Image and Vision Computing 27 (2009) 1007–1022

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Dynamic visual tracking control of a mobile robot with image noise and occlusion robustness

Chi-Yi Tsai *, Kai-Tai Song

Department of Electrical and Control Engineering, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 300, Taiwan, ROC

ARTICLE INFO

Article history:
Received 26 July 2007
Received in revised form 16 June 2008
Accepted 28 August 2008

Keywords:
Visual interaction model
Visual tracking control
Visual state estimation
Nonholonomic mobile robots
Temporary partial/full occlusion

ABSTRACT

This paper presents a robust visual tracking control design for a nonholonomic mobile robot equipped with a tilt camera. This design aims to allow the mobile robot to keep track of a dynamic moving target in the camera's field-of-view; even though the target is temporarily fully occluded. To achieve this, a control system consisting of a visual tracking controller (VTC) and a visual state estimator (VSE) is proposed. A novel visual interaction model is derived to facilitate the design of VTC and VSE. The VSE is responsible for estimating the optimal target state and target image velocity in the image space. The VTC then calculates the corresponding command velocities for the mobile robot to work in the world coordinates. The proposed VSE not only possesses robustness against the image noise, but also overcomes the temporary occlusion problem. Computer simulations and practical experiments of a mobile robot to track a moving target have been carried out to validate the performance and robustness of the proposed system.

1. Introduction

In recent years, computer vision has become a major on-board sensor for autonomous mobile robots. Among the various applications of vision systems, visual tracking plays an important role in autonomous navigation and control. With visual tracking, a robot needs to focus on a target and interact with it accordingly. Thus, the study of *visual tracking control*, i.e. the vision-based robot motion control to track an interesting target, has become an active area in robotics research [4–21]. Based on motion constraints and the scenario of robotic applications, the research on visual tracking control can be categorized into visual servoing for holonomic manipulators and visual tracking for nonholonomic mobile robots. Visual servoing technique for holonomic manipulators and robot hands has been investigated extensively and many powerful tools can be found in the literature [1–3]. However, the results for holonomic manipulators are unsuitable for most mobile robots due to the nonholonomic motion constraints on the mobile platform.

This paper addresses the problem of visual tracking control of a wheeled mobile robot equipped with an on-board monocular camera. Such a visual tracking control task encompasses several interesting issues such as vision-based motion control of mobile robots, estimation of Jacobian matrix, uncertainties caused by image noise and occlusion during visual tracking process. Most researchers focus on the design of visual tracking controllers to track a static object, such as a ground line, landmark, or reference image for the purpose of visual navigation [4–7] or visual regulation (e.g. visual homing) [8–12]. Because tracking a dynamic moving target is an important requirement for many intelligent robots, some efforts focus on visual tracking control design of a moving (non-static) target for the purpose of visual formation control [13,14], visual interception [15,16], visual platooning [17] and human–robot interaction [18]. However, only limited works deal with the external uncertainties during visual tracking process such as image noise, varying illumination and temporary occlusion. To overcome the temporary partial occlusion problem, Malis et al. combined a model-free visual servoing method with a template-based visual tracking algorithm to build a flexible and robust visual tracking control system [19]. In their work, the visual tracking algorithm aims to estimate an optimal homography matrix between the reference pattern and the pattern in current image even if the pattern is partially occluded. In [20], Comport et al. combined robust statistical techniques with visual servoing control law in order to overcome the position uncertainty of image features caused by varying light source and multiple occlusion problems. However, these reported systems might fail in the full occlusion condition due to the requirement on feature matching. To resolve the temporary full occlusion problem, Han et al. proposed a differential approximation approach to measure the velocity of the target in the image frame (termed as *target image velocity*) [21]. When the target is fully occluded, the position of the target in the next image is estimated by the measured target image velocity. However, the estimation result is very sensitive to the image noise due to the shortcoming of the differential approximation approach. In the realization of tracking control schemes, it has been noted that

* Corresponding author.
E-mail addresses: chiyi.ece91g@nctu.edu.tw (C.-Y. Tsai), ktsong@mail.nctu.edu.tw (K.-T. Song).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.08.011

the image noise usually happens during the visual tracking process of mobile robot due to the position-dependent and light-dependent external uncertainties. Therefore, their method cannot efficiently overcome the image noise caused by the external uncertainties. Further, in the estimation of Jacobian matrix, the reported methods usually assume that the target is stationary during visual tracking operation. However, when the target is non-stationary, the existing methods fail to provide a suitable solution.

From the above discussion, we note that it is still a challenge to develop a robust control scheme against the image noise and temporary (partial/full) occlusion uncertainties during visual tracking tasks. These problems have not yet been addressed, which motivated us to design a robust visual tracking control system to track a dynamic moving target and overcome the image noise and temporary occlusion. To achieve this, a novel image-based camera–object visual interaction model is derived in order to help the estimation of Jacobian matrix under non-stationary target condition. Next, a robust visual tracking control system, which consists of a visual tracking controller (VTC) and a visual state estimator (VSE), is developed based on the proposed visual interaction model. The VSE aims to estimate the optimal target state and target image velocity from image plane directly, and the VTC then calculates the corresponding control velocities for the mobile robot. The main advantages of the proposed method compared to the existing methods are summarized in the following.

(1) The proposed method allows the robot to track a predictable as well as unpredictable moving target. Compared to the method given in [16], where only straight motion of a unicycle in 2D space is reported, the proposed method can handle holonomic target motion in 3D space.
(2) Based on Kalman filtering algorithm [22], the proposed method provides the best linear estimate of target state from the observed image sequence, which contains both random noise and temporary occlusion uncertainties. Therefore, the proposed system is robust to the uncertainties of image noise and temporary occlusion.

Note that the main issue addressed in this paper is the vision-based estimation problem for robotic visual tracking control applications. For a discussion on vision-based control problem (e.g. the analysis of system parametric robustness), please refer to [23] for the technical details.

The rest of this paper is organized as follows. Section 2 derives the proposed visual interaction model. Section 3 presents the results of VTC design. In Section 4, the design of VSE is presented to estimate the optimal system state in the image plane for handling the uncertainties caused by image noise and temporary occlusion. In Section 5, computer simulations are employed to validate the system robustness against random image noise. Practical experiments using a robot to track a moving target have been conducted to verify the system robustness against temporary occlusion. Section 6 concludes this paper.

2. Camera–object visual interaction model

The basic assumptions of the proposed method are listed as follows:

(1) The proposed VSE has the same basic assumptions as Kalman filters in terms of Gaussian distribution uncertainty, smoothness motion, and uniform sampling rate.
(2) The on-board camera is supposed to be a calibrated pinhole camera. Because the proposed VTC possesses some degree of robustness against parametric uncertainty [23], a simple linear camera calibration method [24] can be used to estimate the intrinsic parameters of the camera.
(3) In the derivations below, the width of target is supposed to be *a priori* known constant to simplify the depth estimation process. However, any algorithm or sensor which provides the depth information can be utilized to combine with the proposed method.

In the following, the visual interaction model between a mobile robot and a dynamic moving target is derived. We first introduce a kinematics model of the wheeled mobile robot in relation to a target discussed in this paper. The mathematical derivations of the proposed model are then presented and explained.

2.1. Kinematics model of wheeled mobile robot and target

Fig. 1 illustrates the model of wheeled mobile robot and target considered in the nonholonomic visual tracking control problem. The wheeled mobile robot is equipped with a tilt camera to track a dynamic-motion target, which is supposed to be a well-recognizable object with appropriate dimensions in the image plane and zero angular motion relative to the robot. The tilt camera is mounted on top of the mobile robot and its optical-axis faces the target of interest, for instance, a human face. Fig. 1(a) shows the model of the wheeled mobile robot and the target in the world coordinate frame F_f (see Fig. 2), in which the motion of the target is supposed to be holonomic such that

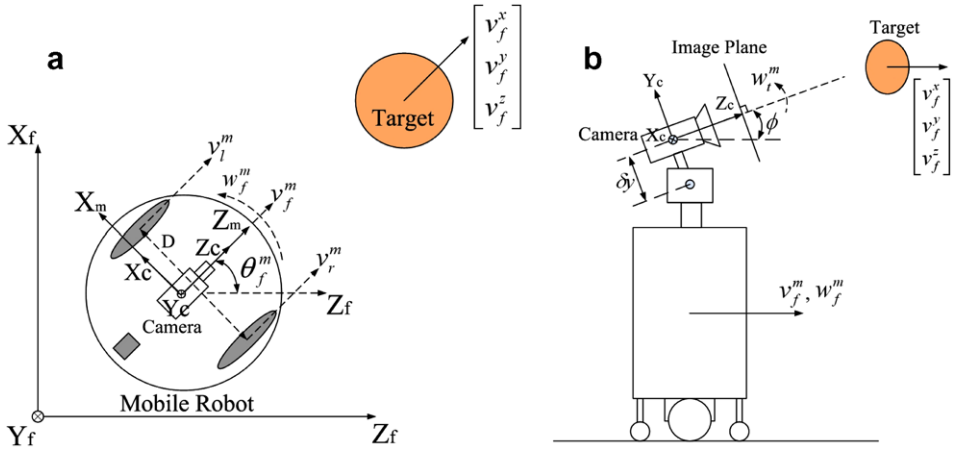

Fig. 1. (a) A model of the wheeled mobile robot and the target in the world coordinate frame. (b) Side view of the wheeled mobile robot with a tilt camera mounted on top of it.

$$\dot{X}_f^t = V_f^t, \tag{1}$$

where $X_f^t = [x_f^t \quad y_f^t \quad z_f^t]^T$ and $V_f^t = [v_f^x \quad v_f^y \quad v_f^z]^T$ denote, respectively, the position and velocity of target in world coordinates.

Fig. 1(b) is the side view of the scenario under consideration, in which the tilt angle ϕ gives the relationship between camera coordinate frame F_c and the mobile coordinate frame F_m. The kinematics of the wheeled mobile robot is described by [25]

$$\dot{X}_f^m = \begin{bmatrix} v_f^m \sin\theta_f^m \\ 0 \\ v_f^m \cos\theta_f^m \end{bmatrix} \quad \text{and} \quad \begin{matrix} \dot{\theta}_f^m = w_f^m \\ \dot{\phi} = w_t^m \end{matrix}, \tag{2}$$

where $X_f^m = [x_f^m \quad y_f^m \quad z_f^m]^T$ is the position of the mobile robot in the world coordinates, (θ_f^m, ϕ) represent the orientation angle of the mobile robot and the tilt angle of the onboard camera, w_t^m is the tilt velocity of the camera, and (v_f^m, w_f^m) are the linear and angular velocities of the mobile robot. In practice, (v_f^m, w_f^m) can be used to calculate the velocity of each wheel of the mobile robot such that

$$v_l^m = v_f^m - (\mathbf{D}\cdot w_f^m)/2 \quad \text{and} \quad v_r^m = v_f^m + (\mathbf{D}\cdot w_f^m)/2, \tag{3}$$

where (v_l^m, v_r^m) are the left- and right-wheel velocities, respectively, and D represents the distance between the two drive wheels. In the rest of this paper, the target model (1) and mobile robot model (2) will be utilized to derive the visual interaction model and to design the visual tracking control system.

2.2. Kinematics of camera–object interaction in the camera coordinate frame

Fig. 2 illustrates the relationship between world, camera and image coordinate frames. Let $X_f = X_f^t - X_f^m$ denote the related position between mobile robot and the target in the world coordinate frame. In order to describe a mobile robot interacting with the target in the image coordinate frame, a visual interaction model has been derived by transferring the kinematics of X_f from the world coordinate frame into the image coordinate frame. This subsection presents the transformation of the kinematics of X_f from world coordinate frame into camera coordinate frame.

As shown in Fig. 2, $X_c = [x_c \quad y_c \quad z_c]^T$ denotes the related position in the camera coordinate frame and can be calculated by the coordinate transformation such that

$$X_c = \mathbf{R}(\phi, \theta_f^m)X_f - \delta Y, \tag{4}$$

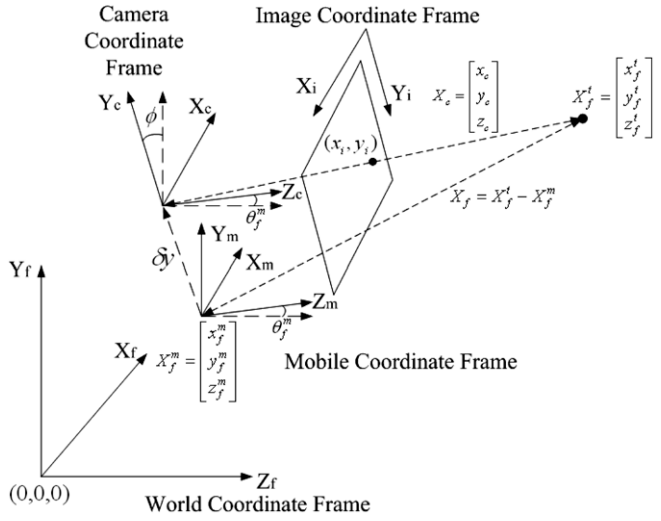

Fig. 2. World (subscript f), camera (subscript c) and image (subscript i) coordinate frames of robotic visual interaction.

where

$$\mathbf{R}(\phi, \theta_f^m) = \mathbf{R}(\phi)\mathbf{R}(\theta_f^m) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}\begin{bmatrix} \cos\theta_f^m & 0 & -\sin\theta_f^m \\ 0 & 1 & 0 \\ \sin\theta_f^m & 0 & \cos\theta_f^m \end{bmatrix},$$

$\delta Y = [0 \quad \delta y \quad 0]^T$.

δy is the distance between the center of robot tilt platform and the onboard camera. Because $\delta Y = [0 \quad \delta y \quad 0]^T$ is a constant translational vector, the derivative of (4) becomes

$$\dot{X}_c = \left[\frac{\partial \mathbf{R}(\phi, \theta_f^m)}{\partial \phi}\dot{\phi} + \frac{\partial \mathbf{R}(\phi, \theta_f^m)}{\partial \theta_f^m}\dot{\theta}_f^m\right]X_f + \mathbf{R}(\phi, \theta_f^m)\dot{X}_f, \tag{5}$$

where

$$\frac{\partial \mathbf{R}(\phi, \theta_f^m)}{\partial \phi} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}\mathbf{R}(\phi, \theta_f^m) \equiv \mathbf{\Psi}_1\mathbf{R}(\phi, \theta_f^m),$$

$$\frac{\partial \mathbf{R}(\phi, \theta_f^m)}{\partial \theta_f^m} = \begin{bmatrix} 0 & \sin\phi & -\cos\phi \\ -\sin\phi & 0 & 0 \\ \cos\phi & 0 & 0 \end{bmatrix}\mathbf{R}(\phi, \theta_f^m) \equiv \mathbf{\Psi}_\phi\mathbf{R}(\phi, \theta_f^m).$$

Substituting (1), (2) and (4) into (5), the kinematics of interaction between the robot and the target in the camera frame can be obtained such that

$$\begin{aligned} \dot{X}_c &= (\mathbf{\Psi}_1 w_t^m + \mathbf{\Psi}_\phi w_f^m)\mathbf{R}(\phi, \theta_f^m)X_f + \mathbf{R}(\phi, \theta_f^m)\dot{X}_f \\ &= \mathbf{A}_c X_c + \mathbf{B}_c u + \mathbf{R}(\phi, \theta_f^m)V_f^t, \end{aligned} \tag{6}$$

where

$$\mathbf{A}_c = \mathbf{\Psi}_1 w_t^m + \mathbf{\Psi}_\phi w_f^m = \begin{bmatrix} 0 & w_f^m \sin\phi & -w_f^m \cos\phi \\ -w_f^m \sin\phi & 0 & -w_t^m \\ w_f^m \cos\phi & w_t^m & 0 \end{bmatrix} \quad \text{and}$$

$$\mathbf{B}_c = \begin{bmatrix} 0 & \delta y \sin\phi & 0 \\ \sin\phi & 0 & 0 \\ -\cos\phi & 0 & \delta y \end{bmatrix}.$$

$u = [v_f^m \quad w_f^m \quad w_t^m]^T$ is the control velocity of the mobile robot and on-board tilt camera. In the following, the kinematics model (6) will be used to derive the interaction model in the image coordinate frame.

2.3. Kinematics of camera–object interaction in the image coordinate frame

In this section, the related position X_c is transformed into the image coordinate frame for deriving the visual interaction model based on (6). We first define the system state in the image plane for the estimator and controller design. Fig. 3 illustrates the definition of observed and desired system state in the image plane. In Fig. 3, x_i and y_i are, respectively, the horizontal and vertical coordinates of the centroid of a target in the image plane and d_x is the width of the target in the image plane. Similar to human's visual tracking behavior, the purpose of the visual tracking control design is to control the centroid position and width of target from an initial state into the desired state in the image plane.

In the following, the visual interaction model is derived using (6) and the selected system state. Based on the pinhole camera model, the diffeomorphism (see [26] for detailed description) in the image plane can be defined by standard projection equations [24] such that:

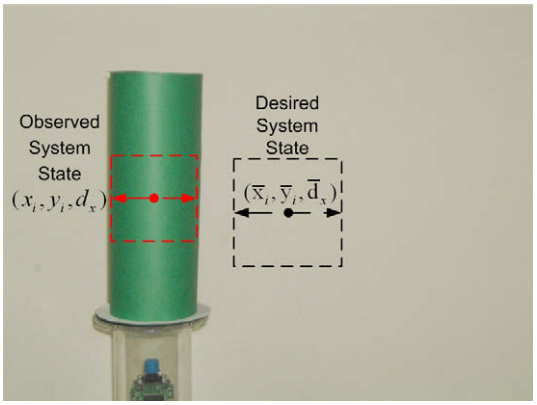

Fig. 3. The definition of observed and desired system states in the image plane.

$$X_i = [x_i \quad y_i \quad d_x]^{\mathrm{T}} = [-k_x x_c \quad k_y y_c \quad k_x k_w z_c]^{\mathrm{T}},$$
$$k_x = f_x/z_c, \quad k_y = f_y/z_c, \quad k_w = W/z_c, \tag{7}$$

where (f_x, f_y) represent fixed focal length along the image x-axis and y-axis, respectively, and W denotes the actual width of the target. By taking the derivative of (7), the kinematic relationship between image and camera coordinate frames can be found such that

$$\dot{X}_i = \mathbf{P}_i \dot{X}_c, \quad \mathbf{P}_i = \begin{bmatrix} -k_x & 0 & -k_x f_x^{-1} x_i \\ 0 & k_y & -k_y f_y^{-1} y_i \\ 0 & 0 & -k_x f_x^{-1} d_x \end{bmatrix}. \tag{8}$$

Substituting (6) and (7) into (8), the kinematic relationship between robot and target in the image coordinate frame can be modeled such that

$$\dot{X}_i = \mathbf{A}_i X_i + \mathbf{B}_i u + C_i, \tag{9}$$

where $\mathbf{A}_i = \mathrm{diag}(A_1, A_2, A_1)$,

$$A_1 = -\frac{k_x}{f_x}(v_f^x \cos\phi \sin\theta_f^m + v_f^y \sin\phi + v_f^z \cos\phi \cos\theta_f^m),$$
$$A_2 = -\frac{k_y}{f_y}(v_f^x \cos\phi \sin\theta_f^m + v_f^y \sin\phi + v_f^z \cos\phi \cos\theta_f^m),$$
$$C_i = \begin{bmatrix} k_x(v_f^z \sin\theta_f^m - v_f^x \cos\theta_f^m) \\ k_y(v_f^y \cos\phi - v_f^x \sin\phi \sin\theta_f^m - v_f^z \sin\phi \cos\theta_f^m) \\ 0 \end{bmatrix},$$
$$\mathbf{B}_i = \begin{bmatrix} \frac{k_x}{f_x} x_i \cos\phi & (\frac{x_i^2 + f_x^2}{f_x})\cos\phi - \frac{f_x}{f_y}(k_y \delta y + y_i)\sin\phi & -\frac{x_i(k_y \delta y + y_i)}{f_y} \\ k_y(\sin\phi + \frac{y_i}{f_y}\cos\phi) & \frac{f_y}{f_x} x_i(\sin\phi + \frac{y_i}{f_y}\cos\phi) & -\frac{y_i^2 + f_y^2 + k_y y_i \delta y}{f_y} \\ \frac{k_x}{f_x} d_x \cos\phi & \frac{x_i d_x}{f_x}\cos\phi & -\frac{d_x(k_y \delta y + y_i)}{f_y} \end{bmatrix},$$

where $diag(a,b,c)$ denotes a 3×3 diagonal matrix with diagonal elements a, b, and c. The elements of system matrix $\mathbf{A}_i$ and vector C_i are dependent on the robot pose and target velocity, and the elements of control matrix $\mathbf{B}_i$ are dependent on the robot pose and current system state.

The visual interaction model (9) indicates that the elements of system matrix $\mathbf{A}_i$ and vector C_i are functions of target velocity. Thus, expression (9) can be rewritten as

$$\dot{X}_i = (\mathbf{A}_i X_i + C_i) + \mathbf{B}_i u = \mathbf{J}_i V_f^t + \mathbf{B}_i u, \tag{10}$$

where

$$\mathbf{J}_i = \begin{bmatrix} -k_x(\frac{x_i}{f_x}\cos\phi \sin\theta_f^m + \cos\theta_f^m) & -k_x \frac{x_i}{f_x}\sin\phi & -k_x(\frac{x_i}{f_x}\cos\phi\cos\theta_f^m - \sin\theta_f^m) \\ -k_y(\frac{y_i}{f_y}\cos\phi \sin\theta_f^m + \sin\phi\sin\theta_f^m) & -k_y(\frac{y_i}{f_y}\sin\phi - \cos\phi) & -k_y(\frac{y_i}{f_y}\cos\phi\cos\theta_f^m + \sin\phi\cos\theta_f^m) \\ -k_x \frac{d_x}{f_x}\cos\phi \sin\theta_f^m & -k_x \frac{d_x}{f_x}\sin\phi & -k_x \frac{d_x}{f_x}\cos\phi\cos\theta_f^m \end{bmatrix}.$$

Expression (10) shows that the visual interaction model consists of two parts: the part of describing target motion $\dot{X}_i^t \equiv [\dot{x}_i^t \quad \dot{y}_i^t \quad \dot{d}_x^t]^{\mathrm{T}} = \mathbf{J}_i V_f^t$, and the effect of mobile robot motion $\dot{X}_i^m \equiv [\dot{x}_i^m \quad \dot{y}_i^m \quad \dot{d}_x^m]^{\mathrm{T}} = \mathbf{B}_i u$. Thus, (10) can be rewritten as a dual-Jacobian equation such that

$$\dot{X}_i = \dot{X}_i^t + \dot{X}_i^m = \mathbf{J}_i V_f^t + \mathbf{B}_i u, \tag{11}$$

where the matrix $\mathbf{J}_i$, termed as *target image Jacobian*, transforms the target velocity V_f^t into target image velocity $\dot{X}_i^t$; matrix $\mathbf{B}_i$, termed as *robot image Jacobian*, transforms the mobile robot control velocity u into robot image velocity $\dot{X}_i^m$. In other words, the image velocity $\dot{X}_i$ is caused by a combination of target image velocity $\dot{X}_i^t$ and robot image velocity $\dot{X}_i^m$. Therefore, the visual interaction between robot and target in the image coordinate frame can be modeled as a *dual-Jacobian* visual interaction model (11), which combines the motion effect of mobile robot together with the moving target.

Remark 1. The scalars $k_x = f_x/z_c$ and $k_y = f_y/z_c$ in (7) depend on the depth information between camera and target. The estimation of depth information is a demanding task in visual tracking control design; especially when a single camera is used. Thus, an algorithm or sensor which provides the depth information is usually adopted during the visual tracking process. In order to simplify the depth estimation process, an alternative is to assume that the width of target is known *a priori*. Therefore, the scalars k_x and k_y can be calculated using the state variable d_x directly based on the fact that $k_x = f_x/z_c = d_x/W$ and $k_y = k_x f_y/f_x$, where W denotes the width of target for a specific target.

3. Visual tracking controller (VTC)

In this section, a visual tracking control law based on the proposed dual-Jacobian visual interaction model (11) for tracking a target of interest in the image plane is derived exploiting feedback linearization and pole placement approaches.

3.1. Dynamic error state model

In order to control the system state variables from an initial state to the desired state, an error state model is formed to facilitate the tracking controller design. The error state in the image plane is defined as

$$X_e = [x_e \quad y_e \quad d_e]^{\mathrm{T}} = [\bar{x}_i - x_i^* \quad \bar{y}_i - y_i^* \quad \bar{d}_x - d_x^*]^{\mathrm{T}} = \bar{X}_i - X_i^*, \tag{12}$$

where $\bar{X}_i = [\bar{x}_i \quad \bar{y}_i \quad \bar{d}_x]^{\mathrm{T}}$ is the vector of fixed desired states in the image plane (as shown in Fig. 3); $X_i^* = [x_i^* \quad y_i^* \quad d_x^*]^{\mathrm{T}}$ is the vector of the estimated states from the VSE (see Section 4). Using the error state (10), a *dynamic error state model* in the image plane can be derived by taking the derivative of (12) such that

$$\dot{X}_e = -\dot{X}_i^t - \dot{X}_i^m = -\mathbf{J}_i V_f^t - \mathbf{B}_i u. \tag{13}$$

With the new error state X_e, the visual tracking control problem is transformed into a stability problem. If X_e converges to zero, then the visual tracking control problem is solved.

3.2. Feedback linearization and pole placement

Based on the dynamic error state model (13), we choose the feedback linearization control law for both the mobile robot and the tilt camera such that

$$u = \begin{bmatrix} v_f^m & w_f^m & w_t^m \end{bmatrix}^T = \mathbf{B}_i^{-1}(\mathbf{K}_g X_e - \mathbf{J}_i V_f^t), \quad (14\text{-}1)$$

$$= \mathbf{B}_i^{-1}(\mathbf{K}_g X_e - \dot{X}_i^t), \quad (14\text{-}2)$$

where (v_f^m, w_f^m, w_t^m) are the control inputs for robot linear velocity, robot angular velocity and camera tilt velocity, respectively; $\mathbf{K}_g$ is a 3×3 gain matrix. We choose the gain matrix such hat

$$\mathbf{K}_g = \mathrm{diag}(\alpha_1, \alpha_2, \alpha_3), \quad (15)$$

in which $(\alpha_1, \alpha_2, \alpha_3)$ are the three positive constants. Substituting (14) and (15) into (13), the closed-loop linearized system is obtained as

$$\dot{X}_e = -\mathbf{K}_g X_e = -\mathrm{diag}(\alpha_1, \alpha_2, \alpha_3) X_e. \quad (16)$$

Expression (16) indicates that

$$X_e(t) = \mathrm{diag}(e^{-\alpha_1 t}, e^{-\alpha_2 t}, e^{-\alpha_3 t}) X_e(0). \quad (17)$$

Because $(\alpha_1, \alpha_2, \alpha_3) > 0$ are positive constants, the estimated system state $X_i^*(t)$ will converge exponentially to the desired system state $\bar{X}_i$. This implies that if the controller u and the gain matrix $\mathbf{K}_g$ are chosen as given in (14) and (15), respectively, the closed-loop visual tracking system described in (13) will be transformed into an asymptotically stable linear time-invariant (LTI) system and the visual tracking control problem is solved.

Remark 2. Although the proposed image control law (14) results in a smooth convergence in the image plane, it still has to prove that the robot should follow the target. The proof of this problem is presented in Appendix.

3.3. Singularity analysis

The feedback linearization control law (14) poses a singularity problem of matrix $\mathbf{B}_i$. By directly computing the determinant of matrix $\mathbf{B}_i$, the sole singularity of matrix $\mathbf{B}_i$ can be found such that

$$f_y = (y_i + S d_x) \tan\phi, \quad (18)$$

where $S = (f_y \delta y)/(f_x W)$ is a fixed scalar factor. Since $k_x = f_x/z_c = d_x/W$ and $k_y = k_x f_y/f_x$, (18) can be rewritten such that

$$\tan\phi = \frac{f_y}{y_i + k_y \delta y}. \quad (19)$$

Moreover, since $f_y = k_y z_c$ and $y_i = k_y y_c$, (19) equals

$$\tan\phi = \frac{z_c}{y_c + \delta y}. \quad (20)$$

As shown in Fig. 4, let ϕ' be the angle related to the location of the target, we have the following geometric relationship:

$$\tan(\phi + \phi\prime) = \frac{z_c}{y_c + \delta y}. \quad (21)$$

From (20) and (21), it is clear that the matrix becomes $\mathbf{B}_i$ singular when ϕ' equals to 0 or π. The physical meaning of these conditions is that the target is directly above or directly below the robot, and the robot will not be able to approach the target in any way due to insufficient degree-of-freedom. Therefore, the robot will stop tracking temporarily under such circumstances.

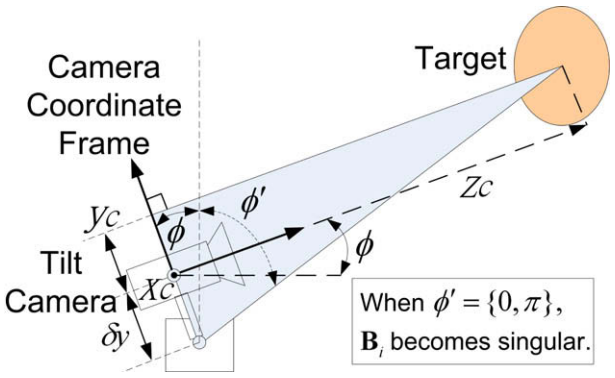

Fig. 4. Physical meaning of the singularity condition (18).

4. Visual state estimator

In Section 3.2, the visual tracking control law (14) requires information about target 3D velocity V_f^t or target image velocity $\dot{X}_i^t$. If V_f^t is known, the first visual tracking control law (14-1) only needs an estimate of target status X_i to calculate the control signal u. However, in practical applications, it is difficult to estimate V_f^t on-line in real time when using only one camera. In this situation, the second visual tracking control law (14-2) provides a useful solution which only needs the target image velocity $\dot{X}_i^t$ in the image plane. In this section, two VSEs will be proposed in order to estimate the necessary information for the VTC. The first VSE is developed under the condition that the target velocity V_f^t is known, and the second VSE is designed by releasing this condition, which will facilitate more general applications of the proposed tracking control scheme in image plane.

4.1. VSE design with target velocity information (VSE-WTV)

In the case that the target 3D velocity V_f^t is known, the VSE-WTV can be designed based on the system model (9) to estimate the optimal target status X_i in the image plane. To achieve this, a propagation model is required in order to facilitate the design of VSE-WTV. This subsection presents the derivation of the required propagation model and the proposed VSE-WTV algorithm.

4.1.1. Propagation model for VSE-WTV

The first step in the derivation of the propagation model is to discretize the system model (9) into the corresponding discrete form. By the definition $\dot{x}(t) = \lim_{T \to 0}[x(t) - x(t-T)]/T$, where T denotes the sampling time of the digital system, we can approximate the system model (9) as

$$X_i^p[n] = (\mathbf{I}_3 + T\mathbf{A}_i) X_i^*[n-1] + T\mathbf{B}_i u_{n-1} + TC_i \quad \text{for } n = 1, 2, \ldots, \quad (22)$$

where $X_i^p[n]$ is the propagated system state at a sample instant n, $\mathbf{I}_3$ is a 3×3 identity matrix, $X_i^*[n-1] = \begin{bmatrix} x_i^* & y_i^* & d_x^* \end{bmatrix}^T$ denotes the estimated system state at a sample instant $n-1$, and $u_{n-1} = \begin{bmatrix} v_f^m & w_f^m & w_t^m \end{bmatrix}^T$ is the control signal at a sample instant $n-1$.

The second step is to analyze the covariance matrix of discrete-time propagation equation (22). We first introduce the following variables:

$$x_i^* = x_i + \delta x_i, \quad y_i^* = y_i + \delta y_i, \quad d_x^* = d_x + \delta d_x, \quad v_f^{m*} = v_f^m + \delta v_f^m,$$
$$w_f^{m*} = w_f^m + \delta w_f^m, \quad w_t^{m*} = w_t^m + \delta w_t^m, \quad (23)$$

where (x_i, y_i, d_x) are the system state variables, (x_i^*, y_i^*, d_x^*) denote the estimated state variables and $(\delta x_i, \delta y_i, \delta d_x)$ are the corresponding state estimation errors. Based on the velocity transformation (3), the estimated velocities $(v_{lc}^m, v_{rc}^m, w_{tc}^m)$ and velocity commands

(v_l^m, v_r^m, w_t^m) can be obtained by $(v_f^{m*}, w_f^{m*}, w_t^{m*})$ and (v_f^m, w_f^m, w_t^m), respectively, such that

$$\begin{bmatrix} v_{lc}^m \\ v_{rc}^m \\ w_{tc}^m \end{bmatrix} = \begin{bmatrix} 1 & -D/2 & 0 \\ 1 & D/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_f^{m*} \\ w_f^{m*} \\ w_t^{m*} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} v_l^m \\ v_r^m \\ w_t^m \end{bmatrix} = \begin{bmatrix} 1 & -D/2 & 0 \\ 1 & D/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_f^m \\ w_f^m \\ w_t^m \end{bmatrix}. \tag{24}$$

Using (24), the estimated velocity errors $(\delta v_f^m, \delta w_f^m, \delta w_t^m)$ can be obtained by the velocity inverse transformation

$$\begin{bmatrix} \delta v_f^m \\ \delta w_f^m \\ \delta w_t^m \end{bmatrix} = \begin{bmatrix} v_f^{m*} - v_f^m \\ w_f^{m*} - w_f^m \\ w_t^{m*} - w_t^m \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ -1/D & 1/D & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{lc}^m - v_l^m \\ v_{rc}^m - v_r^m \\ w_{tc}^m - w_t^m \end{bmatrix}. \tag{25}$$

Substituting (23) into (22) and canceling common terms, the state estimation errors can be approximated by neglecting the higher-order terms in the discrete-time error propagation equation such that

$$\delta X_i[n] = \mathbf{A}_\delta \delta X_i[n-1] + T\mathbf{B}_i \delta u_{n-1}, \tag{26}$$

where $\delta X_i[n-1] = [\,\delta x_i \quad \delta y_i \quad \delta d_x\,]^T$ is the error propagation state, $\delta u_{n-1} = [\,\delta v_f^m \quad \delta w_f^m \quad \delta w_t^m\,]^T$ is the estimated velocity error, and

$$\mathbf{A}_\delta = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

where

$$\begin{aligned}
a_{11} &= 1 + T\left\{A_1 + \frac{1}{f_x}\left[(k_x v_f^{m*} + 2x_i^* w_f^{m*})\cos\phi - k_x\left(\delta y + \frac{y_i^*}{k_y}\right) w_t^{m*}\right]\right\} \\
a_{12} &= -T\frac{f_x}{f_y}\left(w_f^{m*}\sin\phi + \frac{x_i^*}{f_x} w_t^{m*}\right), \\
a_{13} &= \frac{T}{d_x^*}\left[x_i^* A_1 + k_x\left(\frac{x_i^*}{f_x} v_f^{m*}\cos\phi - \delta y w_f^{m*}\sin\phi - \frac{x_i^*\delta y}{f_x} w_t^{m*} + v_f^z \sin\theta_f^m - v_f^x \cos\theta_f^m\right)\right], \\
a_{21} &= T\frac{f_y}{f_x}\left(\sin\phi + \frac{y_i^*}{f_y}\cos\phi\right) w_f^{m*}, \\
a_{22} &= 1 + T\left[A_2 + \frac{v_f^{m*}}{f_y}\cos\phi + \frac{x_i^*}{f_x} w_f^{m*}\cos\phi - \frac{1}{f_y}(2y_i^* + k_y\delta y) w_t^{m*}\right], \\
a_{23} &= \frac{T}{d_x^*}\left\{y_i^* A_2 + k_y\left[\left(\sin\phi + \frac{y_i^*}{f_y}\cos\phi\right) v_f^{m*} - \frac{y_i^*\delta y}{f_y} w_t^{m*} + v_f^y\cos\phi - v_f^x \sin\phi\sin\theta_f^m - v_f^z\sin\phi\cos\theta_f^m\right]\right\}, \\
a_{31} &= T\frac{d_x^*}{f_x} w_f^{m*}\cos\phi, \quad a_{32} = -T\frac{d_x^*}{f_y} w_t^{m*}, \\
a_{33} &= 1 + T\left[2A_1 + \left(\frac{2k_y}{f_y} v_f^{m*} + \frac{x_i^*}{f_x} w_f^{m*}\right)\cos\phi - \frac{(2k_y\delta y + y_i^*) w_t^{m*}}{f_y}\right].
\end{aligned}$$

Because $\delta X_i[n-1]$ and $\delta u[n-1]$ are uncorrelated, the covariance matrix propagation equation can be obtained by adopting (26) such that

$$\mathbf{P}_n = E\left\{\delta X_i[n]\delta X_i^T[n]\right\} = \mathbf{A}_\delta \mathbf{P}_{n-1}^* \mathbf{A}_\delta^T + T^2 \mathbf{B}_i \mathbf{W}_{n-1} \mathbf{B}_i^T, \tag{27}$$

where $\mathbf{W}_{n-1} = E\{\delta u[n-1]\delta u^T[n-1]\}$ is the covariance matrix of the estimated velocity error. Applying (22) and (27), the system state and the corresponding covariance matrix can be propagated.

4.1.2. Observation and correction for VSE-WTV

In this section, the propagated system state and the propagation covariance matrix will be corrected using the observation from the camera

$$Z_n = \mathbf{I}_3 X_i[n] + \delta Z_n,$$

where $\delta Z_n \sim N(0, \mathbf{R}_n)$ denotes Gaussian observation uncertainty with zero mean and covariance matrix $\mathbf{R}_n$ at a sample instant n. The correction procedure is given by

$$X_i^*[n] = X_i^p[n] + \mathbf{K}_n\left\{Z_n - X_i^p[n]\right\} \tag{28}$$

and

$$\mathbf{P}_n^* = (\mathbf{I}_3 - \mathbf{K}_n)\mathbf{P}_n, \tag{29}$$

where $\mathbf{K}_n$ is the Kalman gain matrix given by

$$\mathbf{K}_n = \mathbf{P}_n(\mathbf{P}_n + \mathbf{R}_n)^{-1}. \tag{30}$$

Finally, the corrected system state $X_i^*[n]$ and the corresponding covariance matrix $\mathbf{P}_n^*$ are the optimal estimates at a sample instant n.

4.1.3. Summary of the proposed VSE-WTV algorithm

Based on the propagation equations (22) and (27) and the correction equations (28)–(30), the VSE-WTV can be summarized as follows:

(1) Assume that the initial position of target is located in the field-of-view of the camera, then initialize the estimated system state $X_i^*[0]$ and the propagation covariance matrix $\mathbf{P}_0$ by the first observation such that $X_i^*[0] = Z_0$ and $\mathbf{P}_0 = \mathbf{R}_0$. The proposed VTC starts to work.
(2) Compute the propagated system state $X_i^p[n]$ and the corresponding covariance matrix $\mathbf{P}_n$ using (22) and (27), respectively.
(3) If the target to be tracked is detected in the observed image, then compute the Kalman gain matrix $\mathbf{K}_n$ using (30); otherwise set $X_i^*[n] = X_i^p[n]$ and $\mathbf{P}_n^* = \mathbf{P}_n$, go to step 5.
(4) Correct the estimated state vector $X_i^*[n]$ and the corresponding covariance matrix $\mathbf{P}_n^*$ using (28) and (29), respectively.
(5) Let $X_i^*[n-1] = X_i^*[n]$ and $\mathbf{P}_{n-1}^* = \mathbf{P}_n^*$, then go to step 2.

Remark 3. Because the observation uncertainty usually varies with the conditions of target motion (such as orientation and rotation of the target) and working environment (such as light variation and occlusion), the corresponding covariance matrix $\mathbf{R}_n$ would be time-varying for different operating conditions. In order to overcome this problem, a real-time self-tuning algorithm to choose a suitable observation covariance matrix $\mathbf{R}_n$ in varying environmental conditions has been proposed in [27] and is employed in this work. More technical details can be found in [27].

4.2. VSE design without target velocity information (VSE-WoTV)

The VSE-WoTV aims to estimate the optimal target status X_i and target image velocity $\dot{X}_i^t$ from image space directly without the information of target 3D-velocity V_t^t. In this case, the dual-Jacobian visual interaction model (11) plays an essential role in the estimator design. The same procedure presented in Section 4.1 will be adopted in the design of VSE-WoTV algorithm.

4.2.1. Propagation model for VSE-WoTV

To derive the required propagation model for the design of VSE-WoTV, we first discretize the system model (11) into the discrete form such that

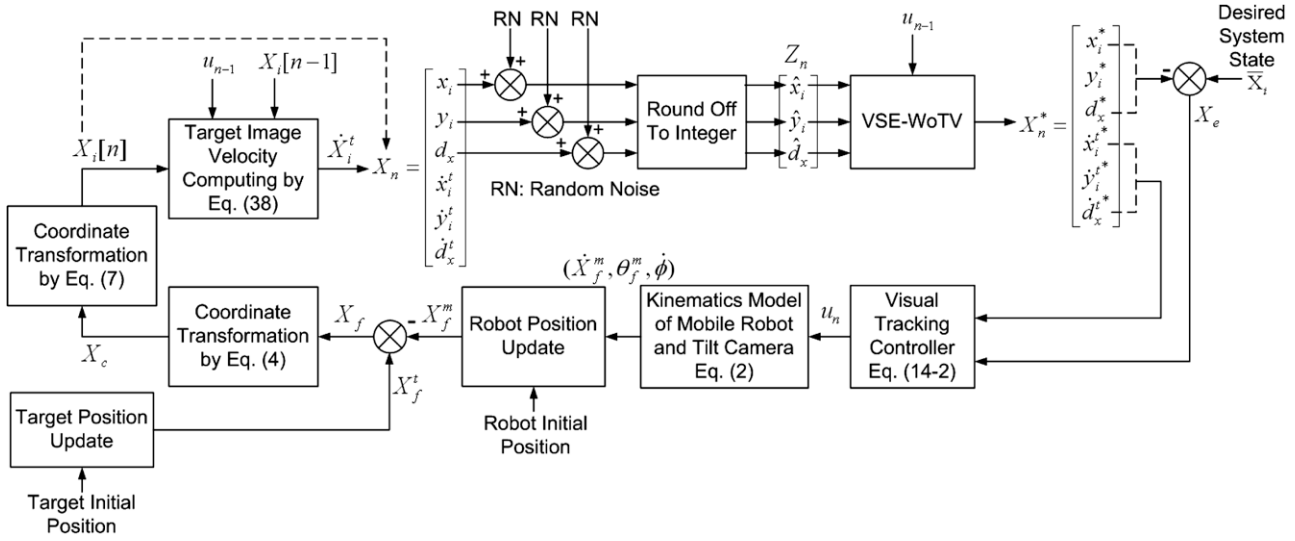

Fig. 5. Simulation setup for the performance evaluation of the proposed visual tracking control system.

Table 1
Parameters used in the simulations

Symbol	Quantity	Description
(f_x, f_y)	(294, 312) pixels	Camera focal length in retinal coordinates
W	12 cm	Width of the target
D	40 cm	Distance between two drive wheels
δy	10 cm	Distance between the center of robot tilt platform and the onboard camera
T	35 ms	Sampling period of the control system
$(\bar{x}_i, \bar{y}_i, \bar{d}_x)$	(0, 0, 35)	Desired system state in the image plane
$(\alpha_1, \alpha_2, \alpha_3)$	(5/4, 3, 1/2)	Positive control gains used in the experiments
$\mathbf{Q}_0$	diag(1, 1, 1, 4, 4, 4)	Initial covariance matrix
K_n	15	Noise gain
ρ	0.75	Constant threshold value

$$X_i^p[n] = X_i^*[n-1] + T\dot{X}_i^t[n-1] + T\mathbf{B}_i u_{n-1} \quad \text{for } n = 1, 2, \ldots \tag{31}$$

Suppose that the target motion is close to a smooth motion in a sampling period, then the target image velocity can be approximated as a constant velocity between two consecutive sample instants

$$\dot{X}_i^t[n] = \dot{X}_i^t[n-1]. \tag{32}$$

Based on (31) and (32), the propagation equation of VSE-WoTV is given by

$$X_n^p = \begin{bmatrix} \mathbf{I}_3 & T\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} X_{n-1}^* + \begin{bmatrix} T\mathbf{B}_i \\ \mathbf{0}_3 \end{bmatrix} u_{n-1} \equiv \mathbf{A}_{\text{est}} X_{n-1}^* + \mathbf{B}_{\text{est}} u_{n-1}, \tag{33}$$

where $(X_n^p)^{\mathrm{T}} = [\,(X_i^p[n])^{\mathrm{T}} \quad (\dot{X}_i^t[n])^{\mathrm{T}}\,]$ is the propagated system state at a sample instant n, $\mathbf{0}_3$ is a 3×3 zero matrix, and $(X_{n-1}^*)^{\mathrm{T}} = [\,(X_i^*[n-1])^{\mathrm{T}} \quad (\dot{X}_i^t[n-1])^{\mathrm{T}}\,]$ denotes the estimated system state at a sample instant $n-1$.

Next, the covariance matrix of propagation equation (33) at a sample instant n is given by

$$\mathbf{P}_n = \mathbf{A}_{\text{est}} \mathbf{P}_{n-1}^* \mathbf{A}_{\text{est}}^{\mathrm{T}} + \mathbf{Q}_{n-1}, \tag{34}$$

where $\mathbf{P}_{n-1}^*$ is the estimated covariance matrix at a sample instant $n-1$, and $\mathbf{Q}_n$ is the covariance matrix of the Gaussian propagation uncertainty. Note that (32) is an oversimplified assumption and will induce propagation error when the target motion is not smooth. However, this kind of error can be compensated by the observation information.

Remark 4. A major difference between VSE-WTV and VSE-WoTV is that the propagation covariance matrix of VSE-WoTV includes the covariance matrix of the Gaussian propagation uncertainty, $\mathbf{Q}_n$. The main reason is that if $V_{\mathrm{f}}^{\mathrm{t}}$ is known *a priori*, the prediction of the target state would be more precise with small uncertainty. Thus, the covariance matrix $\mathbf{Q}_n$ can be approximated by the matrix $T^2\mathbf{B}_i\mathbf{W}_{n-1}\mathbf{B}_i^{\mathrm{T}}$ in the propagation covariance matrix of VSE-WTV. On the other hand, if $V_{\mathrm{f}}^{\mathrm{t}}$ is unknown, the uncertainty of the target prediction state would become larger. Therefore, the propagation covariance matrix of VSE-WoTV should take the covariance matrix $\mathbf{Q}_n$ into account.

4.2.2. Observation and correction for VSE-WoTV

Since the observed image only contains information about the target status X_i in each sample instant, the observation model of the VSE-WoTV is given by

$$Z_n = [\,\mathbf{I}_3 \quad \mathbf{0}_3\,]X_n + \delta Z_n \equiv \mathbf{H}_{\text{est}} X_n + \delta Z_n, \tag{35}$$

where $\delta Z_n \sim N(0, \mathbf{R}_n)$ is the observation uncertainty with zero mean and covariance matrix $\mathbf{R}_n$. Based on Eq. (33) to (35), the optimal estimate and the corresponding covariance matrix at a sample instant n are given by

$$X_n^* = X_n^p + \mathbf{K}_n(Z_n - \mathbf{H}_{\text{est}} X_n^p) \quad \text{and} \quad \mathbf{P}_n^* = (\mathbf{I}_6 - \mathbf{K}_n\mathbf{H}_{\text{est}})\mathbf{P}_n, \tag{36}$$

where $\mathbf{K}_n = \mathbf{P}_n\mathbf{H}_{\text{est}}^{\mathrm{T}}(\mathbf{H}_{\text{est}}\mathbf{P}_n\mathbf{H}_{\text{est}}^{\mathrm{T}} + \mathbf{R}_n)^{-1}$ is the Kalman gain matrix, and $\mathbf{I}_6$ is a 6×6 identity matrix.

4.2.3. Summary of the proposed VSE-WToV algorithm

Combining the propagation equations (33) and (34) with the correction equation (36), the processing steps of VSE-WoTV are summarized as follows:

(1) Choose an initial covariance matrix $\mathbf{Q}_0$.
(2) Assume that the initial position of the target is located in the field-of-view of the camera, then initialize the estimated system state X_0^* and propagation covariance matrix $\mathbf{P}_0$ by the first observation such that $X_0^* = [\,Z_0^{\mathrm{T}} \quad 0 \quad 0 \quad 0\,]^{\mathrm{T}}$ and $\mathbf{P}_0 = \mathbf{I}_6$.
(3) Compute the ideal propagated state X_n^p using (33) and the corresponding propagation covariance matrix $\mathbf{P}_n$ using (34).
(4) If the target is detected in the observed image, then compute the Kalman gain matrix $\mathbf{K}_n$ and calculate the optimal estimate X_n^* with the corresponding covariance matrix $\mathbf{P}_n^*$ using (36); else set $X_n^* = X_n^p$ and $\mathbf{P}_n^* = \mathbf{P}_n$; go to step 5.

(5) Let $X^*_{n-1} = X^*_n$, $\mathbf{P}^*_{n-1} = \mathbf{P}^*_n$ and $\mathbf{Q}_{n-1} = \mathbf{Q}_0$; go to step 3.

5. Simulation and experimental results

Several interesting computer simulations and practical experiments are presented in this section to validate the tracking performance and robustness of the proposed visual tracking control system. First, MATLAB was used to verify the tracking performance of the proposed VTC and the estimation performance of the proposed VSE-WoTV. Next, two experiments were performed on an experimental mobile robot to validate the robustness against the occlusion uncertainty. Since the estimation without velocity information is more difficult to demonstrate compared with that of known velocity information, only the simulation results of the controller response of this part is presented. Practical experimental results of both VSE-WTV and VSE-WoTV are illustrated using a video clip and recorded photos.

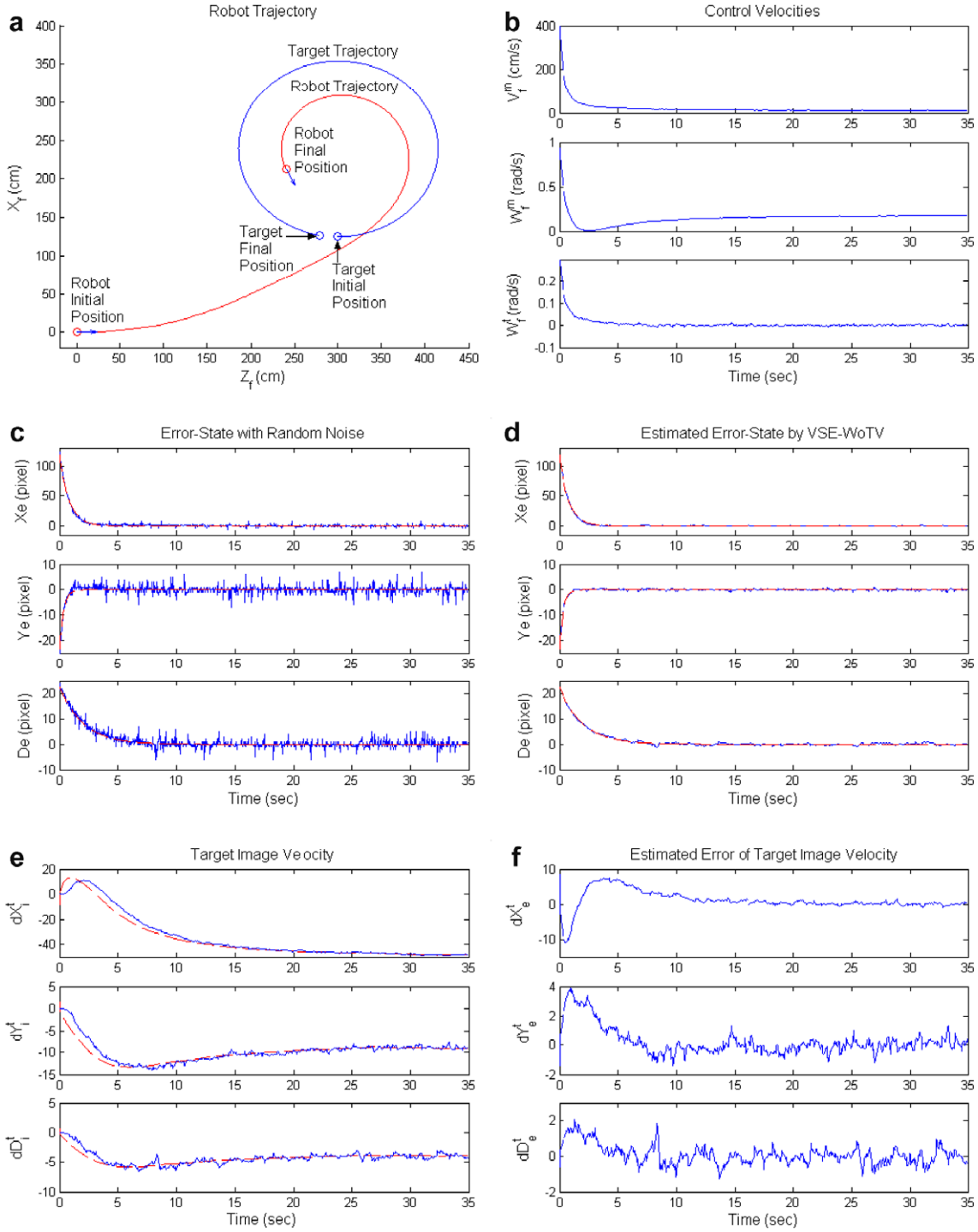

Fig. 6. The computer simulation results of the proposed VTC combined with VSE-WoTV. (a) Robot trajectory in the world coordinate frame. (b) Control velocities of the center point and the tilt camera of tracking robot. (c) Tracking errors with random noise. (d) Tracking errors estimated by VSE-WoTV. (e) Estimated target image velocity. (f) Estimation errors.

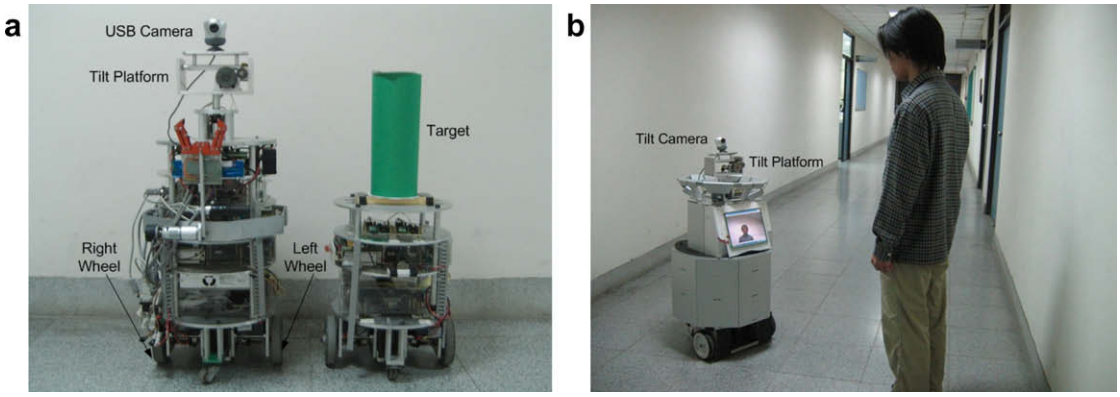

Fig. 7. Experimental mobile robots used to test the tracking performance of the proposed visual tracking control system. (a) Experimental robots used in experiment 1. The left robot is called *tracking robot*, and the right one is called *target robot*. (b) Experimental robot used in experiment 2.

Table 2
Parameters used in the experiments

Symbol	Quantity		Description
	Experiment 1	Experiment 2	
(f_x, f_y)	(294,312) pixels	(393.4,391.8) pixels	Camera focal length in retinal coordinates
W	12 cm	12 cm	Width of the target
D	30 cm	40 cm	Distance between two drive wheels
δy	10 cm	10 cm	Distance between the center of robot tilt platform and the onboard camera
T	100 ms	100 ms	Sampling period of the control system
$(\bar{x}_i, \bar{y}_i, \bar{d}_x)$	(0,0,35)	(0,0,35)	Desired system state in the image plane
$(\alpha_1, \alpha_2, \alpha_3)$	(5/16,6/8, 4/16)	(5/4,3,1/2)	Positive control gains used in the experiments
$\mathbf{Q}_0$	x	diag(5,5,5, 20,20,20)	Initial covariance matrix

x, Do not care.

5.1. Simulation results

In order to evaluate the performance of the proposed visual tracking control system, a simulation environment has been setup using MATLAB. Fig. 5 shows the architecture of the simulation setup. In Fig. 5, X_n,which includes the target state $X_i[n]$ and target image velocity $\dot{X}_i^t$, denotes the ideal state needed to be estimated by the VSE-WoTV. $X_i[n]$ is obtained from the coordinate transformations (4) and (7), and $\dot{X}_i^t$ is calculated by (11) such that

$$\dot{X}_i^t = \dot{X}_i - \dot{X}_i^m = \frac{X_i[n] - X_i[n-1]}{T} - \mathbf{B}_i u_{n-1}. \tag{37}$$

The observation signal Z_n is obtained by the rounding off the value of $X_i[n]$ with random noise (RN) to an integer. In this paper, the random noise is given by

$$RN = \begin{cases} K_n \sigma_1 (0.5 - \sigma_2) & \text{if } (\sigma_3 < \rho), \\ (1 + \sigma_1)(0.5 - \sigma_2) & \text{otherwise}, \end{cases} \tag{38}$$

where $K_n > 1$ is the noise gain, $\sigma_i \in [0,1]$, $i = 1$–3, are three random signals with uniform distribution, and $\rho \in [0,1]$ is a constant threshold value. Expression (38) indicates that the intensity of the noise is time-varying and dependent on a random condition. If the condition $(\sigma_3 < \rho)$ is satisfied, then the random noise will have large noise gain; otherwise the random noise will only have noise gain smaller than 2. Thus, the threshold value ρ determines the probability of appearing large observation noise. This kind of random noise usually occurs during practical visual tracking process of the mobile robot, since the intensity of the observation uncertainty usually is position-dependent and light-dependent. The parameters used in the simulations are listed in Table 1.

Fig. 6 presents the computer simulation results of the proposed visual tracking control system. Fig. 6(a) shows the robot trajectory in the world coordinate frame. In the simulation, the motion of the target is set as a circular path with velocity

$$(v_f^x, v_f^y, v_f^z) = (v_f^t \sin\theta_f^t, 0, v_f^t \cos\theta_f^t),$$

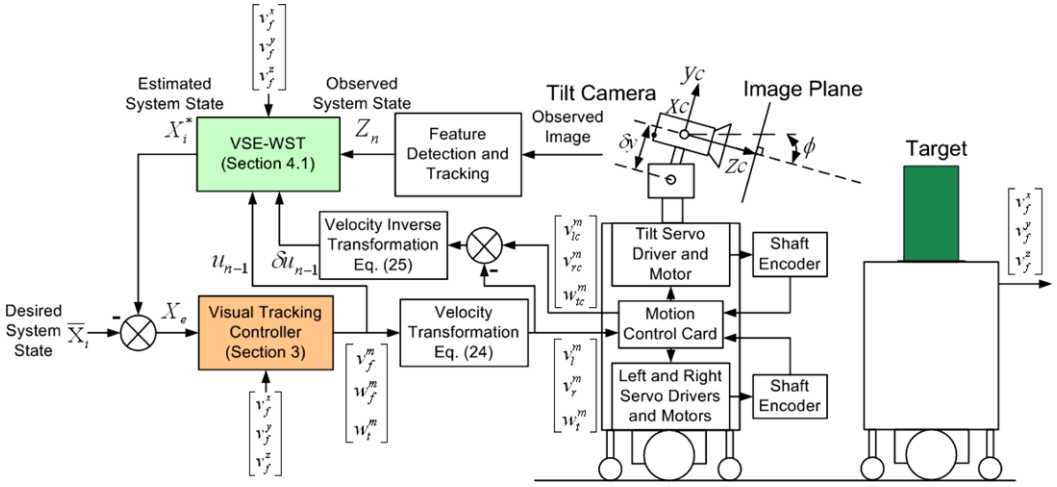

Fig. 8. Block diagram of the visual tracking control system tested in the first experiment.

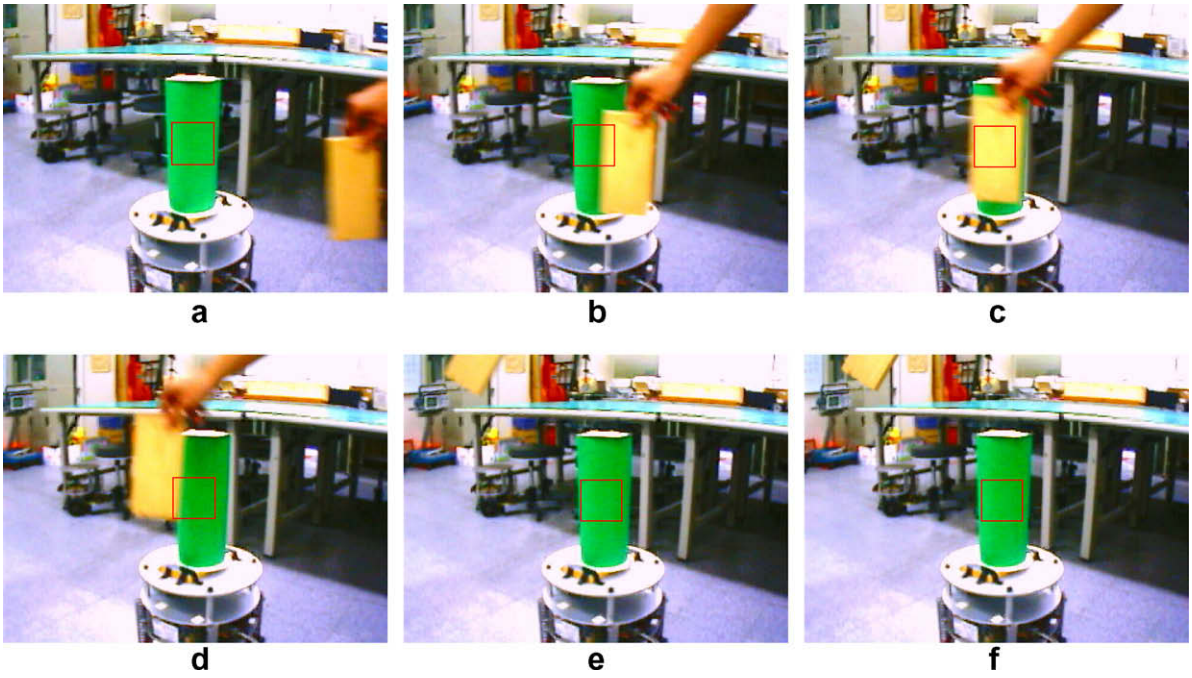

Fig. 9. Experimental results of tracking a moving target when it is temporarily partially occluded. (a) Before partial occlusion. (b)–(d) Partial occlusion occurred. (e)–(f) After partial occlusion, the moving target is still under tracking.

where $v_f^t = 20$ cm/s and $\theta_f^t(\text{new}) = \theta_f^t(\text{old}) + (T\pi/18)$ rad with $\theta_f^t(0) = 0$. From Fig. 6(a), we see that the motion trajectory of the tracking robot is also a circular path as a result of following the target. Fig. 6(b) shows the control velocities of the robot center point and the tilt camera. Simulation results reveal that the linear and angular velocities of tracking robot converge to constant values when the tracking errors decay to zero. Therefore, the tracking robot keeps tracking the target continuously. Fig. 6(c) shows the tracking errors with random noise (39), and Fig. 6(d) is the corresponding tracking errors estimated by the VSE-WoTV. In Fig. 6(c) and (d), the dotted lines illustrate the ideal tracking errors while the solid lines show the observation and estimation results of tracking errors. A comparison of Fig. 6(c) with Fig. 6(d) shows that the random noise in each error state is removed sufficiently, especially the error states y_e and d_e. Thus, the robustness of the proposed VSE-WoTV against the random noise uncertainty is verified. Moreover, in Fig. 6(d), each error state converges to zero exponentially, which validates the tracking performance of the proposed VTC. Fig. 6(e)

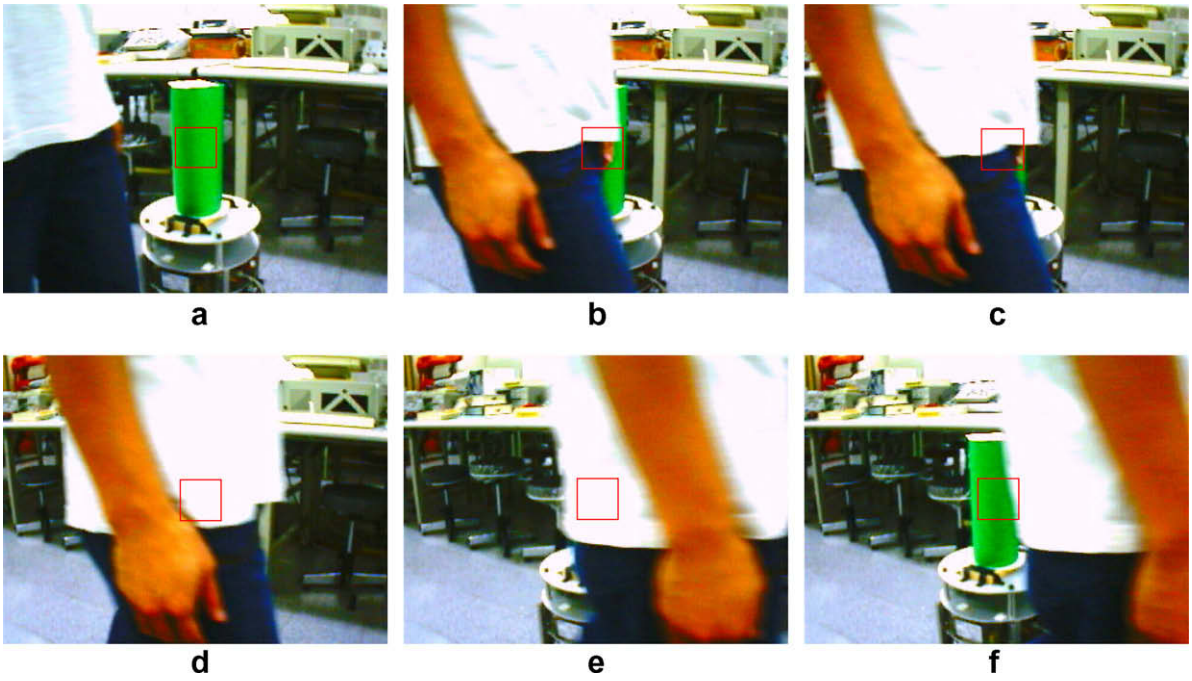

Fig. 10. Experimental results of tracking a moving target when it is temporarily fully occluded. (a) Before full occlusion. (b)–(e) Full occlusion occurred. The moving target is estimated only using the prediction information. (f) After fully occlusion, the moving target is still under tracking.

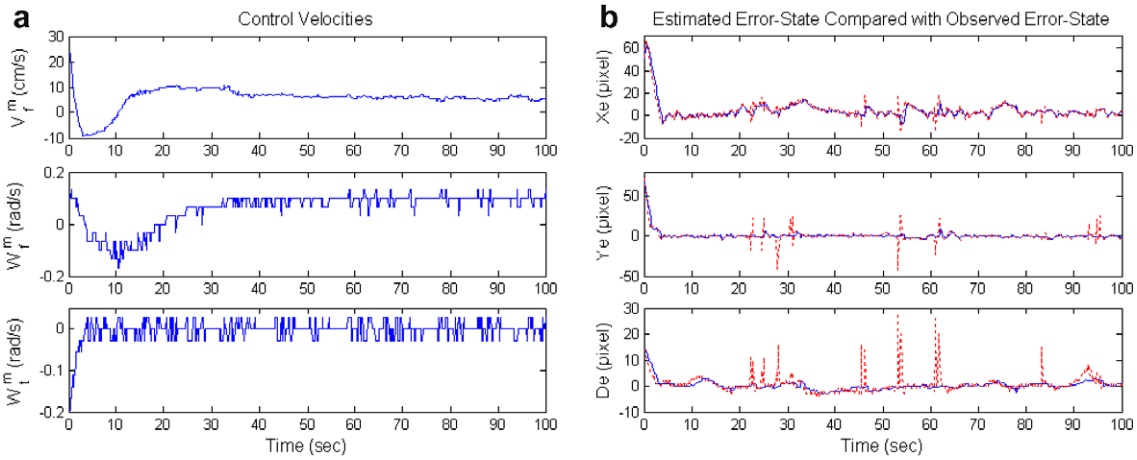

Fig. 11. Experimental results of the proposed VTC combined with the VSE-WTV. (a) Command velocities of mobile robot and tilt camera. (b) Estimated (solid lines) and observed (dotted lines with spikes) tracking errors.

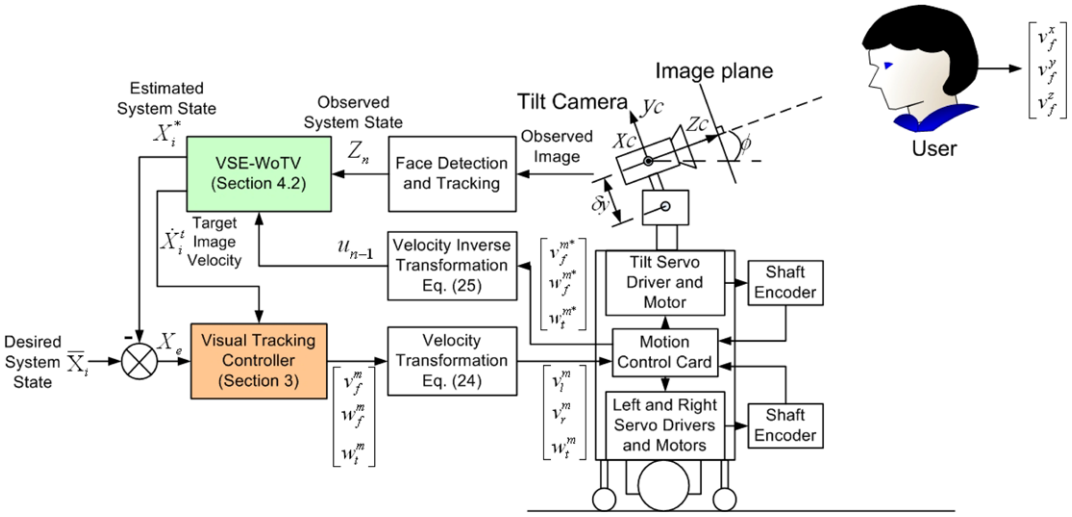

Fig. 12. Block diagram of the visual tracking control system used in the second experiment.

and (f), respectively, presents the estimation results and the estimation errors of target image velocity from the VSE-WoTV. In Fig. 6(e), the dotted lines indicate the ideal target image velocity while the solid lines show the estimation results of target image velocity. It is clear that each estimate converges to the corresponding ideal value. This result also can be seen in Fig. 6(f), which shows that each estimation error converges to zero efficiently. Therefore, these simulation results validate the estimation performance of the proposed VSE-WoTV.

5.2. Experimental results

Two experiments have been carried out using an experimental mobile robot to validate the performance of the proposed control scheme: the first experiment is to track a moving object, and the second one is to track a moving person. Fig. 7(a) and (b) shows the mobile robots of experiments 1 and 2, respectively. In Fig. 7(a), the left robot (called *tracking robot*) is equipped with a USB camera and a tilt camera platform to track another robot, on which a cylindrical object of interest was installed (called *target robot*). Fig. 7(b) shows another experimental robot constructed to serve as a test bed for the study of visual tracking of a moving target without its velocity information. Table 2 tabulates the parameters used for the tracking robot in the experiments. Note that the processing time of the proposed visual tracking control system (including target detection, estimation and control computations) is less than 50 ms. This means that the overall tracking system is of acceptable computational load and can track the target in real time. However, the sampling period of the control system, T, was set to 100 ms in the experiments due to other image processing computations such as image compression and storage. In the following, the experimental results of visual tracking with occlusion are presented to validate the system performance and robustness.

5.2.1. Experiment 1: Visual tracking of a moving object

This section presents the experimental results of tracking a moving target with the target temporarily partially and fully occluded. Fig. 8 shows the block diagram of the implemented visual tracking control system in experiment 1. In this experiment, we set another robot as the moving target with *a priori* known motion velocity in order to verify the performance of the VSE-WTV proposed in Section 4.1. A cylindrical object was placed on the robot

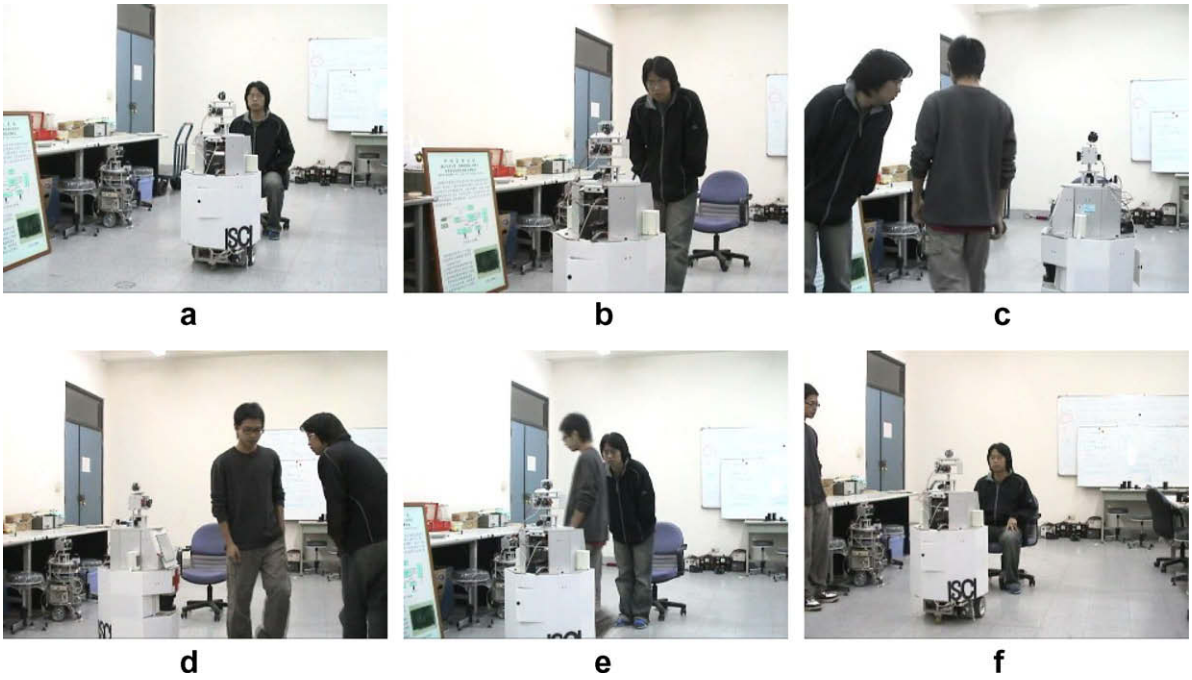

Fig. 13. Experimental results of tracking a moving person. (a)–(f) Image sequence recorded from a DV camera. (a)–(b) The tracking robot started to track the user. (c)–(e) Full occlusion occurred when another person walked across temporarily. (f) The tracking robot still tracked the user after full occlusion.

to facilitate easy recognition of the target, which is moving along a circular path with velocity

$$(v_f^x, v_f^y, v_f^z) = (v_f^t \sin\theta_f^t, 0, v_f^t \cos\theta_f^t),$$

where $v_f^t = 10.5$ cm/s and $\theta_f^t(\text{new}) = \theta_f^t(\text{old}) + 0.01$ rad with $\theta_f^t(0) = \pi$. The information on the target velocity is then used in VSE-WTV to estimate the state of the target and overcome the occlusion problem even if the target is temporarily fully occluded.

Fig. 9 illustrates the partial occlusion experiment recorded by the tilt camera on-board the tracking robot (the robot with a camera). Fig. 9(a) shows the tracked target before partial occlusion. In Fig. 9(b)–(d), the moving target is temporarily partially occluded by a moving object. Fig. 9(e) and (f) shows that the moving target is still tracked after partial occlusion. In Fig. 10, the target was fully blocked by a moving person. Fig. 10(a) shows the tracked target before full occlusion. In Fig. 10(b)–(e), the moving target is temporarily fully occluded by a passing person. Since the target would not be observable in the observed image, the VSE-WTV estimated the moving target only using prediction information. Hence, the moving target is still tracked even though it is unobservable. Fig. 10(f)

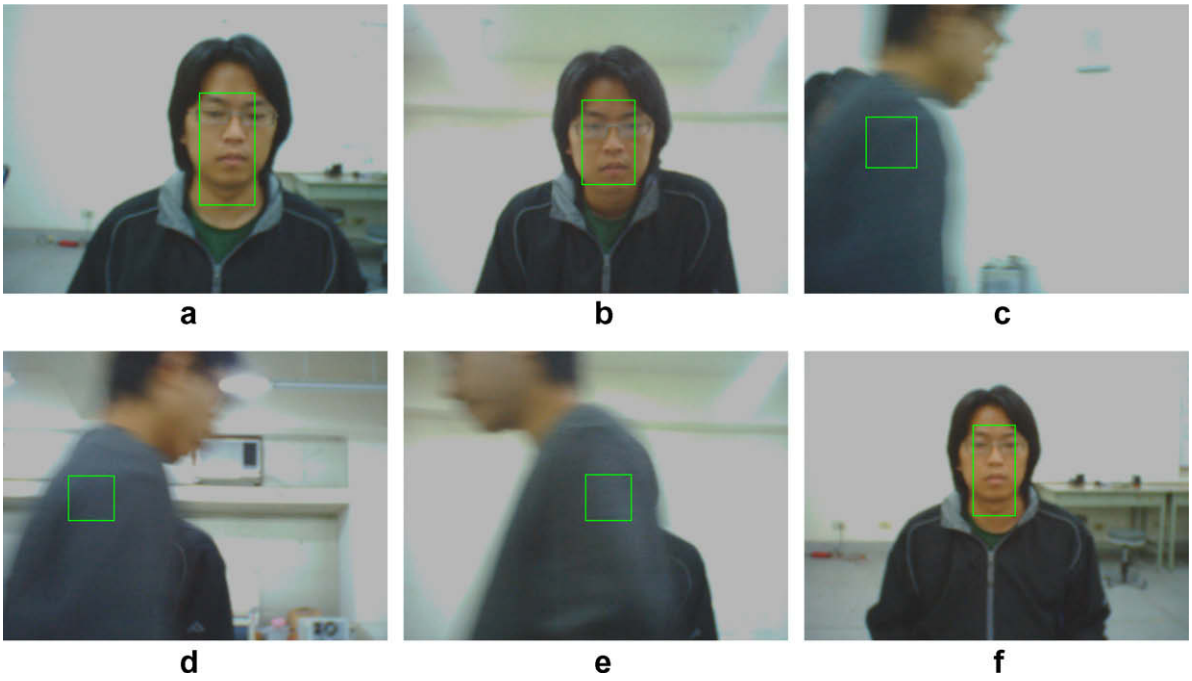

Fig. 14. Experimental results of tracking a moving person. (a)–(f) Image sequence recorded from on-board USB camera. (a)–(b) The tracking robot started to track the user. (c)–(e) Full occlusion occurred when another person walked across temporarily. (f) The tracking robot still tracked the user after full occlusion.

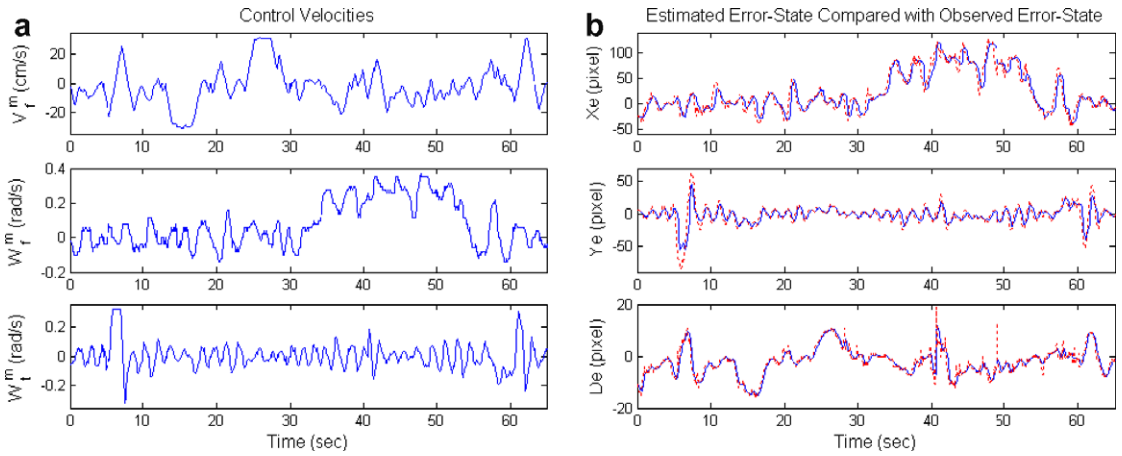

Fig. 15. Experimental results of the proposed VTC combined with the VSE-WoTV. (a) Command velocities of the mobile robot and the tilt camera. (b) Estimated (solid lines) and observed (dotted lines with spikes) tracking errors.

shows that the moving target is tracked successfully after full occlusion.

Fig. 11 presents the recorded experimental results of tracking a moving object. Fig. 11(a) shows the control velocities of both mobile robot and tilt camera. Fig. 11(b) shows a comparison between the estimated error states (the solid lines) and the observed ones (the dotted lines with spikes). From Fig. 11(b), we see that the random noise caused by the temporary occlusion effect is removed efficiently by utilizing the proposed VSE-WTV. Therefore, based on the above occlusion experiments, the robust estimation performance of the proposed visual tracking control system is verified. A video clip of the experimental results is available online [28].

5.2.2. Experiment 2: Visual tracking of a moving person

In this section, the tracking performance of the proposed VTC combined with the VSE-WoTV is demonstrated by tracking a moving person. In order to detect and track the person in the image plane, the visual tracking control system is combined with a real-time face detection and tracking algorithm presented in our previous work [29]. Fig. 12 illustrates the complete visual tracking system which encompasses the face detection/tracking algorithm, the VSE-WoTV described in Section 4.2 and the VTC presented in Section 3. Because the velocity of human motion is unknown, the VSE-WoTV works to estimate the image velocity instead of the motion velocity for the VTC used and thus overcome the temporary occlusion problem.

Figs. 13 and 14 show the recorded images of the mobile robot interacting with the moving person in the experiment. Fig. 13(a)–(f) show the recorded photos of the experimental scenario, and Fig. 14(a)–(f) are the corresponding pictures recorded by the on-board USB camera. In the beginning, the person sat on a stool, and the robot started to track his face using the proposed visual tracking control system (Figs. 13(a) and 14(a)). Next, the person stood up to walk around in the room, and the mobile robot kept following and tracking the person's face by the tilt camera (Figs. 13(b) and 14(b)). When the person was walking, another person passed between the tracked person and the robot temporarily (Fig. 13(c)–(e)). Thus, in Fig. 14(c)–(e), the person's face was temporarily fully blocked by the passing person. Based on the proposed VSE-WoTV algorithm, the propagation information will dominate the estimation results in this situation even if the target is fully unobservable. Therefore, the VSE-WoTV still estimated the positions and velocities of the person's face in the image plane successfully even during the temporary full occlusion conditions. Finally,

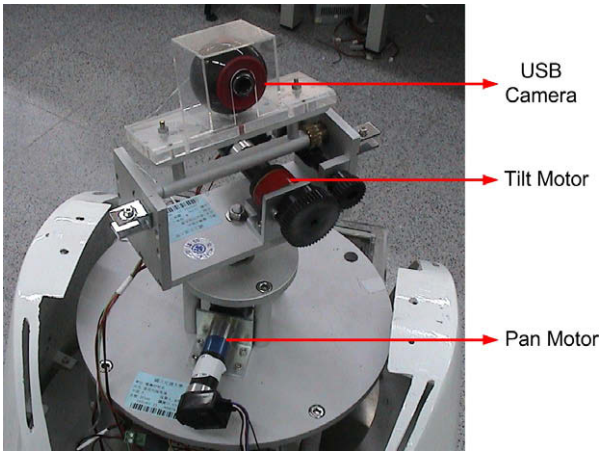

Fig. 16. Experimental pan-tile platform used to demonstrate the robust property of the proposed visual tracking scheme.

the person sat down on the stool, and the robot tracked him continuously.

Fig. 15 presents the recorded experimental results of tracking a moving person. Fig. 15(a) shows the control velocities of both mobile robot and tilt camera. Fig. 15(b) compares the estimated tracking errors (solid lines) with the observed ones (dotted lines with spikes). From Fig. 15(b), it is clear that the random noise is also removed efficiently by the proposed VSE-WoTV algorithm. This occlusion experiment validates the robust estimation performance of the proposed visual tracking control system. A video clip of this experiment is available online in [28].

Remark 5. The main differences between the proposed method and the existing video color object tracking (VCOT) methods, such as CamShift algorithm [30], are twofold. First, the existing VCOT methods usually suppose that the target has located in the camera's field of view and do not consider the camera motion effect. On the contrary, the proposed method considers both camera and target motion effects to increase the tracking performance and system robustness. Second, the existing VCOT methods usually do not deal with the temporary full occlusion problem. In contrast, the proposed method uses the propagation information to deal with the temporary full occlusion problem. Moreover, the propagation covariance matrix can be used to evaluate the reliability of the tracking state under the situation of full occlusion.

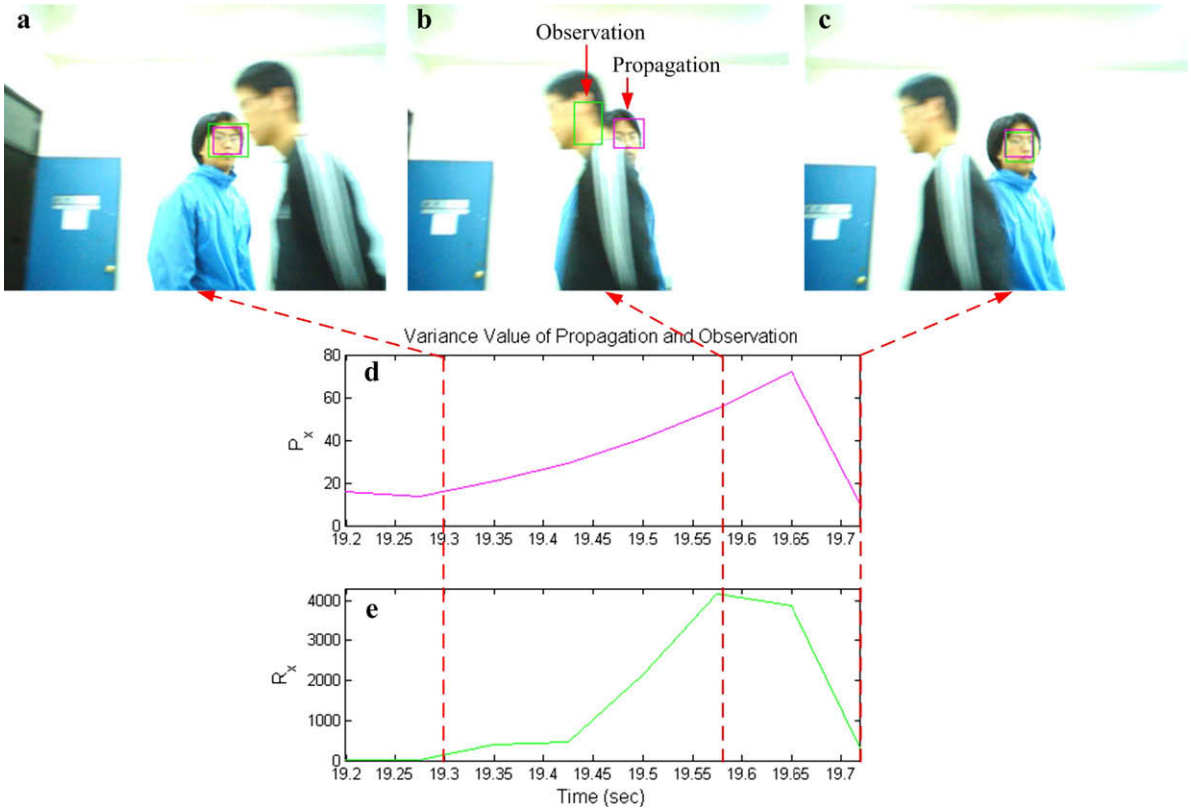

Fig. 17. The experimental results of occlusion using VSE-WoTV. (a)–(c) Recorded camera view, observation states and propagation states, (d) variance of propagation states, (e) variance of observation states.

For example, if one of the diagonal values of the propagation covariance matrix is larger than a preset threshold, then the propagation is not reliable, and thus the visual tracking control system should be stopped and reinitialized.

5.2.3. Experiment of occlusion-robustness property

Since the current VSE design is based on the Kalman filter algorithm, the estimation performance is dependent on the accuracy of covariance matrices $\mathbf{P}_n$ and $\mathbf{R}_n$. In order to demonstrate this property, the proposed visual tracking control system is extended to control a pan-tile camera platform in this experiment. Fig. 16 shows the experimental pan-tile platform equipped with a camera to track the face of a user. The control velocities of pan-tile platform can be computed by simplifying the proposed control law (14) such that

$$\begin{bmatrix} w_f^{pan} \\ w_f^{tilt} \end{bmatrix} = \begin{bmatrix} B_{12} & B_{13} \\ B_{22} & B_{23} \end{bmatrix}^{-1} \begin{bmatrix} \alpha_1 x_e - \dot{x}_i^t \\ \alpha_2 y_e - \dot{y}_i^t \end{bmatrix}, \tag{39}$$

where w_f^{pan} is the pan control velocity, w_f^{tilt} is the tilt control velocity, and B_{mn} denotes an element of matrix $\mathbf{B}_i$ corresponding to the mth row and nth column.

Fig. 17 presents the experimental results. Fig. 17(a)–(c) shows the recorded images, in which the green and magenta windows indicate the observation and propagation, respectively. Fig. 17(d) and (e), respectively, illustrates the variance value of state x_i in propagation and observation covariance matrices. Since the face tracking algorithm employed in the current system only uses the skin color to detect the human face in a local search window, the algorithm will track another person's face which moves across the user's face and camera. In this situation, the variance value of observation covariance matrix will increase greatly due to the rapid change in the observation. Thus, in Fig. 17(e), we see that the variance value of the observed state x_i (denoted by R_x) increases rapidly due to the sudden change in observation. On the other hand, the variance value of the propagated state x_i (denoted by P_x) increases smoothly but much smaller than the observed one. Therefore, after the correction step in Kalman filtering algorithm, the propagation state will dominate the estimation result, which tracks the correct user's face. Finally, in Fig. 17(c), the face tracking

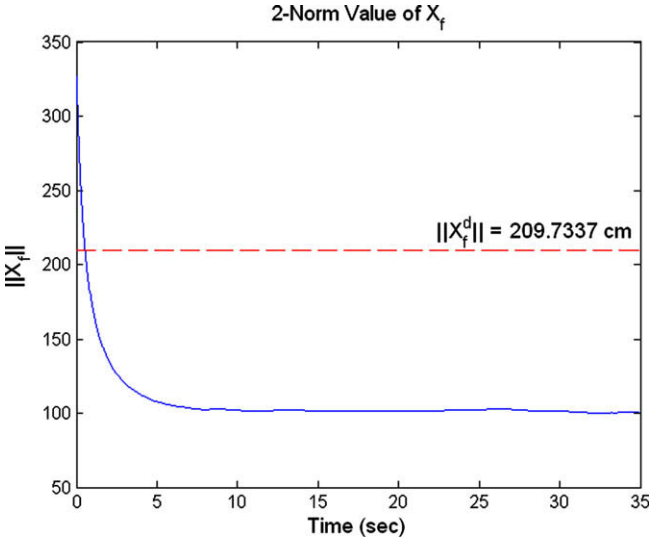

Fig. 18. Simulation result of the distance between mobile robot and motion target, $||X_f||$.

algorithm detects the human face close to the estimation result, and the observation is corrected. A video clip of this experiment is available online in [28].

Remark 6. If there is an object with similar feature and motion to target, then the proposed VSE may track to this object when it moves across the target and camera. However, this problem can be resolved by combining an object recognition algorithm with visual tracking algorithm. In this paper, we do not cover the object recognition problem and only focus the topic on visual tracking problem.

6. Conclusion and future work

A novel visual tracking control model of visual interaction between a mobile robot and a dynamic moving object in the image plane has been derived in this paper. Based on the control model, a tracking controller is proposed to resolve visual tracking control of a dynamic moving target with asymptotical convergence. Based on the proposed visual interaction model, two visual state estimators have been proposed to estimate the optimal system state from the noisy observation and temporary occlusion during visual tracking operation. The merit of this design is that the image processing procedures are much simplified for the desired motion control in the world coordinates due to image-based computation. Simulation and experimental results of tracking a moving target validate the performance and robustness of the proposed control schemes.

Since the performance of the current VSE design has some restrictions due to the assumptions of the Kalman filter (e.g. Gaussian distribution uncertainty, smoothness motion, and uniform sampling rate), the future work will focus on developing other types of VSE, such as neural-networks based VSE, to solve this problem and improve the accuracy of the visual estimation results.

Acknowledgements

The authors thank Fu-Sheng Huang, Chen-Yang Lin, and Chun-Wei Chen for their assistance in the experiments. This work was supported by the Ministry of Economic Affairs under Grant 95-EC-17-A-04-S1-054 and the National Science Council of Taiwan, ROC under Grant NSC 95-2218-E-009-024.

Appendix

In this appendix, we will show that when the system state X_i converges to the desired system state $\bar{X}_i$, the mobile robot follows the moving target. Recall the diffeomorphism defined in (7)

$$X_i = \mathbf{P}_c^i X_c, \tag{A1}$$

where $\mathbf{P}_c^i = \mathrm{diag}(-k_x, k_y, k_x k_w)$. Suppose that the system state X_i has converged to the desired system state $\bar{X}_i$, we then have the following results based on (A1)

$$\bar{X}_i = \mathbf{P}_c^i X_c^d \tag{A2}$$

and

$$X_c^d = [\,x_c^d \quad y_c^d \quad z_c^d\,]^T = \mathbf{R}(\phi, \theta_f^m) X_f^d - \delta Y, \tag{A3}$$

where X_c^d and X_f^d, respectively, are the related position between mobile robot and motion target in camera and world coordinate frame when $X_i = \bar{X}_i$. Because $\mathbf{P}_c^i$ is invertible, the following relation between X_f^d and $\bar{X}_i$ can be obtained by substituting (A3) into (A2) such that

$$X_f^d = \mathbf{R}^T(\phi, \theta_f^m)[(\mathbf{P}_c^i)^{-1}\bar{X}_i + \delta Y]. \tag{A4}$$

Let $\|A\|$ denote the 2-norm value of vector or matrix A. The key idea is that if $\|X_f^d\|$ is bounded, it implies that the mobile robot has followed the motion target. Using (A4), $\|X_f^d\|$ is given by

$$\begin{aligned}\|X_f^d\| = \|\mathbf{R}^T(\phi, \theta_f^m)[(\mathbf{P}_c^i)^{-1}\bar{X}_i + \delta Y]\| &\leqslant \|\mathbf{R}^T(\phi, \theta_f^m)\| \cdot \|(\mathbf{P}_c^i)^{-1}\bar{X}_i + \delta Y\| \\ &\leqslant \|(\mathbf{P}_c^i)^{-1}\bar{X}_i\| + \|\delta Y\| \\ &\leqslant \|(\mathbf{P}_c^i)^{-1}\| \cdot \|\bar{X}_i\| + \|\delta Y\|. \end{aligned} \tag{A5}$$

Because of $\|(\mathbf{P}_c^i)^{-1}\| = \|\mathrm{diag}(-k_x^{-1}, k_y^{-1}, k_x^{-1}k_w^{-1})\| = z_c\|\mathrm{diag}(-f_x^{-1}, f_y^{-1}, f_x^{-1}W^{-1})\| = z_c\sqrt{\lambda_{\max}}$, where $z_c = f_x W/d_x$ and $\lambda_{\max} = \max(f_y^{-1}, f_x^{-1}W^{-1})$, we have the following result:

$$\|X_f^d\| \leqslant \frac{f_x W}{\bar{d}_x}\sqrt{\lambda_{\max}}\|\bar{X}_i\| + \|\delta Y\|. \tag{A6}$$

From (A6), it is clear that $\|X_f^d\|$ is bounded, and hence the proof is completed.

We use the simulation presented in Section 5.1 as an example to explain the physical meaning of (A6). By using the parameters listed in Table 1, the term on the right-hand side of (A6) can be calculated by

$$\|X_f^d\| \leqslant \frac{294 \times 12}{35}\sqrt{0.0032} \times 35 + 10 = 209.7337 \quad (\mathrm{cm}), \tag{A7}$$

which means that when X_i converges to $\bar{X}_i$, the distance between mobile robot and motion target is bounded to 209.7337 cm. Fig. 18 shows the simulation result of the distance between mobile robot and motion target. In Fig. 18, the solid line presents the 2-norm value of X_f, and the dotted line denotes the bounded distance calculated in (A7). From Fig. 18, we see that the distance between mobile robot and target finally converges to about 100 cm, which is satisfied in the bounded condition (A7). Because the target is always moving and the distance between the robot and target is bounded, this implies that the robot has followed the target as we expected.

References

[1] S. Hutchinson, G.D. Hager, P.I. Corke, A tutorial on visual servo control, IEEE Transactions Robotics and Automation 12 (5) (1996) 651–670.
[2] F. Chaumette, S. Hutchinson, Visual servo control part I: basic approaches, IEEE Robotics and Automation Magazine 13 (4) (2006) 82–90.
[3] F. Chaumette, S. Hutchinson, Visual servo control part II: advanced approaches, IEEE Robotics and Automation Magazine 14 (1) (2007) 109–118.
[4] Y. Ma, J. Košecká, S.S. Sastry, Vision guided navigation for a nonholonomic mobile robot, IEEE Transactions on Robotics and Automation 15 (3) (1999) 521–536.
[5] J.-B. Coulaud, G. Campion, G. Bastin, M.D. Wan, Stability analysis of a vision-based control design for an autonomous mobile robot, IEEE Transactions on Robotics 22 (5) (2006) 1062–1069.
[6] H. Zhang, J.P. Ostrowski, Visual motion planning for mobile robots, IEEE Transactions on Robotics and Automation 18 (2) (2002) 199–208.
[7] T. Nierobisch, W. Fischer, F. Hoffmann, Large view visual servoing of a mobile robot with a pan-tilt camera, in: Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 2006, pp. 3307–3312.
[8] J. Chen, W.E. Dixon, D.M. Dawson, M. McIntyre, Homography-based visual servo tracking control of a wheeled mobile robot, IEEE Transactions on Robotics 22 (2) (2006) 407–416.
[9] Y. Fang, W.E. Dixon, D.M. Dawson, P. Chawda, Homography-based visual servo regulation of mobile robots, IEEE Transactions on System, Man, and Cybernetics-Part B: Cybernetics 35 (5) (2005) 1041–1049.
[10] G.L. Mariottini, G. Oriolo, D. Prattichizzo, Image-based visual servoing for nonholonomic mobile robots using epipolar geometry, IEEE Transactions on Robotics 23 (1) (2007) 87–100.
[11] G.L. Mariottini, D. Prattichizzo, G. Oriolo, Image-based visual servoing for nonholonomic mobile robots with central catadioptric camera, in: Proceedings IEEE International Conference on Robotics and Automation, Orlando, FL, USA, 2006, pp. 538–544.
[12] G. López-Nicolás, C. Sagüés, J.J. Guerrero, D. Kragic, P. Jensfelt, Switching visual control based on epipoles for mobile robots, Journal of Robotics and Autonomous Systems 56 (7) (2008) 592–603.
[13] A.K. Das, R. Fierro, V. Kumar, J.P. Ostrowski, J. Spletzer, C.J. Taylor, A vision-based formation control framework, IEEE Transactions on Robotics and Automation 18 (5) (2002) 813–825.
[14] R. Vidal, O. Shakernia, S. Sastry, Following the flock [formation control], IEEE Robotics and Automation Magazine 11 (4) (2004) 14–20.

[15] J.A. Borgstadt, N.J. Ferrier, Interception of a projectile using a human vision-based strategy, in: Proceedings IEEE International Conference on Robotics and Automation, San Francisco, USA, 2000, pp. 3189–3196.
[16] L. Freda, G. Oriolo, Vision-based interception of a moving target with a nonholonomic mobile robot, Journal of Robotics and Autonomous Systems 55 (6) (2007) 419–432.
[17] H.Y. Wang, S. Itani, T. Fukao, N. Adachi, Image-based visual adaptive tracking control of nonholonomic mobile robots, in: Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, Maui, Hawaii, USA, 2001, pp. 1–6.
[18] C.-Y. Tsai, K.-T. Song, Face tracking interaction control of a nonholonomic mobile robot, in: Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 2006, pp. 3319–3324.
[19] E. Malis, S. Benhimane, A unified approach to visual tracking and servoing, Journal of Robotics and Autonomous Systems 52 (1) (2005) 39–52.
[20] A.I. Comport, É. Marchand, F. Chaumette, Statistically robust 2-D visual servoing, IEEE Transactions on Robotics and Automation 22 (2) (2006) 416–421.
[21] Y. Han, H. Hahn, Visual tracking of a moving target using active contour based SSD algorithm, Journal of Robotics and Autonomous Systems 53 (3–4) (2005) 265–281.
[22] J.D. Schutter, J.D. Geeter, T. Lefebvre, H. Bruyninckx, Kalman filters: a tutorial, Journal A 40 (4) (1999) 52–59.
[23] C.-Y. Tsai, K.-T. Song, Visual tracking control of a wheeled mobile robot with a system model and velocity quantization robustness, IEEE Transactions on Control Sysrems Technology, accepted for publication.
[24] W.I. Grosky, L.A. Tamburino, A unified approach to the linear camera calibration problem, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (7) (1990) 663–671.
[25] T.-C. Lee, C.-Y. Tsai, K.-T. Song, Fast parking control of mobile robots: a motion planning approach with experimental validation, IEEE Transactions on Control Systems Technology 12 (5) (2004) 661–676.
[26] J.-J.E. Slotine, W. Li, Applied Nonlinear Control, Prentice-Hall, Englewood Cliffs, NJ, 1991.
[27] C.-Y. Tsai, K.-T. Song, X. Dutoit, H. Van Brussel, M. Nuttin, Robust mobile robot visual tracking control system using self-tuning Kalman filter, in: Proceedings IEEE International Symposium on Computational Intelligence in Robotics and Automation, Jacksonville, Florida, 2007, pp. 161–166.
[28] The video website. Available: <http://isci.cn.nctu.edu.tw/video/RVTCS_IVC/>.
[29] K.-T. Song, J.-S. Hu, C.-Y. Tsai, C.-M. Chou, C.-C. Cheng, W.-H. Liu, C.-H. Yang, Speaker attention system for mobile robots using microphone array and face tracking, in: Proceedings IEEE International Conference on Robotics and Automation, Orlando, FL, 2006, pp. 3624–3629.
[30] G.R. Bradski, S. Clara, Computer vision face tracking for use in a perceptual user interface, Intel Technology Journal 2 (2) (1998) 1–15.