

Article

# A Study on Vision-Based Backstepping Control for a Target Tracking System

Thinh Huynh <sup>1</sup>, Minh-Thien Tran <sup>1</sup>, Dong-Hun Lee <sup>2</sup>, Soumayya Chakir <sup>2,\*</sup> and Young-Bok Kim <sup>1,\*</sup>

<sup>1</sup> Department of Smart Robot Convergence and Application Engineering, Pukyong National University, Nam-gu, Busan 48513, Korea; huynhtinh@hcmute.edu.vn (T.H.); minhthien@pukyong.ac.kr (M.-T.T.)

<sup>2</sup> Department of Mechanical System Engineering, Pukyong National University, Nam-gu, Busan 48513, Korea; iottoi@pukyong.ac.kr

\* Correspondence: soumayya.chakir@gmail.com (S.C.); kpjiwoo@pknu.ac.kr (Y.-B.K.); Tel.: +82-51-629-6197 (Y.-B.K.)

**Abstract:** This paper proposes a new method to control the pose of a camera mounted on a two-axis gimbal system for visual servoing applications. In these applications, the camera should be stable while its line-of-sight points at a target located within the camera's field of view. One of the most challenging aspects of these systems is the coupling in the gimbal kinematics as well as the imaging geometry. Such factors must be considered in the control system design process to achieve better control performances. The novelty of this study is that the couplings in both mechanism's kinematics and imaging geometry are decoupled simultaneously by a new technique, so popular control methods can be easily implemented, and good tracking performances are obtained. The proposed control configuration includes a calculation of the gimbal's desired motion taking into account the coupling influence, and a control law derived by the backstepping procedure. Simulation and experimental studies were conducted, and their results validate the efficiency of the proposed control system. Moreover, comparison studies are conducted between the proposed control scheme, the image-based pointing control, and the decoupled control. This proves the superiority of the proposed approach that requires fewer measurements and results in smoother transient responses.

**Keywords:** backstepping control; gimbal system; imaging geometry; line-of-sight; visual servoing



**Citation:** Huynh, T.; Tran, M.-T.; Lee, D.-H.; Chakir, S.; Kim, Y.-B. A Study on Vision-Based Backstepping Control for a Target Tracking System. *Actuators* **2021**, *10*, 105. <https://doi.org/10.3390/act10050105>

Academic Editor: Paolo Di Giamberardino

Received: 7 April 2021  
Accepted: 18 May 2021  
Published: 19 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Vision-based motion control refers to the use of computer vision as feedback to control the motion of a system. The topic is becoming more and more popular, thanks to the development of intelligent computer vision techniques. In general, the objective of vision-based control is to minimize the error between the actual image measurement of the visual features vector and its desired value. Numerous studies have reviewed several aspects of this topic, especially the two-part tutorial series by Chaumette and Hutchinson, [1,2]. Two control approaches are named depending on the value of the image measurement: image-based visual servo if the value is immediately available in the visual data, or position-based visual servo if the value contains a set of 3D parameters estimated from the visual data. Obviously, the visual data are acquired from a camera or a set of cameras, which is either fixed in the workspace or directly mounted on the controlled system. Applications of visual servoing are becoming popular in robotics, where the pose of a robot manipulator end-effector is estimated and controlled with 6 degrees-of-freedom (DoF), such as in [3–5]. In outdoor applications, thanks to their ability to perceive the surrounding environment, camera systems have been widely applied for controlling the motion wheeled robots [6,7] and attitudes of unmanned vehicles [8–10]. Gimbale mechanisms with multiple rotational axes are preferred to carry the camera if its pose needs controlling. Since a gimbal system can only control the angular motions of the camera (the number of controllable DoFs is equal to the number of independent gimbal channels), the control objectives are different

from classical robotics. Instead, the fundamental goal is to direct the camera's line-of-sight (LOS) to a target and keep it in its field of view (FOV). Thus, two control actions are required: detecting the target in optical imagery and controlling the gimbal for target tracking. Additionally, the camera-based visual data can be used to control behaviors of relative systems, for example, a stereo camera unit carried by a gimbal on a vessel that keeps track of the target and measures the distance for a berthing aid system [11]. Although many aspects of visual servoing and gimbal motion control have been studied individually for decades ([12–14]), the issue of vision-based control is seldom tackled.

To detect targets in optical imagery, an image tracker is required. In particular, the tracker detects the presence of a target in the camera's FOV and then tracks its projection in the image plane frame-by-frame. Thus, the tracker measures the error between the LOS orientation and target location. Image processing techniques and computer vision algorithms are implemented to fulfill these objectives. Traditional visual tracking approaches rely on deterministic feature search. For instance, the Kanade–Lucas–Tomasi tracker [15–17], that prefers corner features in an image patch, and the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm in [18,19], which is based on color distribution within a video. However, they have difficulty tracking a target when it appears temporarily or when the background is contaminated. To cope with natural image changes, modern tracking methods are trained online with sample patches updated at every new detection. Especially, discriminative tracking methods, such as the Tracking–Learning–Detection (TLD) [20] and the Kernelized Correlation Filter (KCF) [21], have enhanced their performance significantly. However, limitations of the reliability and robustness of the tracker heavily affect the efficiency of the vision-based control system. For instance, when the transient response is fast, the image may become blurred, the visual tracking may simply fail to measure the location of the target, and thus, resulting in missing data samples. From a control engineering standpoint, the tracker is a sampled device [13], therefore, missing data samples generates delays in the tracking loop control, and the system stability may be affected.

To achieve the desired performances, the motion controller should overcome not only the tracker's limitations but also the nonlinearities in gimbal dynamics and the imaging geometry. Besides, the gimbal's kinematics are complex due to the intertwined trigonometric functions [12]. The system dynamics are highly nonlinear, the nonlinearities are mainly induced by unbalanced masses, parameter uncertainties, and torque disturbances ([22–27]). Additionally, difficulties may also come from the imaging geometry couplings along with the gimbal system's lack of controllable directions and measurable variables [28,29]. Furthermore, there are non-ideal practical factors, such as delay time, measurement noise, and the camera shuttle speed.

Meanwhile, most of the available studies using visual data to control the gimbal are rather simple. For instance, decoupled approaches, where all the coupling terms are neglected, were used in [11,13,30]. In [31], desired orientations of the gimbal channels were calculated from their forward kinematics, and then a piece-wise linear controller was implemented. However, due to the nature of the underactuated gimbal system, these approaches are unable to achieve effective results. One of the significant studies in vision-based motion control is the one by Hurák et al. ([28,29]). The authors proposed an image-based pointing controller designed based on the image-based visual servo approach, in which the camera's angular rates were the control inputs. The designed control law took into account the couplings and limitations of the gimbal channels, but at the same time, it required a lot of measurements. Besides, the stability of the designed control system is questionable, which may result in unwanted responses. In one of the latest studies related to the topic, X. Liu et al. [32] treated unknown imaging geometry, angular rate errors, and uncalibrated camera parameters as lumped disturbances. The authors proposed a combination of a disturbance observer and a model predictive controller to reduce the number of required measurements; disturbance rejection and good tracking performances were obtained. It is worth noting that any estimation from an observer is always lagging

behind the true values for some time. Along with the delay in image streaming and computer vision processing, the observer may lead to performance deterioration and system instability.

Therefore, in this paper, a new vision-based tracking control scheme for a two-axis gimbal system was designed. A visual camera served as a payload for the gimbal and provides visual information for the control system. This type of system is mainly used in aerial surveillance. The control objective was to bring a target to the center of the image plane with zero steady-state error and smooth transient response so that effective visual tracking was achieved. To fulfill these tasks, the proposed control scheme was designed following these steps. First, the complete model of the system was derived, where the gimballed mechanism dynamics, imaging kinematics, and the actuation model were all taken into consideration. Secondly, the required rotation angles to track a predefined target were determined from a new perspective with the decoupling effects were fully isolated. Then, the proposed controller was designed based on the backstepping procedure ensuring the system stability and tracking performances. Finally, simulations and experiments were conducted for validation. The proposed controller requires fewer measurements, and it ought to perform better than the benchmark controller in Hurák's research. In short, the contributions of the paper are summarized as follows:

- A complete model is derived representing the target tracking system with the gimballed mechanism and image measurements.
- A new control scheme for visual servoing systems is proposed. The novelty of this approach is that the couplings in both the gimbal kinematics and imaging geometry are decoupled using a new technique, namely the calculation of additional orientation. Then, the vision-based target tracking system can be expressed with recursive structures of separate SISO systems. Therefore, conventional control schemes can be easily implemented.
- The stability of the closed-loop system is analyzed. Simulation and experimental results are presented and discussed; thus, the effectiveness of the proposed system is validated.

Accordingly, the remainder of the paper is organized as follows. Characteristics of the gimbal motion and the imaging geometry are analyzed and presented in Section 2. Then, to improve the control performance of gimbal systems, a new control scheme is introduced in Section 3. Section 4 details the implementation of the visual tracking algorithm KCF for the image tracker. Simulation and experimental comparison studies were conducted, and the results are presented in Section 4. Finally, conclusions are drawn in Section 5.

## 2. System Model

In this study, a two-axis gimbal system mounted on an aerial vehicle is considered. The camera's aperture was assumed to coincide with the rotational center of the inner channel. Let us assign the coordinate system of five frames for the system as follows. First, a coordinate frame ( $B$ ) was fixed to the platform carrying the gimbal with its  $Z$ -axis pointing upward and  $X$ -axis pointing forward. This frame was brought into the rotational center of the tilt channel by the three dashed orthogonal vectors, as depicted in Figure 1. Two more frames, namely ( $P$ ) and ( $T$ ), were associated with the pan and tilt channels, respectively.  $Y_p$ - and  $Y_t$ -axes of the  $P$  and  $T$  frames coincided, while the  $Z_p$ -axis of the  $P$  frame coincided with the  $Z$ -axis of frame  $B$ . The camera was carried by the tilt channel, and so the camera frame ( $C$ ) was fixed to frame  $T$ . However, its  $Z_c$ -axis was now denoting the camera's LOS, as illustrated by the coordinate  $OX_cY_cZ_c$  in Figure 1. Finally, the image coordinate frame was at an appropriate focal position along the optical axis from the origin of the camera frame. The planar orthogonal to the LOS is called the image plane, in which the  $x$ - and  $y$ -axes are the image coordinates, as shown in Figure 1.

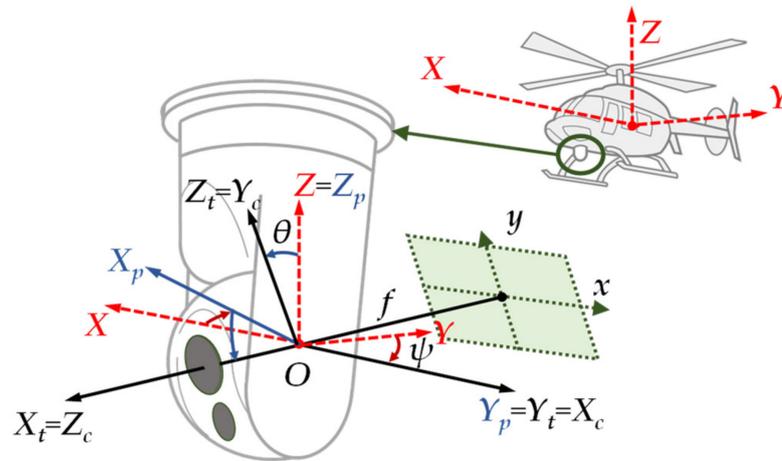


Figure 1. Two-axis gimbal with defined coordinates.

2.1. System Kinematics

The following kinematic relation of the gimbal system is the result of the transformation from frame *B* to frame *T*, and it is given by:

$$\begin{bmatrix} \omega_{tx} \\ \omega_{ty} \\ \omega_{tz} \end{bmatrix} = \begin{bmatrix} \cos \theta (\omega_{bx} \cos \psi + \omega_{by} \sin \psi) - \sin \theta (\omega_{bz} + \omega_{\psi}) \\ -\omega_{bx} \sin \psi + \omega_{by} \cos \psi + \omega_{\theta} \\ \sin \theta (\omega_{bx} \cos \psi + \omega_{by} \sin \psi) + \cos \theta (\omega_{bz} + \omega_{\psi}) \end{bmatrix} \tag{1}$$

where  $\omega_t = [\omega_{tx} \ \omega_{ty} \ \omega_{tz}]^T$  and  $\omega_b = [\omega_{bx} \ \omega_{by} \ \omega_{bz}]^T$  are the angular rate vectors of the tilt channel and the platform, respectively.  $\psi$  is the relative rotation between the outer gimbal and the platform about the *Z*-axis, whereas  $\theta$  is the relative motion between the inner and the outer gimbal about the *Y<sub>p</sub>*-axis, while  $\omega_{\psi}$  and  $\omega_{\theta}$  are their respective rates. The rotation between the camera frame *C* and the tilt channel frame *T* is given by the following rotation matrix:

$$R_C^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{2}$$

Pre-multiplying Equation (1) by  $R_C^T$  results in the transformation from the base to the camera frame. Similarly, by a translation equal to the lens focal length along the *Z<sub>c</sub>*-axis, the kinematics of the image frame is thus obtained. However, it is preferable to consider the relationship between the projected point on the image plane and the gimbal motion.

In the pinhole camera model, the camera aperture can be described as a point and no lenses were used to focus light. That is, a first-order approximation of the mapping from an object in the three-dimensional space to its projection on the image plane is obtained. If the location of a 3-D point in the camera frame is given by a vector  $P = [X_c \ Y_c \ Z_c]^T$ , and its projection on the image plane is  $p = [x \ y]^T$ , the pinhole camera model yields the following projective transformation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix} \tag{3}$$

with *f* is the focal distance. Time-derivatives of *x* and *y* result in the velocities of the projected point, which highly depends on the focal length, the motion of the observed target, and the spatial velocities of the camera. Assume that the focal length is constant, then the time variation of the projected point is related to the spatial motion of the camera

through an interaction matrix ([1,29]). Considering the fixed rotation between the camera frame  $C$  and the tilt channel frame  $T$  (Equation (2)), the following equation is derived:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = L v_t + e_d \quad (4)$$

where  $L$  is the interaction matrix between the camera motions and the projected point.  $v_t$  is the spatial velocity vector of the tilt channel, which includes the instantaneous linear velocity of the tilt frame origin  $v_t = [v_{tx} \ v_{ty} \ v_{tz}]^T$  and  $\omega_t$  the angular velocity.  $e_d$  is the term expressing the influence of the target motion on the velocity of the projected point. The matrix and vectors above mentioned are given by:

$$L = \begin{bmatrix} \frac{x}{Z} & \frac{f}{Z} & 0 & y & -\frac{xy}{f} & \frac{(f^2+x^2)}{f} \\ \frac{y}{Z} & 0 & \frac{f}{Z} & -x & -\frac{f^2+y^2}{f} & \frac{xy}{f} \end{bmatrix} \quad (5)$$

$$v_t = [v_t^T \ \omega_t^T]^T$$

The last three columns in the interaction matrix represent the relation between the gimbal orientation and the motion of the projected point. Equations (1) and (5) show the total kinematic coupling between the gimbal motion and the projection in the image plane. The first row of Equation (1) indicates a redundant roll motion generated by the two other gimbal orientations, and the third column in Equation (5) reveals the influence of this motion on the projection onto the image plane. The third row of Equation (1), especially, shows that the required motion of the outer gimbal increases in proportion with the secant of the angle  $\theta$ . Therefore, the gimbal mechanism cannot be globally asymptotically stabilized because when the angle  $\theta$  approaches 90 [deg] or  $-90$  [deg], the gimbal is completely out of control (this phenomenon is known as the gimbal lock, which has been thoroughly analyzed in [12,14]). Common solutions for this problem are feedforward approaches, additional channels, or safety mechanisms to prevent the tilt channel from reaching 90 [deg] or  $-90$  [deg]. The two-axis gimbal system used for this study is equipped with the latter.

## 2.2. System Dynamics and Actuation Model

When applying Euler's equation of motion for rigid body dynamics, the torque relationships of the inner and outer gimbals are derived as:

$$\begin{aligned} J_t \dot{\omega}_t + [\omega_t \times J_t \omega_t] &= T_t \\ J_p \dot{\omega}_p + [\omega_p \times J_p \omega_p + T_{t/p}] &= T_p \end{aligned} \quad (6)$$

The subscript  $t$  denotes the tilt channel, while subscript  $p$  is used to denote the pan channel.  $J$  is the inertia matrix,  $T$  is the applied torque, and  $T_{t/p}$  is the tilt gimbal's torque as observed from the coordinate frame of the pan gimbal. Equation (6) represents the dynamics of both channels expressed in their respective coordinate axes. As aforementioned, the pan channel rotates about the  $Z$ -axis, whereas the tilt rotates about the  $Y_p$ -axis. Thus, only two motion equations are considered:

$$\begin{aligned} J_{ty} \dot{\omega}_{ty} + (J_{tx} - J_{tz}) \omega_{tx} \omega_{tz} &= T_{ty} - K_t \omega_\theta \\ J_{pz} \dot{\omega}_{pz} + (J_{py} - J_{px}) \omega_{px} \omega_{py} + T_{t/pz} &= T_{pz} - K_p \omega_\psi \end{aligned} \quad (7)$$

The external torques acting on each channel consist of the driving torque ( $T_{ty}$  and  $T_{pz}$ ) and the friction torques. The latter is assumed to be proportional to the relative speed, and the coefficients of proportionality are  $K_t$  and  $K_p$ . In a typical gimbal system, the driving torques are generated by the stabilization controllers whose references are generated by the tracking loops ([12,13,20]). The actuators in this study, in particular, are servo systems: each one of them has an integrated controller responsible for speed control. The closed-loop

configuration of this structure is illustrated in Figure 2, and it is used as the stabilization loop for the gimbal control system. Then, the tracking loop model is derived taking into consideration both the system dynamics and the speed control loops.

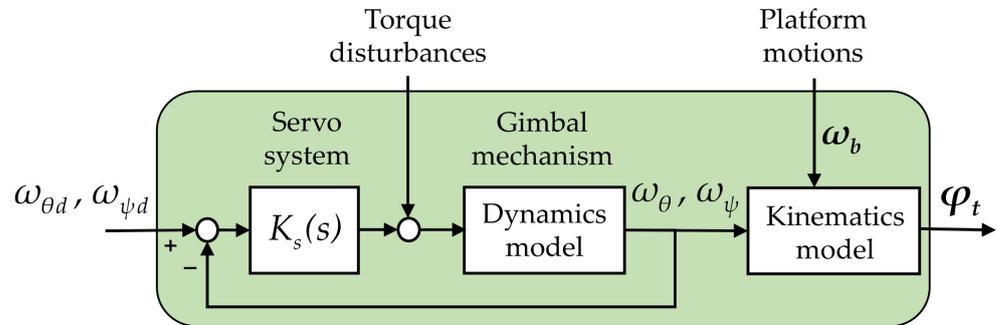


Figure 2. Block diagram of the gimbal with the stabilization controller.

Proportional controllers for speed control are used in the stabilization loops, thus the driving torques are:

$$\begin{aligned} T_{ty} &= P_t(\omega_{\theta d} - \omega_{\theta}) + d_t \\ T_{pz} &= P_p(\omega_{\psi d} - \omega_{\psi}) + d_p \end{aligned} \tag{8}$$

where  $P_t$  and  $P_p$  are the proportional control gains,  $\omega_{\theta d}$  and  $\omega_{\psi d}$  are the desired rate commands,  $d_t$  and  $d_p$  are the matched disturbances due to measurement errors, actuator saturation, etc.  $\omega_{\theta d}$  and  $\omega_{\psi d}$  are now the control signals that the tracking controllers compute.

From Equations (1), (4), (7), and (8), the system model is rearranged and given by:

$$\begin{aligned} \dot{\boldsymbol{\varphi}} &= \boldsymbol{\omega} \\ \begin{bmatrix} \dot{x} & \dot{y} \end{bmatrix}^T &= L_{\omega 1} \boldsymbol{\omega} + L_{\omega 2} \omega_{tx} + L_v \boldsymbol{v}_t + e_d \\ \dot{\boldsymbol{\omega}} &= \boldsymbol{B} \boldsymbol{u} - \boldsymbol{K} \boldsymbol{\omega} + \boldsymbol{d} \end{aligned} \tag{9}$$

where  $\boldsymbol{\varphi} = [\varphi_{ty} \ \varphi_{tz}]^T$  is the angular position of the tilt and pan channels, and  $\boldsymbol{\omega} = [\omega_{ty} \ \omega_{tz}]^T$  is the corresponding angular rates vector.  $L_{\omega 1}$ ,  $L_{\omega 2}$  and  $L_v$  are elements of the interaction matrix  $L$  compatible with  $\boldsymbol{\omega}$ ,  $\omega_{tx}$  and  $\boldsymbol{v}_y$ , respectively.  $\boldsymbol{u} = [\omega_{\theta d} \ \omega_{\psi d}]^T$  is the rate commands vector,  $\boldsymbol{B} = \text{diag}\{B_1, B_2\}$  and  $\boldsymbol{K} = \text{diag}\{K_1, K_2\}$  are system parameter matrices, and  $\boldsymbol{d} = [d_1 \ d_2]^T$  is the vector of unpredictable disturbances. The  $L$  matrix components and the system parameter matrices are respectively expressed as:

$$L_{\omega 1} = \begin{bmatrix} -\frac{xy}{f} & \frac{(f^2+x^2)}{f} \\ -\frac{f^2+y^2}{f} & \frac{xy}{f} \end{bmatrix}, L_{\omega 2} = \begin{bmatrix} y \\ -x \end{bmatrix}, L_v = \begin{bmatrix} \frac{x}{Z} & \frac{f}{Z} & 0 \\ \frac{y}{Z} & 0 & \frac{f}{Z} \end{bmatrix} \tag{10}$$

$$K_1 = \frac{K_t + P_t}{J_{ty}}, K_2 = \frac{K_p}{J_{pz} + J_{tz} \cos^2 \theta + J_{tx} \sin^2 \theta}, B_1 = \frac{P_t}{J_{ty}}, B_2 = \frac{P_p \cos \theta}{J_{pz} + J_{tz} \cos^2 \theta + J_{tx} \sin^2 \theta} \tag{11}$$

and the elements of the disturbances vector are:

$$\begin{aligned} d_1 &= \frac{1}{J_{ty}} (d_t - (J_{tx} - J_{tz}) \omega_{tx} \omega_{tz} + (K_t + P_t) \omega_{py}) \\ d_2 &= \frac{1}{J_{pz} + J_{tz} \cos^2 \theta + J_{tx} \sin^2 \theta} \{ (K_p + \cos \theta P_p) (\omega_{px} \sin \theta + \omega_{bz} \cos \theta) - J_{pz} (-\dot{\omega}_{px} \sin \theta - \omega_{px} \dot{\theta} \cos \theta + \omega_{pz} \dot{\theta} \sin \theta) \\ &+ [J_{tx} (\dot{\omega}_{px} \cos \theta - \omega_{px} \dot{\theta} \sin \theta - \tan \theta (-\dot{\omega}_{px} \sin \theta - \omega_{px} \dot{\theta} \cos \theta)) - \frac{\dot{\theta}}{\cos^2 \theta} (\omega_{tz} - \omega_{px} \sin \theta)] + (J_{tz} - J_{ty}) \omega_{ty} \omega_{tz} \\ &\sin \theta \cos \theta - \cos \theta (J_{py} - J_{px}) \omega_{px} \omega_{py} - [(J_{ty} - J_{tx}) \omega_{tx} \omega_{ty}] \cos^2 \theta + d_p - \cos \theta P_p \omega_{tz} \} \end{aligned} \tag{12}$$

$B$  is nonsingular within the operating range of the system, thus invertible. Neglecting the disturbances term, the proportional gains of the integrated controllers should be chosen such that  $K = B$ . Thus, the actuation system—consisting of the gimbaled mechanism and the two integrated actuators—works as a low pass filter ensuring the actual speeds are the filtered version of the command rates.

### 3. Vision-Based Tracking Control System

In general, the mapping from the 3D to 2D coordinates using the pinhole camera model allows the determination of the target location in the camera coordinates, as in Equation (3). Then, using the inverse tangent function, the desired rotations of the gimbal camera are derived. Nevertheless, Equations (1) and (5) show the unwanted effects of the motion of one actuator on the other orientations. In this section, the coupling effect in visual servoing is analyzed from a new viewpoint, where the required rotation of the gimbal camera is fully determined. Then, the proposed controller is designed based on the backstepping procedure. Due to many practical limitations, the actuation is not an ideal low pass filter. The backstepping technique is well known to be effective in dealing with uncertainties, hence ensures system performance and stability. On the other hand, an image-based pointing and tracking control scheme ([28,29]), and a decoupled approach [13] are implemented for a comparison study.

#### 3.1. Vision-Based Backstepping Control

##### 3.1.1. Calculation of Additional Orientation

Let  $\varphi_a = [\varphi_{tya} \quad \varphi_{tza}]^T$  be the additional angles about  $Y_t$ - and  $Z_t$ -axes that the inner gimbal needs to point at the predefined target. This value is known as the difference between the angular positions of the target and the camera's LOS. The computer vision tracker defines  $\varphi_a$  from the location of the projected point in the image plane. Thus, when the LOS is pointing at the target, its projection is at the center of the image plane, that is  $\varphi_a = [0 \quad 0]^T$ . Otherwise, the tilt and/or pan motions of the gimbal are needed.

In Figure 3, point A expressed by the vector  $[x_A \quad y_A]^T$  is the projection of the target onto the horizontal axis of the image plane. From the pinhole camera model, the image plane is tangent to a spherical surface whose radius equals the focal length of the camera. Moreover, the image plane's center point is the intersection point of the LOS and the spherical surface. Besides, the relative angle between the tilt and the pan gimbals along the  $Y_p$ -axis is  $\theta$ . This makes the rotation planes of the pan gimbal and the camera different. Thus, the additional rotation about the camera's  $Y_c$ -axis brings the center of the image plane to point A. On the other hand, the pan motion makes the image plane's center point below point A, hence a tilt motion is required even though  $y_A = 0$ .

As shown in Figure 3, the additional orientation of the camera's plane is easily obtained with:

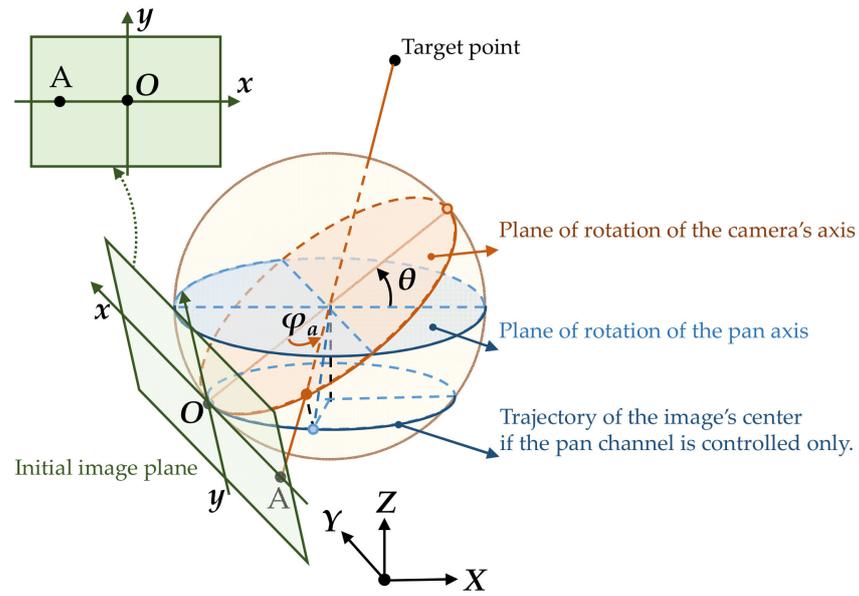
$$\tan \varphi_a = \frac{X_A}{Z_A} = -\frac{x_A}{f} \quad (13)$$

From the coordinate allocation, this orientation is equivalent to the rotation about the  $Z_t$ -axis of the tilt channel frame.

$$\tan \varphi_{tza} = \tan \varphi_a \quad (14)$$

while the additional rotation about the  $Y_t$ -axis of the tilt channel is required to be as the following:

$$\varphi_{tya} = \theta - \arcsin(\cos \varphi_a \sin \theta) \quad (15)$$



**Figure 3.** Relationship between gimbal orientations and location of the target’s projection on the image plane.

The third subscript  $a$  of the left terms in Equations (14) and (15) indicates the additional rotations, while the first and second subscripts denote the channels as well as the axis in which the values are measured. In general, when the target is expressed by its position  $\mathbf{p} = [x \ y]^T$  in the image plane, the rotation angles of both channels are computed as:

$$\begin{aligned} \varphi_{tza} &= -\arctan\left(\frac{x}{y}\right) \\ \varphi_{tya} &= \theta - \arctan\left(\frac{y}{f}\right) - \arcsin(\cos \varphi_a \sin \theta) \end{aligned} \tag{16}$$

### 3.1.2. Backstepping Controller Design

For a target located within the camera’s FOV and the gimbal’s operating range, the control objective is to bring its projection to the center of the image frame, which is equivalent to  $[\varphi_{tya} \ \varphi_{tza}]^T \rightarrow [0 \ 0]^T$ . Note that  $\varphi_a = \varphi - \varphi_d$  is the difference between  $\varphi$  and  $\varphi_d$  which are the angular position of the tilt channel and the target position that the LOS should point at, respectively. Interestingly, the value of  $\varphi_d$  is not only dependent on the location of the target, but also the translational motion of the camera. However, due to the characteristics of outdoor visual servoing applications, it is reasonable to assume that  $\varphi_d$  is a slow time-varying value, i.e.,  $\dot{\varphi}_d \approx 0$ . Besides, the additional angular rate is defined as follows:

$$\omega_a = -\Lambda \varphi_a - \omega \tag{17}$$

Accordingly, the following control law is proposed:

$$[\omega_{\theta d} \ \omega_{\psi d}]^T = \mathbf{B}^{-1} \left[ \mathbf{K} \omega + (\mathbf{\Gamma} + \mathbf{\Lambda}) \omega_a - (\mathbf{I} - \mathbf{\Lambda}^2) \varphi_a \right] \tag{18}$$

$\mathbf{\Lambda}$  and  $\mathbf{\Gamma}$  are positive definite diagonal matrices. Equation (18) gives the rate command vector to send to the gimbal system. Note that the inverse of matrix  $\mathbf{B}$  contains a secant gain correction,  $1/\cos \theta$ . The proposed control scheme is illustrated in Figure 4.

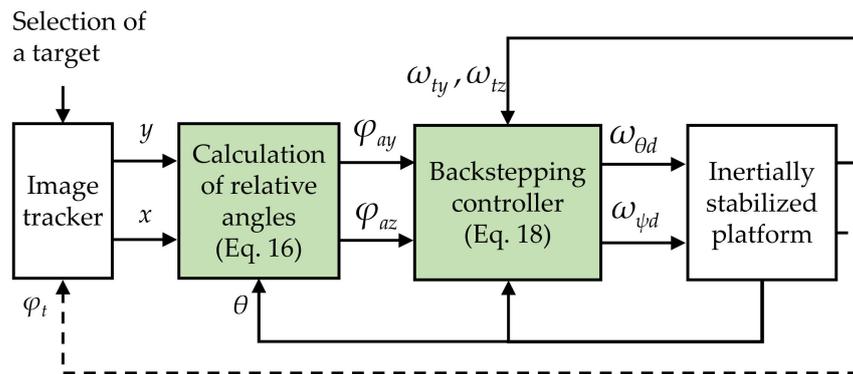


Figure 4. Schematic of the proposed control.

To analyze the system stability of the proposed control scheme, let us consider a positive definite function and its time-derivative as follows:

$$\begin{aligned} V &= \boldsymbol{\varphi}_a^T \boldsymbol{\varphi}_a + \boldsymbol{\omega}_a^T \boldsymbol{\omega}_a \\ \dot{V} &= \boldsymbol{\varphi}_a^T \dot{\boldsymbol{\varphi}}_a + \boldsymbol{\omega}_a^T \dot{\boldsymbol{\omega}}_a \end{aligned} \tag{19}$$

where

$$\begin{aligned} \dot{\boldsymbol{\varphi}}_a &= \dot{\boldsymbol{\varphi}} - \dot{\boldsymbol{\varphi}}_d = -\boldsymbol{\Lambda} \boldsymbol{\varphi}_a - \boldsymbol{\omega}_a \\ \dot{\boldsymbol{\omega}}_a &= \boldsymbol{\Lambda} (\boldsymbol{\Lambda} \boldsymbol{\varphi}_a + \boldsymbol{\omega}_a) + \mathbf{K} \boldsymbol{\omega} - \mathbf{B} \mathbf{u} - \mathbf{d} \end{aligned} \tag{20}$$

By substituting Equation (20) and the control law in Equation (18) into the expression of  $\dot{V}$ , we obtain:

$$\begin{aligned} \dot{V} &= \boldsymbol{\varphi}_a^T (-\boldsymbol{\Lambda} \boldsymbol{\varphi}_a - \dot{\boldsymbol{\varphi}}_d) + \boldsymbol{\omega}_a^T [\boldsymbol{\Lambda} \dot{\boldsymbol{\varphi}}_d - \boldsymbol{\Gamma} \boldsymbol{\omega}_a - \mathbf{d}] \\ &= -\boldsymbol{\varphi}_a^T \boldsymbol{\Lambda} \boldsymbol{\varphi}_a - \boldsymbol{\omega}_a^T \boldsymbol{\Gamma} \boldsymbol{\omega}_a + \boldsymbol{\omega}_a^T \mathbf{d} \end{aligned} \tag{21}$$

According to Young’s inequality for products:

$$\boldsymbol{\omega}_a^T \mathbf{d} \leq \frac{1}{2\delta^2} \boldsymbol{\omega}_a^T \boldsymbol{\omega}_a + \frac{\delta^2}{2} \mathbf{d}^T \mathbf{d} \tag{22}$$

Applying the inequality (22) to Equation (21) results in:

$$\dot{V} \leq -\boldsymbol{\varphi}_a^T \boldsymbol{\Lambda} \boldsymbol{\varphi}_a - \boldsymbol{\omega}_a^T \left( \boldsymbol{\Gamma} - \frac{1}{2\delta^2} \mathbf{I} \right) \boldsymbol{\omega}_a + \frac{\delta^2}{2} \mathbf{d}^T \mathbf{d} \tag{23}$$

$\delta$  is an arbitrary value, which will be chosen such that  $\left( \boldsymbol{\Gamma} - \frac{1}{2\delta^2} \mathbf{I} \right)$  is positive definite (where  $\mathbf{I}$  is the identity matrix). Hence, the system is proved to be input-to-state stable (ISS) with  $V$  an ISS-Lyapunov function for the system, [33]. That is, the system is asymptotically stable without the disturbances, and the control errors are bounded if the disturbances are bounded.

### 3.2. Image-Based Pointing Control

The image-based pointing and tracking control implemented in this section has been previously discussed in the work of Hurák et al., [28,29]. The objective of the pointing-tracking controller is to set the reference rates for the motion system. A classical image-based motion control uses the inverse of the interaction matrix to ensure a decoupled decrease of the error between the current value of the visual feature and its desired value.

Since only the rotations are controllable in the gimbal system, Equation (4) can be rewritten as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \mathbf{L}_v \mathbf{v}_t + \mathbf{L}_\omega \boldsymbol{\omega}_t + \mathbf{e}_d \tag{24}$$

where  $L_v$  contains the first three columns and  $L_\omega$  the last three columns of the interaction matrix  $L$  in Equation (5). Thus, the reference rates of the tilt channel are computed as:

$$\omega_{td} = L_\omega^* \left( -\alpha \begin{bmatrix} x \\ y \end{bmatrix} - L_v v_t \right) + k \begin{bmatrix} f & x & y \end{bmatrix}^T \quad (25)$$

The first term guarantees the asymptotically stable projection, where  $\alpha$  is the exponential rate and  $L_\omega^*$  is the right pseudoinverse of  $L_\omega$ . However, since the two-axis gimbal can only control the pitch and yaw motions, the parameter  $k$  is chosen such that the reference rate for the roll motion is equal to the measured velocity  $\omega_{tx}$ . From Hurák et al. ([14,15]),  $k$  is given by:

$$k = \frac{1}{f^2 + x^2 + y^2} (xv_{tz} - yv_{ty}) + \frac{\omega_{tx}}{f} \quad (26)$$

Then, from Equation (25), the reference rates for the inner gimbal orientations are given as:

$$\begin{bmatrix} \omega_{txd} \\ \omega_{tyd} \\ \omega_{tzd} \end{bmatrix} = \begin{bmatrix} \frac{\alpha y f}{f^2 + x^2 + y^2} + \frac{x \omega_{tx}}{f} - \frac{\omega_{tx}}{f} \frac{y f v_{tx} + f^2 v_{tz} - (x v_{tz} - y v_{ty}) x}{f^2 + x^2 + y^2} \\ -\frac{\alpha x f}{f^2 + x^2 + y^2} + \frac{y \omega_{tx}}{f} + \frac{x f v_{tx} + f^2 v_{ty} + (x v_{tz} - y v_{ty}) y}{f^2 + x^2 + y^2} \end{bmatrix} \quad (27)$$

From the system kinematics in Equation (1), the rate commands for the two gimbal channels are:

$$\begin{bmatrix} \omega_{\theta d} \\ \omega_{\psi d} \end{bmatrix} = \begin{bmatrix} \omega_{tyd} \\ \frac{1}{\cos \theta} \omega_{tzd} \end{bmatrix} \quad (28)$$

Generally, the distance to the target and the gimbal translations and rotations are required to generate the control law. In the case of no translational motions, the third terms of the reference vector elements in Equation (27) are equal to zero. Thus, the control law depends only on the rotations of the gimbal. However, the stability of the system remains questionable since disturbances are neglected and  $k$  is chosen as a function of both the camera motion and the projection's location.

Additionally, the simplistic decoupled controller studied in [13] is implemented for comparison. This control scheme uses a single-input-single-output loop for every channel of the gimbal. Each channel receives the corresponding vertical or horizontal locations of the target in the image plane. That is:

$$\omega_{\theta d} = \alpha y \text{ and } \omega_{\psi d} = \frac{1}{\cos \theta} \alpha x \quad (29)$$

## 4. Implementation, Simulations, and Experiments

### 4.1. Implementation

In this study, the image tracker uses the KCF algorithm. This tracker initializes with a target patch cropped from the target location in the initial frame and gives back its location in each frame of the sequence. The details of the KCF algorithm are presented in [21]. In this paper, a combination of the KCF tracker and a Gaussian kernel working on the histogram of oriented gradients (HOG) was implemented.

In the following simulations and experiments, the sampling time of the control systems and the image tracker was set at 0.05 [s]. From the camera to the control system, the image sequences were streamed using the real-time streaming protocol (RTSP) with a resolution of  $640 \times 360$  (pixel) and a framerate of 20 (fps). The image streaming was significantly influenced by the communication delays—the tracker received an image that was a delayed version of the one taken by the camera. With our experimental apparatus, the average overall delay time was 0.3 [s], and in the worst case, the delay was 0.45 [s]. Besides, motion

sensors and gimbal actuators were also connected to the controller via an RS-232 serial port. However, the resulting delay time was relatively small compared to the image streaming delay, so it was neglected. The detailed control parameters are shown in Table 1.

**Table 1.** Control system parameters.

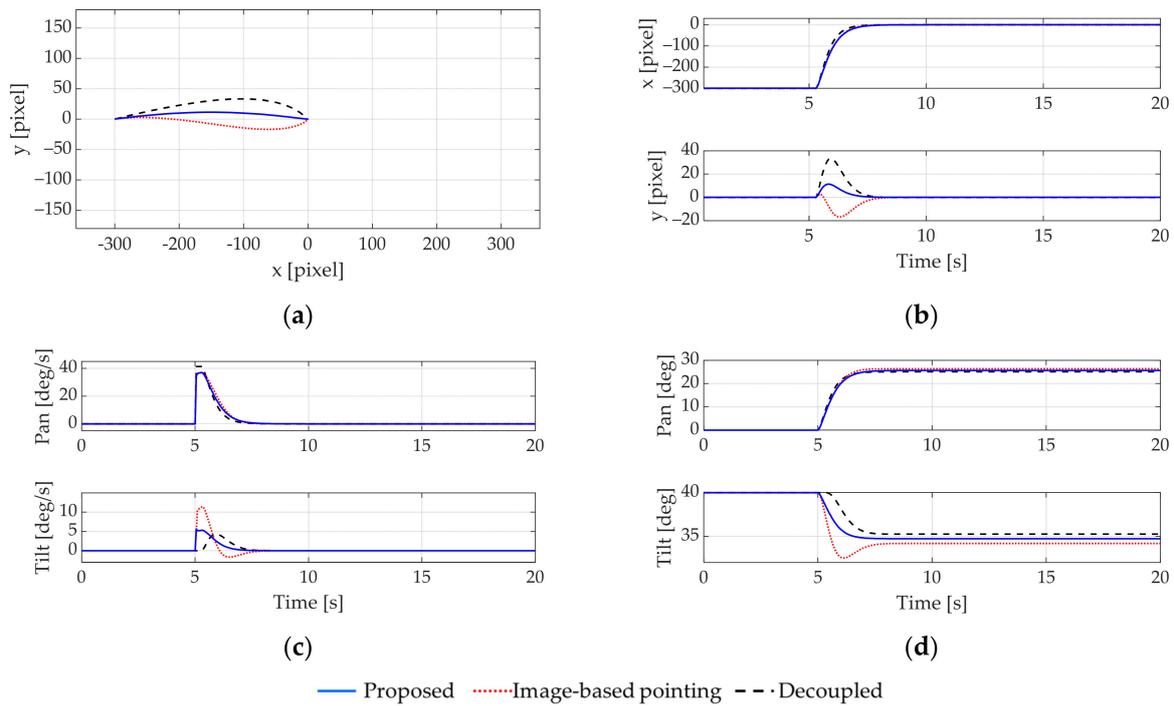
	Proposed (Equation (18))	Image-Based Pointing (Equations (27) and (28))	Decoupled (Equation (29))
Parameters	Controller's gains $\Lambda = \begin{bmatrix} 25 & 0 \\ 0 & 25 \end{bmatrix},$ $\Gamma = \begin{bmatrix} 1.1 & 0 \\ 0 & 1.1 \end{bmatrix}$ System model parameters $K = \begin{bmatrix} 21.9680 & 0 \\ 0 & 30.3969 \end{bmatrix},$ $B = \begin{bmatrix} 22.2485 & 0 \\ 0 & 28.4352 \cos \theta \end{bmatrix}$	Controller's gain $\alpha = 65$	Controller's gain $\alpha = 65$
Sampling time		0.05 [s]	
Video streaming delay time		Average 0.3 [s], worst-case 0.45 [s]	

A fast response was not the priority when tuning parameters due to the limitations of the camera's shutter speed. All the controllers were tuned such that from the zero orientation, the controllers could bring a projected point from an arbitrary location to the center of the image plane with a linear tracking path and reasonable rotational speeds.

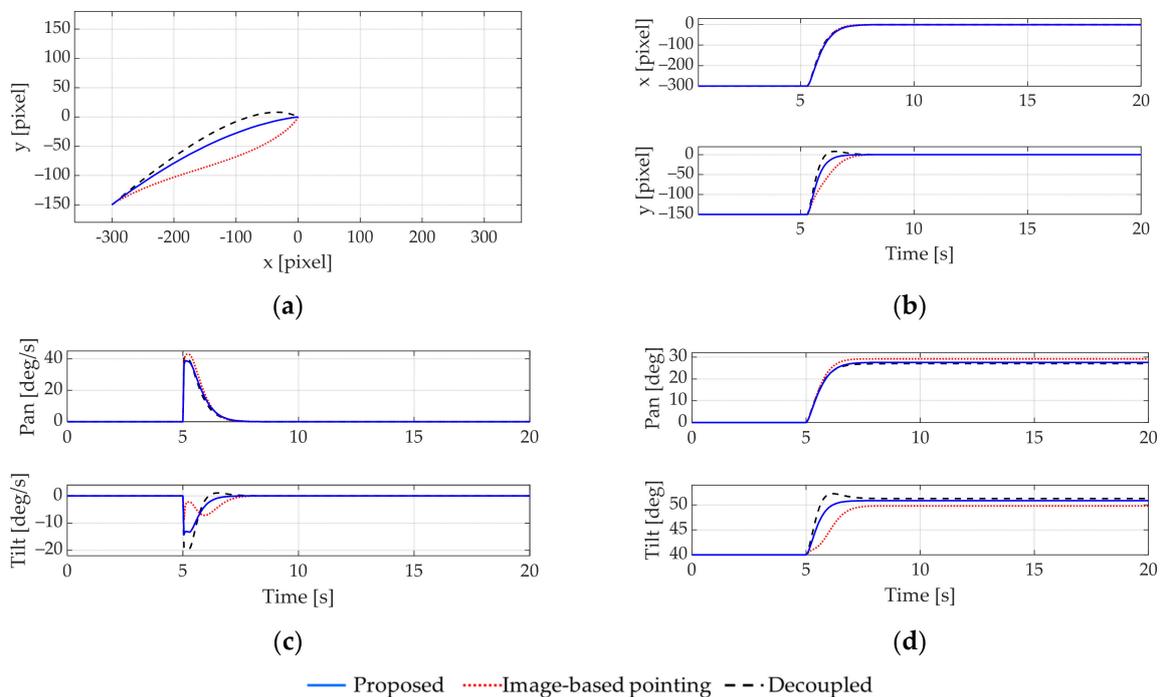
#### 4.2. Simulation Studies

Control systems were simulated with a target located out of the center of the image plane. A comparison between the performance of the three control schemes was carried out. Moreover, to highlight the coupling effects of the gimbal mechanism, the tilt angle was initially set at 40 [deg]. In the first simulation, the target projection is on the horizontal line of the image plane, while the target is located at the bottom-left corner in the second simulation test. Ideally, the controllers should follow a linear path in the image plane since it is the shortest way, and the response in both directions should converge smoothly to zero at the same time with no overshoot. The following figures, Figures 5 and 6, show these simulation results. In each figure, the responses with the proposed controller (Equation (18)), the image-based pointing controller (Equations (27) and (28)), and the decoupled control system (Equation (29)) are plotted and compared. Although the unit of  $x$  and  $y$  is (m), in the following figures they are displayed in [pixel], which corresponds to the displayed images in the user interface. Converting from one unit to the other is simply done by the multiplication of pixel dimensions. All the control actions take place from the 5th second of the simulation time.

Both simulations show that all the controllers struggle to follow a linear path in the image plane. The main reason is the association of time delay and kinematic couplings acting as disturbances on the system. In the first simulation, the initial control signal from the decoupled control scheme is 0, as shown in Figure 5c. Since the decoupled controller cannot anticipate the influence of the coupling at a nonzero tilt angle, it leads to a high peak of the  $y$ -coordinate of the projected point, which results in a bold curve in the image plane. In contrast, nonzero control signals are generated by the proposed controller and the image-based pointing controller; however, a great overshoot can be seen in the response of the latter (Figure 5b,d). The image-based pointing controller is sensitive to disturbances and uncertainties, while the backstepping technique is well known to be robust against them. Thus, the proposed controller performs well with an almost linear tracking path and smooth transient gimbal motions.



**Figure 5.** Simulation 1: Tracking a target point on the horizontal line. (a) Tracking path on the image plane; (b) time response of each coordinate; (c) control signals; (d) orientations of the inner gimbal.

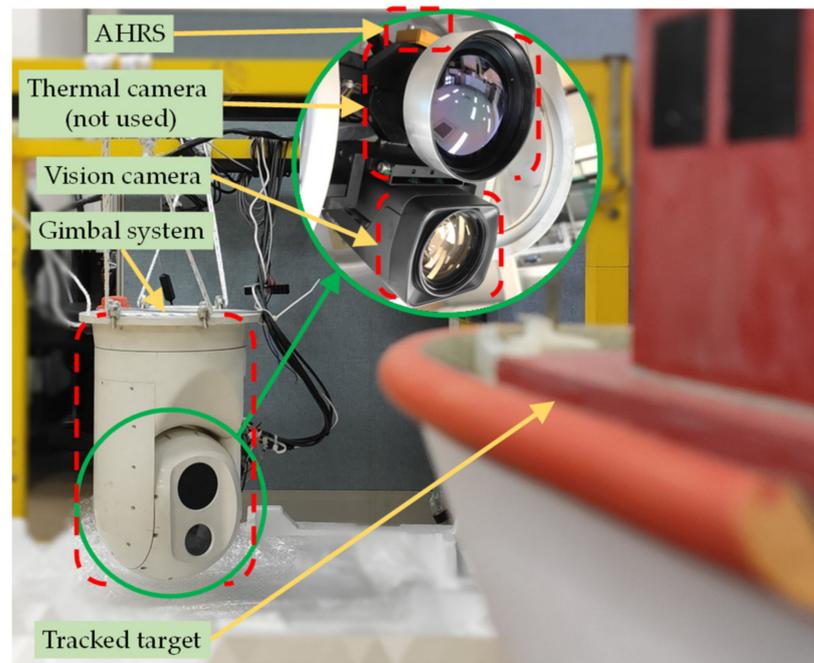


**Figure 6.** Simulation 2: Tracking a target point on the bottom-left corner of the image plane. (a) Tracking path on the image plane; (b) time response of each coordinate; (c) control signals; (d) orientations of the inner gimbal.

The results of the second simulation provide similar findings, that is the proposed controller performed the best among the three control schemes. For instance, an overshoot is recorded in the response of the decoupled controller, and non-smooth motions are obtained with the image-based pointing controller. Meanwhile, good step responses in both directions of the image plane are achieved with the proposed controller.

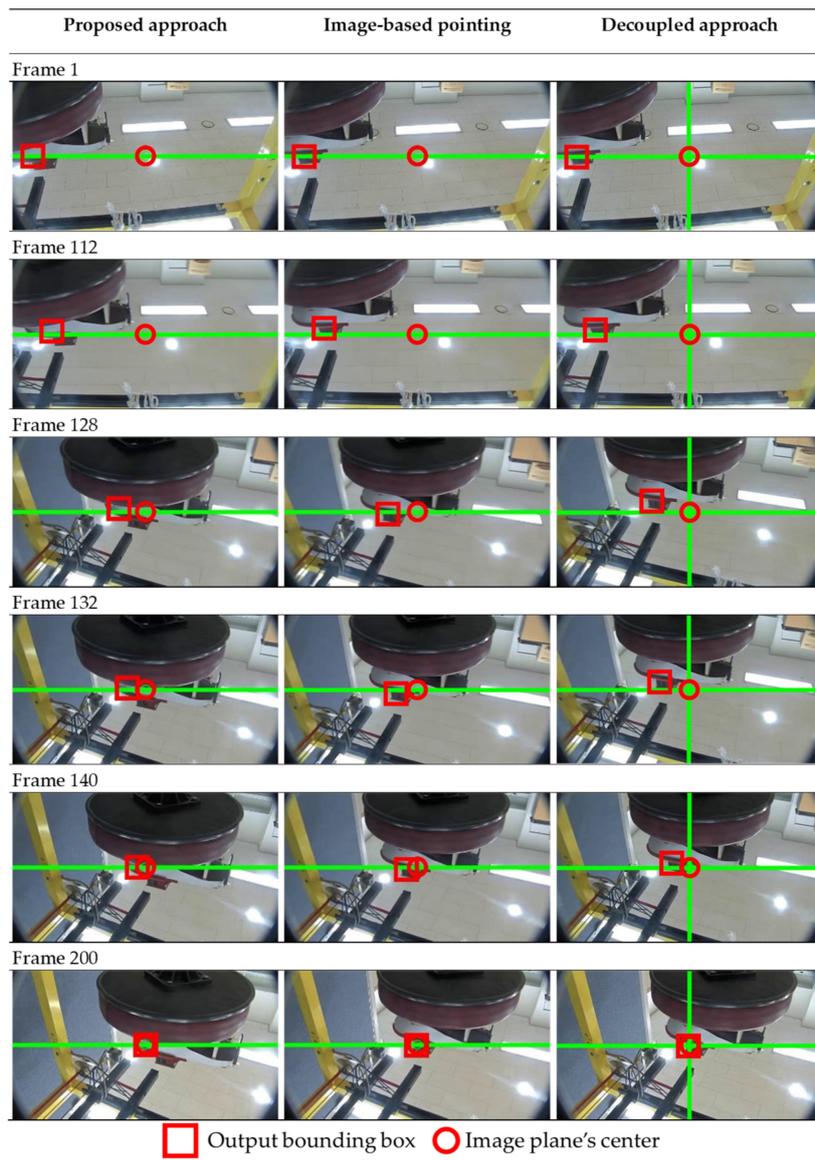
### 4.3. Experiments

The experimental studies were conducted with the 2-axis gimbal system illustrated in Figure 7. In this section, the experimental results of several scenarios are presented. Three studies were conducted in total. The first two were similar to the above-mentioned simulations, and the third scenario was a vision-based tracking in the presence of disturbances. The following figures, Figures 8–10, illustrate their results. The system specifications and the control parameters were addressed in the previous section.

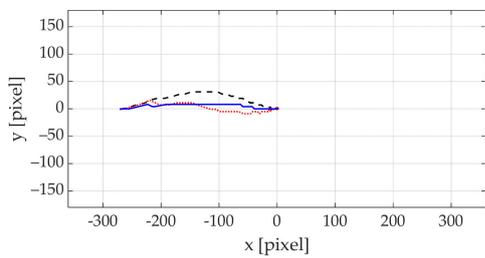


**Figure 7.** The target tracking apparatus.

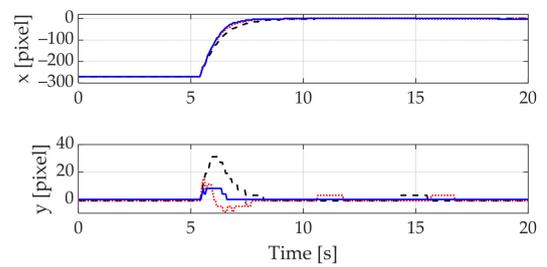
The frame-by-frame visualization of the three controllers' tracking performances is shown in Figure 8a. The horizontal green line indicates the horizontal axis, which cuts through the center of the image (the red circle). The position of the tracked target is represented by its bounding box (the red square) on the image frame. Although all the controllers could bring the target's projection to the center of the image (frame 200), the orientation of the pan actuator created a curvilinear path in the image plane resulting in an upward motion at the beginning of the control action (frame 112). The decoupled control performed worst since the LOS projection continued to deviate further away from the horizontal axis. Both tracking paths resulting from the two other controllers were close to the horizontal line, and the performance of the proposed controller remained better. Meanwhile, with the image-based pointing controller, the location of the projection went above and below the horizontal line indicating the overreaction of the control system. The projection's paths on the image frame as shown in Figure 8b and the time responses along the  $Y_t$ -axis direction in Figure 8c prove this remark. Moreover, the final angular positions of the inner gimbal were not the same even though all the image errors were converging (Figure 8e). The two vector elements in Equations (4) and (5) were determined by three angular rates of the camera's motion (without the translational motion). Since only two rotations were controllable, there were infinite combinations to bring a projected point to the center of the image plane.



(a)

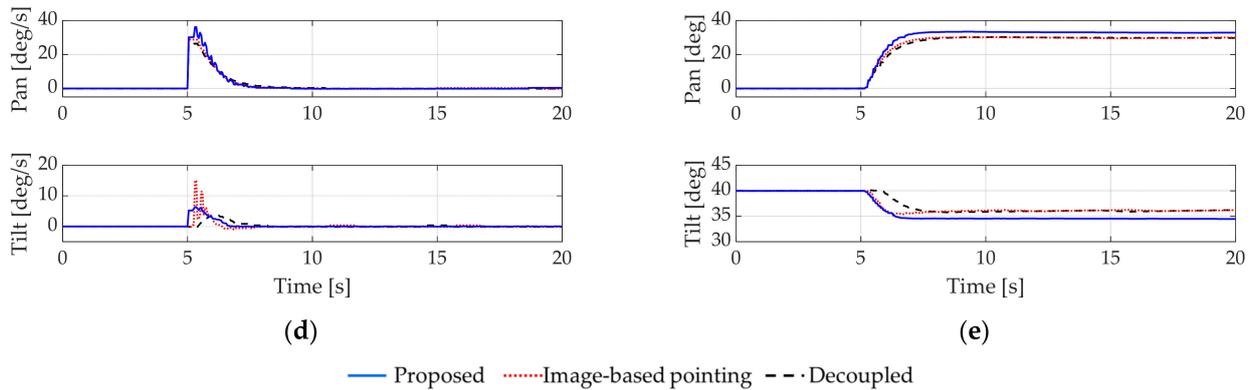


(b)

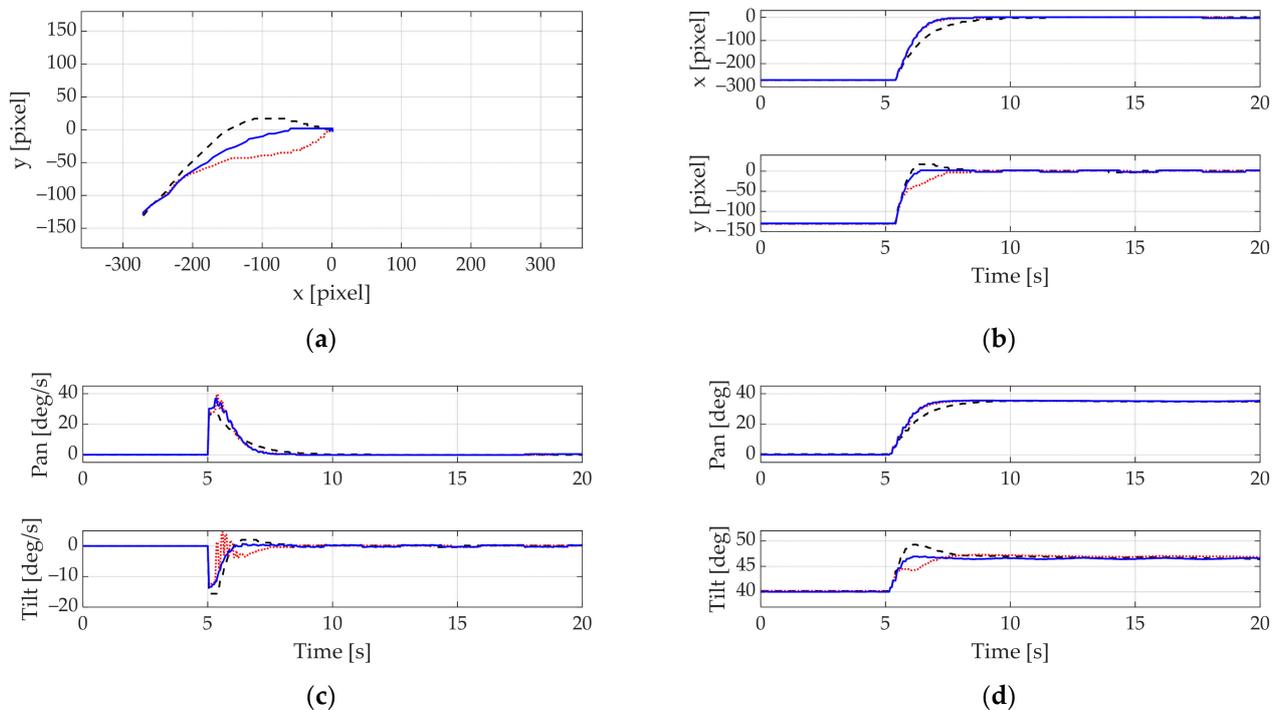


(c)

Figure 8. Cont.

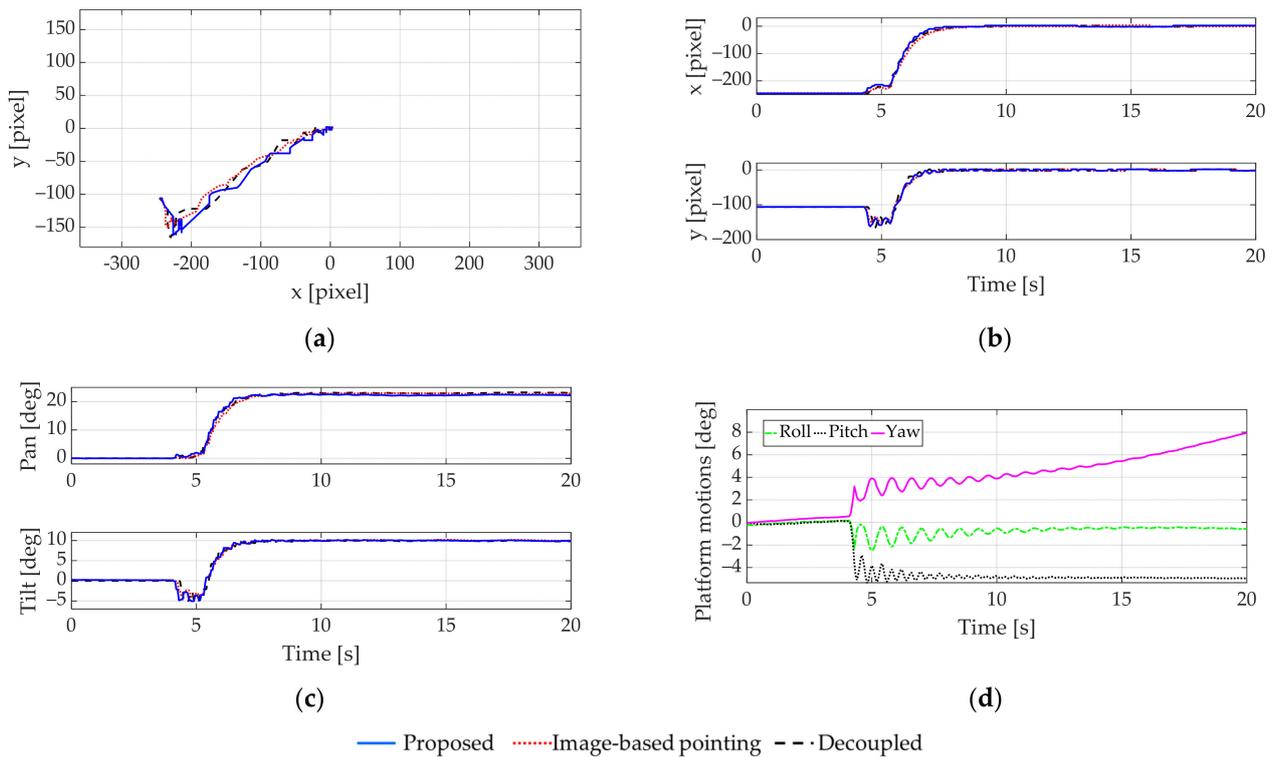


**Figure 8.** Experiment 1: Tracking a target initially located on the horizontal line. (a) Frame-by-frame tracking performance of the proposed control and the image-based pointing control; (b) tracking path on the image plane; (c) time response of each coordinate; (d) control signals; (e) orientations of the inner gimbal.



**Figure 9.** Experiment 2: Tracking a target point on the bottom-left corner of the image plane. (a) Tracking path on the image plane; (b) time response of each coordinate; (c) control signals; (d) orientations of the inner gimbal.

Furthermore, Figure 9 shows that in the second experiment the proposed controller provided the smoothest response among the three control schemes. Although its tracking path was curvilinear (Figure 9a), the transient responses were good in both directions (Figure 9b). Slow response in the horizontal direction and an overshoot in the vertical direction were seen in the performance of the decoupled controller. Moreover, the image-based pointing controller generated bumpy actions in the  $Y_t$ -axis direction despite the good horizontal response.



**Figure 10.** Experiment 3: Vision-based tracking with the presence of disturbance. (a) Tracking path on the image plane; (b) time response of each coordinate; (c) orientations of the inner gimbal; (d) disturbances from motions of the platform.

On the other hand, delay systems usually have bad reputations regarding system stability and performance. The delay puts a fundamental limitation on the achievable bandwidth of the system, which then recommended not to exceed  $\pi/2T$  or  $2/T$ , with  $T$  is the delay time ([34,35]). In the third experiment, the system needed to track a target initially situated at a corner of the image plan from zero initial tilt orientation. Disturbances in experiment 3 were induced by a manual rotation of the gimbal platform. The disturbances were a combination of high frequency damped sinusoidal oscillations and a continuously increasing rotation. Experimental results in Figure 10 show that all controllers could effectively attenuate the latter, but not the former due to limited bandwidths. In practical scenarios when the gimbal system needs to track a moving target, the unpredicted target motion will result in an additional term in the time derivative of its projection (Equation (4)). This can be considered as a disturbance to the system, for which the proposed controller can still be effective if it is within the controllable bandwidth. Thus, the LOS follows the motion of the target estimated by the image tracker.

## 5. Conclusions

In this paper, a novel control scheme was designed and implemented for visual servoing with a two-axis gimbal mechanism. The design procedure was presented systematically, with the couplings in the gimbal kinematics as well as the imaging geometry were addressed as the main difficulties. To overcome these difficulties, the proposed controller computes the required motions from the visual data and then generates the control signals using the backstepping technique. In simulation and experimental studies, the proposed controller was challenged by the non-ideal specifications of the system through different tracking scenarios. Hence, its effectiveness was validated, and its superiority was proved in comparison with other control schemes. On the other hand, practical limitations of the experimental apparatus were also highlighted.

Visual servoing is still a wide-open research field with the purpose of enhancing the reliability of visual tracking algorithms and system robustness. Moreover, deep learning

and 3D imaging are bringing new perspectives and rapid changes to computer vision. Meanwhile, vision-based motion control systems will still be affected by moving targets, disturbances, and delays. For gimbal-based tracking systems, in particular, the effects of the target motion, as well as the disturbances, are complicated because they are highly dependent on the area of application. Therefore, predicting the target motion, attenuating the disturbances, and diminishing the time delay influence are control objectives that require further studies. In future works, the enhancement of the visual servoing controller is necessary to keep up with the developing visual tracking algorithms, whereas the system performance and robustness while tracking moving targets and facing external disturbances should be improved.

**Author Contributions:** Conceptualization, Y.-B.K., and T.H.; methodology, T.H.; software, M.-T.T., S.C., and D.-H.L.; validation, T.H., S.C., and Y.-B.K.; formal analysis, T.H.; writing—original draft preparation, T.H.; writing—review and editing, T.H., M.-T.T., D.-H.L., and S.C.; supervision, Y.-B.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** Following are results of a study on the “Leaders in Industry-university Cooperation +” Project, supported by the Ministry of Education and National Research Foundation of Korea.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing is not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chaumette, F.; Hutchinson, S. Visual servo control. I. Basic approaches. *IEEE Robot. Autom. Mag.* **2006**, *13*, 82–90. [\[CrossRef\]](#)
2. Chaumette, F.; Hutchinson, S. Visual servo control. II. Advanced approaches [Tutorial]. *IEEE Robot. Autom. Mag.* **2007**, *14*, 109–118. [\[CrossRef\]](#)
3. Yun-Hui, L.; Hesheng, W.; Chengyou, W.; Lam, K.K. Uncalibrated visual servoing of robots using a depth-independent interaction matrix. *IEEE Trans. Robot.* **2006**, *22*, 804–817. [\[CrossRef\]](#)
4. Muis, A.; Ohnishi, K. Eye-to-hand approach on eye-in-hand configuration within real-time visual servoing. In Proceedings of the 8th IEEE International Workshop on Advanced Motion Control, Kawasaki, Japan, 28 March 2004; Volume 10, pp. 647–652.
5. Liu, Z.; Wang, F.; Zhang, Y. Adaptive Visual Tracking Control for Manipulator With Actuator Fuzzy Dead-Zone Constraint and Unmodeled Dynamic. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 1301–1312. [\[CrossRef\]](#)
6. Li, B.; Zhang, X.; Fang, Y.; Shi, W. Visual Servoing of Wheeled Mobile Robots Without Desired Images. *IEEE Trans. Cybern.* **2019**, *49*, 2835–2844. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Mekonnen, G.; Kumar, S.; Pathak, P.M. Wireless hybrid visual servoing of omnidirectional wheeled mobile robots. *Robot. Auton. Syst.* **2016**, *75*, 450–462. [\[CrossRef\]](#)
8. Serra, P.; Cunha, R.; Hamel, T.; Cabecinhas, D.; Silvestre, C. Landing of a Quadrotor on a Moving Target Using Dynamic Image-Based Visual Servo Control. *IEEE Trans. Robot.* **2016**, *32*, 1524–1535. [\[CrossRef\]](#)
9. Zheng, D.; Wang, H.; Wang, J.; Chen, S.; Chen, W.; Liang, X. Image-Based Visual Servoing of a Quadrotor Using Virtual Camera Approach. *IEEE/ASME Trans. Mechatron.* **2017**, *22*, 972–982. [\[CrossRef\]](#)
10. Jabbari Asl, H.; Yoon, J. Adaptive vision-based control of an unmanned aerial vehicle without linear velocity measurements. *ISA Trans.* **2016**, *65*, 296–306. [\[CrossRef\]](#)
11. Yuen, H.P.; Choi, Y.W.; Kim, Y.B. Implementation of Tracking-Learning-Detection for Improving of a Stereo-Camera-based Marker-less Distance Measurement System for Vessel Berthing. In Proceedings of the 2020 16th IEEE International Colloquium on Signal Processing and its Applications, Langkawi, Malaysia, 28–29 February 2020; pp. 63–68.
12. Hilkert, J.M. Inertially Stabilized Platform Technology Concepts and Principles. *IEEE Control Syst.* **2008**, *28*, 26–46. [\[CrossRef\]](#)
13. Masten, M.K. Inemally Stabilized Platforms for Optical Imaging Systems Tracking Dynamic Targets with Mobile Sensors. *IEEE Control Syst.* **2008**, *28*, 47–64. [\[CrossRef\]](#)
14. Osborne, J.M.; Fuentes, R. Global Analysis of the Double- Gimbal Mechanism. *IEEE Control Syst.* **2008**, *28*, 44–64. [\[CrossRef\]](#)
15. Jianbo, S.; Tomasi, C. Good features to track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 21–23 June 1994; pp. 593–600.
16. Tomasi, C.; Kanade, T. *Detection and Tracking of Point Features*; Technical Report CMU-CS-91-132; Carnegie Mellon University: Pittsburgh, PA, USA, 1991.

17. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence—Volume 2, Vancouver, BC, Canada, 24–28 August 1981; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1981; pp. 674–679.
18. Exner, D.; Bruns, E.; Kurz, D.; Grundhofer, A.; Bimber, O. Fast and robust CAMShift tracking. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 9–16.
19. Bradski, G.R. Real time face and object tracking as a component of a perceptual user interface. In Proceedings of the Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201), Princeton, NJ, USA, 19–21 October 1998; pp. 214–219.
20. Kalal, Z.; Mikolajczyk, K.; Matas, J. Tracking-Learning-Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1409–1422. [[CrossRef](#)] [[PubMed](#)]
21. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [[CrossRef](#)] [[PubMed](#)]
22. Li, H.; Yu, J. Anti-Disturbance Control Based on Cascade ESO and Sliding Mode Control for Gimbal System of Double Gimbal CMG. *IEEE Access* **2020**, *8*, 5644–5654. [[CrossRef](#)]
23. Abdo, M.; Vali, A.R.; Toloei, A.; Arvan, M.R. Research on the cross-coupling of a two axes gimbal system with dynamic unbalance. *Int. J. Adv. Robot. Syst.* **2013**, *10*, 357. [[CrossRef](#)]
24. Liu, S.; Lu, T.; Shang, T.; Xia, Q. Dynamic Modeling and Coupling Characteristic Analysis of Two-Axis Rate Gyro Seeker. *Int. J. Aerosp. Eng.* **2018**, *2018*, 8513684. [[CrossRef](#)]
25. Ding, Z.; Zhao, F.; Lang, Y.; Jiang, Z.; Zhu, J. Anti-Disturbance Neural-Sliding Mode Control for Inertially Stabilized Platform With Actuator Saturation. *IEEE Access* **2019**, *7*, 92220–92231. [[CrossRef](#)]
26. Mao, J.; Yang, J.; Liu, X.; Li, S.; Li, Q. Modeling and Robust Continuous TSM Control for an Inertially Stabilized Platform With Couplings. *IEEE Trans. Control Syst. Technol.* **2020**, *28*, 2548–2555. [[CrossRef](#)]
27. Kennedy, P.J.; Kennedy, R.L. Direct versus indirect line of sight (LOS) stabilization. *IEEE Trans. Control Syst. Technol.* **2003**, *11*, 3–15. [[CrossRef](#)]
28. Hurák, Z.; Řezáč, M. Combined line-of-sight inertial stabilization and visual tracking: Application to an airborne camera platform. In Proceedings of the IEEE Conference on Decision and Control, Shanghai, China, 16–18 December 2009; pp. 8458–8463.
29. Hurák, Z.; Řezáč, M. Image-based pointing and tracking for inertially stabilized airborne camera platform. *IEEE Trans. Control Syst. Technol.* **2012**, *20*, 1146–1159. [[CrossRef](#)]
30. Antonello, R.; Oboe, R.; Ramello, A.; Ito, K.; Felicini, N.; Cenedese, A. IMU-aided image stabilization and tracking in a HSM-driven camera positioning unit. In Proceedings of the 2013 IEEE International Symposium on Industrial Electronics, Taipei, Taiwan, 28–31 May 2013; pp. 1–7.
31. Bibby, C.; Reid, I. Visual tracking at sea. In Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; Volume 2005, pp. 1841–1846.
32. Liu, X.; Mao, J.; Yang, J.; Li, S.; Yang, K. Robust predictive visual servoing control for an inertially stabilized platform with uncertain kinematics. *ISA Trans.* **2021**. [[CrossRef](#)]
33. Baillieul, J. *Encyclopedia of Systems and Control*; Baillieul, J., Samad, T., Eds.; Springer: London, UK, 2020; ISBN 978-1-4471-5102-9.
34. Åström, K.J. Limitations on control system performance. *Eur. J. Control* **2000**, *6*, 2–20. [[CrossRef](#)]
35. Zhong, Q.C. *Robust Control of Time-Delay Systems*; Springer: London, UK, 2006; ISBN 9781846282645.