MAV is maintained in the field of view. It is noted that the control of the gimbal camera and the observer MAV is not based on the visual detection results. Instead, the control is based on the measurements provided by the RKT GPS and inter-MAV wireless communication. In this way, we can analyze the image and flight data offline and compare the performance of the two approaches of bearing-angle and bearing-only. The acquired images and flight data are saved on the onboard computer during the flight. A dataset of 5,341 images was collected (Fig. 12b) and used to train a tiny-YOLO v4 network. The detection precision of the trained network reaches mAP=99.8%.

The experimental results are shown in Fig. 13. As can be seen, the bearing-angle approach performs well. By contrast, the bearing-only approach only works well before the observer MAV reached the desired position relative to the target MAV. That is because the bearing varies significantly during this process due to the fluctuation of the observer's motion caused by the flight control. However, the bearing-only approach diverges quickly thereafter when the bearing stops varying significantly.

10 Conclusion

Motivated by the limitation of the existing bearing-only approach, this paper proposed and analyzed a novel bearing-angle approach for vision-based target motion estimation. We showed that the observability by the bearing-angle approach is significantly enhanced compared to the bearing-only one. As a result, the requirement of the observer's extra motion for observability enhancement can be significantly relaxed. As we showed in various experiments, the bearing-angle approach can successfully estimate the target's motion in many scenarios where the bearing-only approach fails. The enhanced observability of the bearing-angle approach comes with no additional cost since almost all vision detection algorithms can generate bounding boxes. One assumption adopted in the bearing-angle approach is that the target's physical size is invariant to different viewing angles. Although this assumption is approximately valid in many tasks as demonstrated in this paper, it is meaningful to study how to relax or remove this assumption in the future.

Acknowledgements

The authors would like to thank the anonymous reviewers and the editors for their helpful advice for improving this work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the Hangzhou Key Technology Research and Development Program (Grant 20200416A16), and the Research Center for Industries of the Future at Westlake University (Grant WU2022C027).

References

Aidala V (1979) Kalman filter behavior in bearings-only tracking applications. *IEEE Transactions on Aerospace and Electronic Systems* AES-15(1): 29–39. DOI:10.1109/taes.1979.308793.

Aidala V and Nardone S (1982) Biased estimation properties of the pseudolinear tracking filter. *IEEE Transactions on Aerospace and Electronic Systems* AES-18(4): 432–441. DOI:10.1109/taes.1982.309250.

Anjaly P and Ratnoo A (2018) Observability enhancement of maneuvering target with bearings-only information. *Journal of Guidance, Control, and Dynamics* 41(1): 184–198. DOI:10.2514/1.g003003.

Avidan S and Shashua A (2000) Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(4): 348–357. DOI:10.1109/34.845377.

Bar-Shalom Y and Li X (1998) *Estimation and Tracking: Principles, Techniques, and Software*. Artech House. ISBN 9780964831216.

Bochkovskiy A, Wang CY and Liao HYM (2020) YOLOv4: Optimal speed and accuracy of object detection .

Brighton CH and Taylor GK (2019) Hawks steer attacks using a guidance system tuned for close pursuit of erratically manoeuvring targets. *Nature Communications* 10(1). DOI:10.1038/s41467-019-10454-z.

Doğançay K (2005) Bearings-only target localization using total least squares. *Signal Process* 85(9): 1695–1710. DOI:10.1016/j.sigpro.2005.03.007.

Dressel L and Kochenderfer MJ (2019) Hunting drones with other drones: Tracking a moving radio target. In: *2019 International Conference on Robotics and Automation (ICRA)*. DOI:10.1109/icra.2019.8794243.

Farina A (1999) Target tracking with bearings – only measurements. *Signal Process* 78(1): 61–78. DOI:10.1016/s0165-1684(99)00047-x.

Fogel E and Gavish M (1988) Nth-order dynamics target observability from angle measurements. *IEEE Transactions on Aerospace and Electronic Systems* 24(3): 305–308. DOI:10.1109/7.192098.

Griffin BA and Corso JJ (2021) Depth from camera motion and object detection. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI:10.1109/cvpr46437.2021.00145.

Hammel S, Liu P, Hilliard E and Gong K (1989) Optimal observer motion for localization with bearing measurements. *Computers & Mathematics with Applications* 18(1-3): 171–180. DOI:10.1016/0898-1221(89)90134-x.

He S, Shin HS and Tsourdos A (2019) Trajectory optimization for target localization with bearing-only measurement. *IEEE Transactions on Robotics* 35(3): 653–668. DOI:10.1109/tro.2019.2896436.

He S, Wang J and Lin D (2018) Three-dimensional bias-compensation pseudomeasurement kalman filter for bearing-only measurement. *Journal of Guidance, Control, and Dynamics* 41(12): 2678–2686. DOI:10.2514/1.g003785.

Hoelzer H, Johnson G and Cohen A (1978) Modified polar coordinates-the key to well behaved bearings only ranging. *IR & D Report* : 78–M19.

Kaminski JY and Teicher M (2004) A general framework for trajectory triangulation. *Journal of Mathematical Imaging and Vision* 21(1): 27–41. DOI:10.1023/b:jmiv.0000026555.79056.b8.

Kulikov GY and Kulikova MV (2018) Moore-penrose-pseudo-inverse-based kalman-like filtering methods for estimation of stiff continuous-discrete stochastic systems with ill-conditioned measurements. *IET Control Theory & Applications* 12(16): 2205–2212. DOI:10.1049/iet-cta.2018.5404.

Lee CH, Kim TH and Tahk MJ (2010) Observability analysis of advanced guidance laws with bearing-only measurement. *IFAC Proceedings Volumes* 43(15): 136–141. DOI:10.3182/20100906-5-jp-2022.00024.

Lee CH, Kim TH and Tahk MJ (2015) Biased PNG for target observability enhancement against nonmaneuvering targets. *IEEE Transactions on Aerospace and Electronic Systems* 51(1): 2–17. DOI:10.1109/taes.2014.120103.

Li J, Ning Z, He S, Lee CH and Zhao S (2022) Three-dimensional bearing-only target following via observability-enhanced helical guidance. *IEEE Transactions on Robotics* : 1–18DOI:10.1109/tro.2022.3218268.

Lin X, Kirubarajan T, Bar-Shalom Y and Maskell S (2002) Comparison of EKF, pseudomeasurement, and particle filters

for a bearing-only target tracking problem. In: Drummond OE (ed.) *SPIE Proceedings*. DOI:10.1117/12.478508.

Lingren A and Gong K (1978) Position and velocity estimation via bearing observations. *IEEE Transactions on Aerospace and Electronic Systems* AES-14(4): 564–577. DOI:10.1109/taes.1978.308681.

Ma Y, Soatto S, Kosecká J and Sastry SS (2012) *An Invitation to 3-D Vision*. Springer New York. ISBN 9780387217796.

Meier L, Tanskanen P, Fraundorfer F and Pollefeys M (2011) PIXHAWK: A system for autonomous flight using onboard computer vision. In: *2011 IEEE International Conference on Robotics and Automation*. DOI:10.1109/icra.2011.5980229.

Milne-Thomson LM (2000) *The calculus of finite differences*. AMS Chelsea Pub. ISBN 0-8218-2107-5.

Nardone S and Aidala V (1981) Observability criteria for bearings-only target motion analysis. *IEEE Transactions on Aerospace and Electronic Systems* AES-17(2): 162–166. DOI:10.1109/taes.1981.309141.

Park HS, Shiratori T, Matthews I and Sheikh Y (2015) 3D trajectory reconstruction under perspective projection. *International Journal of Computer Vision* 115(2): 115–135. DOI:10.1007/s11263-015-0804-2.

Ponda S, Kolacinski R and Frazzoli E (2009) Trajectory optimization for target localization using small unmanned aerial vehicles. In: *AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics and Astronautics. DOI:10.2514/6.2009-6015.

Qiu K, Qin T, Gao W and Shen S (2019) Tracking 3-D motion of dynamic objects using monocular visual-inertial sensing. *IEEE Transactions on Robotics* 35(4): 799–816. DOI:10.1109/tro.2019.2909085.

Rothe J, Strohmeier M and Montenegro S (2019) A concept for catching drones with a net carried by cooperative UAVs. In: *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. DOI:10.1109/ssrr.2019.8848973.

Sabet M, Fathi A and Daniali HM (2016) Optimal design of the own ship maneuver in the bearing-only target motion analysis problem using a heuristically supervised extended kalman filter. *Ocean Engineering* 123: 146–153. DOI:10.1016/j.oceaneng.2016.07.028.

Seo MG and Tahk MJ (2015) Observability analysis and enhancement of radome aberration estimation with line-of-sight angle-only measurement. *IEEE Transactions on Aerospace and Electronic Systems* 51(4): 3321–3331. DOI:10.1109/taes.2015.140531.

Shah S, Dey D, Lovett C and Kapoor A (2017) AirSim: High-fidelity visual and physical simulation for autonomous vehicles. In: *Field and Service Robotics*. Springer International Publishing, pp. 621–635. DOI:10.1007/978-3-319-67361-5_40.

Song TL and Um TY (1996) Practical guidance for homing missiles with bearings-only measurements. *IEEE Transactions on Aerospace and Electronic Systems* 32(1): 434–443. DOI:10.1109/7.481284.

Stallard DV (1991) Angle-only tracking filter in modified spherical coordinates. *Journal of Guidance, Control, and Dynamics* 14(3): 694–696. DOI:10.2514/3.20699.

Tekin B, Sinha SN and Fua P (2018) Real-time seamless single shot 6D object pose prediction. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. DOI:10.1109/cvpr.2018.00038.

Vrba M and Saska M (2020) Marker-less micro aerial vehicle detection and localization using convolutional neural networks. *IEEE Robotics and Automation Letters* 5(2): 2459–2466. DOI:10.1109/lra.2020.2972819.

Woffinden DC and Geller DK (2009) Optimal orbital rendezvous maneuvering for angles-only navigation. *Journal of Guidance, Control, and Dynamics* 32(4): 1382–1387. DOI:10.2514/1.45006.

Yang S and Scherer S (2019) CubeSLAM: Monocular 3-d object SLAM. *IEEE Transactions on Robotics* 35(4): 925–938. DOI:10.1109/tro.2019.2909168.

Yoshikawa T (1972) On discrete-time kalman filter in singular case and a kind of pseudo-inverse of a matrix†. *International Journal of Control* 15(6): 1157–1163. DOI:10.1080/00207177208932228.

Yu Q, Shang Y, Zhou J, Zhang X and Li L (2009) Monocular trajectory intersection method for 3D motion measurement of a point target. *Science in China Series E: Technological Sciences* 52(12): 3454–3463. DOI:10.1007/s11431-009-0239-5.

Zhao S, Lin F, Peng K, Dong X, Chen BM and Lee TH (2015) Vision-aided estimation of attitude, velocity, and inertial measurement bias for UAV stabilization. *Journal of Intelligent & Robotic Systems* 81(3-4): 531–549. DOI:10.1007/s10846-015-0206-2.

Zhao S and Zelazo D (2019) Bearing rigidity theory and its applications for control and estimation of network systems: Life beyond distance rigidity. *IEEE Control Systems Magazine* 39(2): 66–83. DOI:10.1109/mcs.2018.2888681.