Robotics and Autonomous Systems 57 (2009) 652–664

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Robust visual tracking control system of a mobile robot based on a dual-Jacobian visual interaction model

Chi-Yi Tsai [a], Kai-Tai Song [a,*], Xavier Dutoit [b], Hendrik Van Brussel [b], Marnix Nuttin [b]

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

[b] Department of Mechanical Engineering, Division PMA, K. U. Leuven, Celestijnenlaan 300B, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 26 March 2007
Received in revised form
6 January 2009
Accepted 7 January 2009
Available online 21 January 2009

Keywords:
Visual tracking control
Visual estimation
Visual interaction model
Kalman filter

ABSTRACT

This paper presents a novel design of a robust visual tracking control system, which consists of a visual tracking controller and a visual state estimator. This system facilitates human–robot interaction of a unicycle-modeled mobile robot equipped with a tilt camera. Based on a novel dual-Jacobian visual interaction model, a robust visual tracking controller is proposed to track a dynamic moving target. The proposed controller not only possesses some degree of robustness against the system model uncertainties, but also tracks the target without its 3D velocity information. The visual state estimator aims to estimate the optimal system state and target image velocity, which is used by the visual tracking controller. To achieve this, a self-tuning Kalman filter is proposed to estimate interesting parameters and to overcome the temporary occlusion problem. Furthermore, because the proposed method is fully working in the image space, the computational complexity and the sensor/camera modeling errors can be reduced. Experimental results validate the effectiveness of the proposed method, in terms of tracking performance, system convergence, and robustness.

1. Introduction

An intelligent robot uses its on-board sensors to collect information from the surroundings and react to the changes of its immediate environment. In recent years, vision systems have been widely used for various intelligent robots, and the research on visual tracking control has gained increasing attention in the area of robotic research [1–25]. In robotics, visual tracking control means vision-based robot motion control to track an interesting target. Based on the motion constraints of the robot, visual tracking control can be classified into two categories: visual servoing for holonomic manipulators and visual tracking for nonholonomic mobile robots. Although visual servoing of holonomic manipulators has been discussed extensively and many results can be found in the literature [1–3], mobile robots are commonly nonholonomic and the visual servoing results for holonomic manipulators are unsuitable for a mobile platform [4].

This paper addresses the problem of visual tracking control of unicycle-modeled or usually termed as wheeled mobile robots equipped with an *on-board* monocular vision system. Due to large number of mobile robot visual tracking control methods, we classify the reported methods into four groups based on the type of the target to be tracked. Many efforts focus on the first group which aims to track a static target, such as a ground line, landmark, or reference image, for the purpose of mobile robot navigation or regulation (so-called homing) [5–17]. For a mobile robot to track a ground line, Ma et al. formulated the visual tracking control problem as controlling the shape of a ground curve in the image plane and proposed a closed-loop vision-guided control system for a nonholonomic mobile robot [5]. Coulaud et al. proposed a simple and stable feedback controller design, which avoids sophisticated image processing and control algorithms, for a mobile robot equipped with a fixed camera to track a line on the ground [6]. In the case of tracking the landmark, the reported controllers usually modify the visual servoing technique to satisfy the nonholonomic constraint for the motion control of the mobile robot [7–10]. In [11], Zhang and Ostrowski utilized an optimal control method to solve the visual motion-planning problem by generating a virtual trajectory in the image plane and the corresponding optimal control signals for the robot to follow. Nierobisch et al. proposed a visual tracking control method for a mobile robot with a pan-tilt camera to track visual reference landmarks in the acquired views during autonomous navigation [12]. Recently, the homography-based [13,14] and epipolar-based [4,15–17] visual tracking control approaches were proposed for a mobile robot equipped with a pinhole or an omni-directional (so-called central catadioptric) camera to track a reference image toward a desired configuration. These two approaches consider the mobile robot visual tracking

* Corresponding author. Tel.: +886 3 5731865; fax: +886 3 5715998.
E-mail addresses: chiyi.ece91g@nctu.edu.tw (C.-Y. Tsai), ktsong@mail.nctu.edu.tw (K.-T. Song), xavier.dutoit@mech.kuleuven.be (X. Dutoit), marnix.nuttin@mech.kuleuven.be (M. Nuttin).

0921-8890/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2009.01.001

control problem as a visual servoing regulation or visual homing problem. In [13], Chen et al. developed a visual tracking controller based on the Euclidean homography to track a desired time-varying trajectory defined by a prerecorded image sequence of a stationary target viewed by the on-board camera as the mobile robot moves. However, the stability of their result is restricted by the non-zero reference velocity condition of the desired trajectory. To overcome this drawback, Fang et al. exploited Lyapunov-based techniques to construct a homography-based visual servoing regulation controller for proving asymptotic regulation of the mobile robot [14]. In [15], Mariottini et al. exploited the epipolar geometry defined by the current and desired camera views to develop a two-step visual servoing regulation controller. They also extend this design to the visual servoing regulation control of a mobile robot with a central catadioptric camera [16]. In [17], Goedemé et al. developed a vision-only navigation and homing system for mobile robots with an omni-directional camera. Their method divides the visual homing operation into two phases and computes visual homing vector based on epipolar geometry estimation. Although these approaches of the first group provide appropriate solutions for static target visual tracking control problem, they cannot guarantee to solve the moving (non-static) target visual tracking control problem.

The second group aims to track other robot teammates in a robot team for the formation control purpose [18,19]. The proposed approaches in this group usually are based on a central catadioptric camera model in order to detect all robot teammates at the same instant. The subject of the third group is to track a predictable moving target, such as a projectile or straight moving ball, for mobile robot interception purpose [20,21]. In [20], Borgstadt et al. utilized a human vision-based strategy to guide a mobile robot to intercept a projectile ball. Similarly, Capparella et al. extended the concept of human-like strategy to develop a vision-based two-level interception approach, which contains a lower level controller to control the on-board pan-tilt camera and a higher lever controller to operate the mobile robot platform, for intercepting a straight moving ball [21]. A common point of the second and third group is that the motion of the interesting target is known and predictable. However, in many robotic applications, a mobile robot requires to track a dynamic and unpredictable motion target, such as a human's face, for the purpose of pursuit or interaction. Thus, the existent methods of the aforementioned two groups are not suitable to solve the dynamic moving target visual tracking control problem.

The purpose of the fourth group is to solve the problem of tracking a dynamic moving target [22–25]. In [22], Wang et al. proposed an adaptive backstepping control law based on an image-based camera-target visual interaction model to track a dynamic moving target with unknown height parameter. Although the approach in [22] guarantees the asymptotic stability of a closed-loop visual tracking control system in tracking a dynamic moving target, the case of tracking a static target cannot be guaranteed due to the non-zero restrictions on the reference velocity of the mobile robot. In [23], Malis et al. integrated template-based visual tracking algorithms and model-free vision-based control techniques to build a flexible and robust visual tracking control system for various robotic applications. Because their visual tracking result is based on the homography estimation, which requires two images of the target pattern to estimate the optimal homography, the reported system only overcomes the partial occlusion problem but fails in the fully occlusion problem. In [24], Han et al. proposed an image-based visual tracking control scheme for a mobile robot to estimate the position of the target in the next image and track the target to the central area of the image. Since their method utilized the differential approximation method to estimate the velocity of the target in the image plane, the estimation result is very sensitive to image noise. Recently, a visual interaction controller had been proposed for a unicycle-modeled mobile robot to track a dynamic moving target such as a human's face [25]. The drawback of this method is that the controller requires the target's 3D motion velocity, which is difficult to estimate when only a monocular camera is used.

From the literature survey, we note that challenges in the mobile robot visual tracking control design is to develop a robust tracking control system to estimate the motion of the moving target and to track the target based on a stability criterion. This problem motivates us to derive a new model for developing a robust visual tracking control system to solve the tracking problem of dynamic moving target in the image plane directly and efficiently. To do so, the visual interaction model described in [25] is extended to derive a novel dual-Jacobian visual interaction model for designing a robust mobile robot visual tracking control system, which encompasses a visual state estimator and a visual tracking controller. The visual state estimator is constructed by a real-time self-tuning Kalman filter and aims to estimate the optimal system state and target motion in the image plane directly for later use by the visual tracking controller. The visual tracking controller then calculates the robot's control velocities in the image plane directly. The main differences between the proposed method and other existent approaches are summarized as follows:

(1) The proposed dual-Jacobian visual interaction model considers not only the effect of mobile robot motion, but also the effect of target motion. Thus, based on the proposed model, the visual tracking control problem of a unicycle-modeled mobile robot for tracking a dynamic moving target can be solved with asymptotic convergence using a single controller. Moreover, the proposed model also considers the kinematics of a tilt camera platform mounted on the mobile robot. Therefore, the applicability of the proposed method is greatly increased.
(2) The proposed visual tracking control system not only possesses some degree of robustness against the system model uncertainties, but also overcomes the unmodeled quantization effect in the velocity commands and the occlusion effect during visual tracking process. This advantage enhances the reliability of the proposed method in practical applications.
(3) The proposed visual tracking control system works fully in the image space. Therefore, compared with position-based [23], homography-based [13,14], and epipole-based [4, 15–17] visual tracking control approaches, the computational complexity and the sensor/camera modeling errors can be much reduced due to the advantages of image-based visual servo control [2].
(4) The proposed self-tuning Kalman filter can automatically choose a suitable observation covariance matrix in varying environmental conditions. This helps to improve the estimation performance when there is noise in the system observation.

To validate the performance and robustness of the proposed control system, computer simulation and experimental studies of tracking a moving target have been conducted. Simulation and experimental results will be presented and discussed to verify the effectiveness of the proposed control system, in terms of tracking performance, system convergence, and robustness. Note that a brief version of the research results has been published in [26]. This paper will present the complete design of the proposed tracking control system, including robustness analysis, computer simulation and experimental validation.

The rest of this paper is organized as follows. Section 2 describes the proposed dual-Jacobian visual interaction model. Section 3 presents the results of visual tracking controller design. Section 4 develops the visual state estimator using Kalman filter with self-tuning algorithm. Simulation and experimental results are reported in Section 5. Extended discussion of several interesting observations will be presented. Section 6 concludes the contributions of this paper.

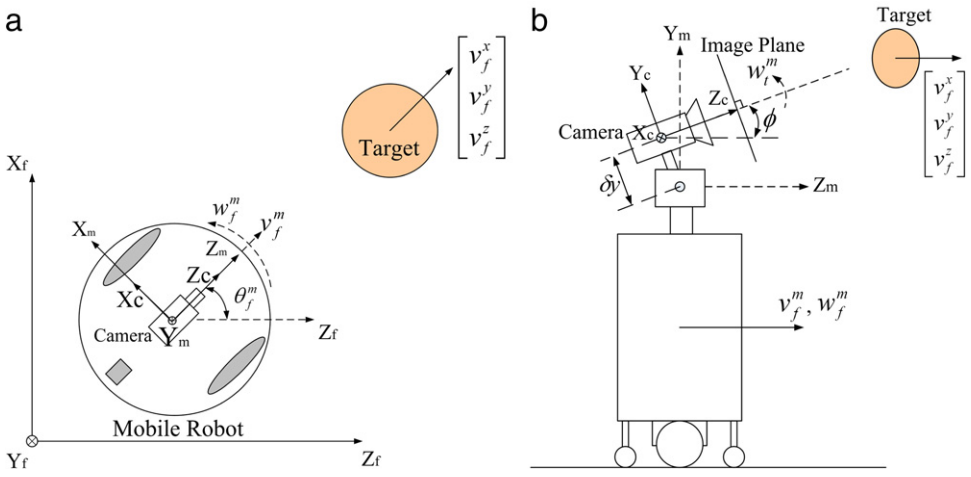

Fig. 1. (a) A model of the unicycle-modeled mobile robot and target in the world coordinate frame. (b) Side view of the mobile robot with a tilt camera mounted on top of its head to track a dynamic target.

2. Dual-Jacobian visual interaction model

As shown in Fig. 1, the considered system is a unicycle-modeled mobile robot equipped with a tilt camera mounted on top of it to track a moving target, such as a human face, in the image plane. The robot can keep the target in the camera's field of view such that the optical axis of this camera faces the interested target. Fig. 1(a) illustrates the model of the unicycle-modeled mobile robot and target in the world coordinate frame F_f, in which the motion of the target is supposed to be holonomic. 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 unicycle-modeled mobile robot and the target can be described [27], respectively, by

$$\begin{bmatrix} \dot{z}_f^m \\ \dot{x}_f^m \\ \dot{\theta}_f^m \\ \dot{\phi}_f^m \end{bmatrix} = \begin{bmatrix} \cos\theta_f^m & 0 & 0 \\ \sin\theta_f^m & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_f^m \\ w_f^m \\ w_t^m \end{bmatrix}, \quad \dot{y}_f^m = 0 \quad \text{and} \quad \begin{bmatrix} \dot{z}_f^t \\ \dot{x}_f^t \\ \dot{y}_f^t \end{bmatrix} = \begin{bmatrix} v_f^z \\ v_f^x \\ v_f^y \end{bmatrix}, \tag{1}$$

where (x_f^m, y_f^m, z_f^m) and (x_f^t, y_f^t, z_f^t) are, respectively, the positions of the mobile robot and the target in Cartesian coordinates. (θ_f^m, ϕ) are the orientation angle of the mobile robot and the tilt angle of the onboard camera. (v_f^m, w_f^m) and w_t^m are, respectively, the linear and angular velocity of the mobile robot and the tilt velocity of the camera. (v_f^x, v_f^y, v_f^z) are the target velocity in Cartesian coordinates.

In order for the mobile robot to interact with the target in the image coordinate frame, a visual interaction model was proposed in authors' previous work [25]. Fig. 2 shows the definition of the observed system state in the image plane used for the visual interaction model. In Fig. 2, x_i and y_i are, respectively, the horizontal and vertical position of the centroid of target in the image plane, and d_x is the width of target in the image plane. Let $X_i = [\,x_i \quad y_i \quad d_x\,]^{\mathrm{T}}$ denote the system state in the image plane, $u = [\,v_f^m \quad w_f^m \quad w_t^m\,]^{\mathrm{T}}$ is the control velocity of the mobile robot and on-board tilt camera, (f_x, f_y) represent fixed focal length along the image x-axis and y-axis, respectively [28], and W stands for the actual width of the target. The visual interaction model between robot and target in the image coordinate frame can be derived as in Box I, where diag() denotes a 3-by-3 diagonal matrix, $k_x = d_x/W$ and $k_y = k_x f_y/f_x$ are two scalars, and δy is the distance between the center of robot tilt platform and the onboard camera. A detailed derivation of the visual interaction model (2) given in Box I is presented in Appendix.

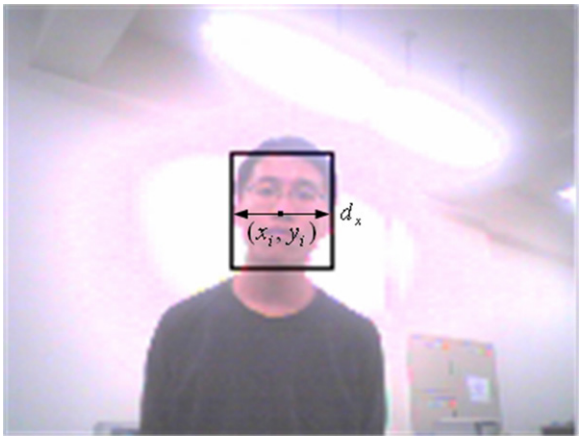

Fig. 2. Definition of the observed system state in the image plane.

The visual interaction model (2) given in Box I indicates that the elements of system matrix $\mathbf{A}_i$ and vector C_i are functions of the target velocity. Thus, expression (2) given in Box I can be rewritten as in Box II, and $V_t = [\,v_f^x \quad v_f^y \quad v_f^z\,]^{\mathrm{T}}$ is the vector of target velocities in Cartesian coordinates. Expression (3) given in Box II shows that the visual interaction model consists of two parts: first, the effect of target motion $\dot{X}_i^t \equiv \left[\dot{x}_i^t \quad \dot{y}_i^t \quad \dot{d}_x^t\right]^{\mathrm{T}} = \mathbf{J}_t V_t$, and second, the effect of mobile robot motion $\dot{X}_i^m \equiv \left[\dot{x}_i^m \quad \dot{y}_i^m \quad \dot{d}_x^m\right]^{\mathrm{T}} = \mathbf{B}_i u$. Thus, the Eq. (3) given in Box II can be rewritten as a dual-Jacobian equation such that

$$\dot{X}_i = \dot{X}_i^t + \dot{X}_i^m = \mathbf{J}_t V_t + \mathbf{B}_i u, \tag{4}$$

where matrix $\mathbf{J}_t$, termed as *target image Jacobian*, transfers the target velocity V_t into target image velocity $\dot{X}_i^t$; matrix $\mathbf{B}_i$, which denotes *robot image Jacobian*, transfers 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 the combination of target image velocity $\dot{X}_i^t$ and robot image velocity $\dot{X}_i^m$. Fig. 3 shows the concept of dual-Jacobian equation (4). Therefore, the visual interaction between robot and target in the image coordinate frame can be modeled as a dual-Jacobian visual interaction model (4).

$$\dot{X}_i = \mathbf{A}_i X_i + \mathbf{B}_i u + C_i \tag{2}$$

where $\mathbf{A}_i = \text{diag}(A_1, A_2, A_1)$, $A_1 = -f_x^{-1} k_x (v_f^x \cos\phi \sin\theta_f^m + v_f^y \sin\phi + v_f^z \cos\phi \cos\theta_f^m)$

$$A_2 = -f_y^{-1} k_y (v_f^x \cos\phi \sin\theta_f^m + v_f^y \sin\phi + v_f^z \cos\phi \cos\theta_f^m)$$

$$\mathbf{B}_i = \begin{bmatrix} f_x^{-1} k_x x_i \cos\phi & f_x^{-1}(x_i^2 + f_x^2)\cos\phi - f_y^{-1} f_x (k_y \delta y + y_i)\sin\phi & -f_y^{-1} x_i (k_y \delta y + y_i) \\ k_y(\sin\phi + f_y^{-1} y_i \cos\phi) & f_x^{-1} f_y x_i (\sin\phi + f_y^{-1} y_i \cos\phi) & -f_y^{-1}(y_i^2 + f_y^2 + k_y y_i \delta y) \\ f_x^{-1} k_x d_x \cos\phi & f_x^{-1} x_i d_x \cos\phi & -f_y^{-1} d_x (k_y \delta y + y_i) \end{bmatrix}$$

$$C_i = \begin{bmatrix} k_x (v_f^z \sin\theta_f^m - v_f^x \cos\theta_f^m) \\ k_y (v_f^y \cos\phi - v_f^x \sin\phi \sin\theta_f^m - v_f^z \sin\phi \cos\theta_f^m) \\ 0 \end{bmatrix}$$

Box I.

$$\dot{X}_i = (\mathbf{A}_i X_i + C_i) + \mathbf{B}_i u = \mathbf{J}_t V_t + \mathbf{B}_i u, \tag{3}$$

where

$$\mathbf{J}_t = \begin{bmatrix} -k_x (f_x^{-1} x_i \cos\phi \sin\theta_f^m + \cos\theta_f^m) & -k_x f_x^{-1} x_i \sin\phi & -k_x (f_x^{-1} x_i \cos\phi \cos\theta_f^m - \sin\theta_f^m) \\ -k_y (f_y^{-1} y_i \cos\phi \sin\theta_f^m + \sin\phi \sin\theta_f^m) & -k_y (f_y^{-1} y_i \sin\phi - \cos\phi) & -k_y (f_y^{-1} y_i \cos\phi \cos\theta_f^m + \sin\phi \cos\theta_f^m) \\ -k_x f_x^{-1} d_x \cos\phi \sin\theta_f^m & -k_x f_x^{-1} d_x \sin\phi & -k_x f_x^{-1} d_x \cos\phi \cos\theta_f^m \end{bmatrix}$$

Box II.

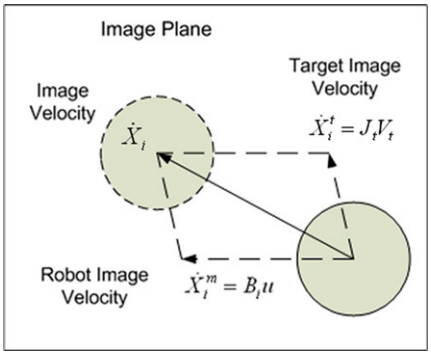

Fig. 3. Depicts the concept of dual-Jacobian visual interaction model (4).

Similar to the human's visual tracking behavior, the purpose of visual tracking control design aims to control the centroid position and width of target from an initial state to the desired state in the image plane. To achieve this, an error state model is introduced to design the tracking controller. Define the error coordinates in the image plane such that

$$\begin{aligned} X_e &= \begin{bmatrix} x_e & y_e & d_e \end{bmatrix}^{\mathrm{T}} = \bar{X}_i - X_i^* \\ &= \begin{bmatrix} \bar{x}_i - x_i^* & \bar{y}_i - y_i^* & \bar{d}_x - d_x^* \end{bmatrix}^{\mathrm{T}}, \end{aligned} \tag{5}$$

where $\bar{X}_i = \begin{bmatrix} \bar{x}_i & \bar{y}_i & \bar{d}_x \end{bmatrix}^{\mathrm{T}}$ is the desired state in the image plane; $X_i^* = \begin{bmatrix} x_i^* & y_i^* & d_x^* \end{bmatrix}^{\mathrm{T}}$ is the estimated system state from a visual state estimator (VSE, see Section 4). If the system state converges to the desired state, the visual tracking control problem is solved. Furthermore, the dynamic error state model in the image plane can be derived directly by taking the derivative of (5). The result is given by

$$\dot{X}_e = -\dot{X}_i^t - \dot{X}_i^m = -\mathbf{J}_t V_t - \mathbf{B}_i u. \tag{6}$$

With the new coordinate X_e, the visual tracking control problem is transformed into a stability problem. In the rest of this paper, the new coordinate X_e will be used to solve the visual tracking control problem. If X_e converges to zero, then the visual tracking control problem is solved.

3. Visual tracking controller (VTC) design

In this section, a visual tracking control law is derived based on the error state model (6) to interact with an interesting target in the image plane by exploiting feedback linearization and pole placement techniques. The robustness analysis of the proposed controller against parametric uncertainties is also presented.

3.1. Feedback linearization and pole placement

According to the dynamic error state model (6), a feedback control law can be obtained by feedback linearization such that

$$u = \mathbf{B}_i^{-1}(\mathbf{K}_g X_e - \mathbf{J}_t V_t) = \mathbf{B}_i^{-1}(\mathbf{K}_g X_e - \dot{X}_i^t), \tag{7}$$

where $\mathbf{K}_g$ is an 3-by-3 gain matrix. Substituting (7) into (6) yields

$$\dot{X}_e = -\mathbf{K}_g X_e. \tag{8}$$

It is clear that if all eigen values of matrix $-\mathbf{K}_g$ are constant and lie inside the left-half complex plane, then the system error state $X_e(t)$ will decay exponentially to zero. Thus, we choose the gain matrix by pole placement such that

$$\mathbf{K}_g = \text{diag}(\alpha_1, \alpha_2, \alpha_3), \tag{9}$$

in which $(\alpha_1, \alpha_2, \alpha_3)$ are three positive constants. Substituting (9) into (8) yields

$$\dot{X}_e = -\mathbf{K}_g X_e = -\text{diag}(\alpha_1, \alpha_2, \alpha_3) X_e. \tag{10}$$

Suppose that the initial error state $X_e(t_0)$ is within the image plane. Then expression (10) indicates that

$$\begin{aligned} X_e(t)|_{t_0, X_e(t_0)} &\equiv X_e(t; t_0, X_e(t_0)) \\ &= \text{diag}(\mathrm{e}^{-\alpha_1(t-t_0)}, \mathrm{e}^{-\alpha_2(t-t_0)}, \mathrm{e}^{-\alpha_3(t-t_0)}) X_e(t_0) \end{aligned}$$

$$\text{for some } t_0 \geq 0. \tag{11}$$

Because $(\alpha_1, \alpha_2, \alpha_3)$ are three positive constants, expression (11) leads to the following inequality:

$$\|X_e(t; t_0, X_e(t_0))\| \leq \mathrm{e}^{-\lambda_{\min}(\mathbf{K}_g)(t-t_0)} \|X_e(t_0)\| \quad \text{for all } t \geq t_0, \tag{12}$$

where $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of matrix $\mathbf{A}$, and $\|B\|$ denotes the 2-norm value of vector B. From (12), it is clear that the system error state satisfies $\|X_e(t; t_0, X_e(t_0))\| \leq \|X_e(t_0)\|$

and $\lim_{t\to\infty}\|X_e(t;t_0,X_e(t_0))\| = 0$, and thus the visual tracking control problem is solved. Summarizing the above discussions, we obtain the following proposition.

Proposition 1. *Suppose that the initial system state X_i is within the image plane. Let $(\alpha_1,\alpha_2,\alpha_3) > 0$ be three positive constants. Consider the dual-Jacobian visual interaction system (4). If the matrix $\mathbf{B}_i$ is nonsingular, then the closed-loop visual tracking system (6) is asymptotically stable by using the control law*

$$u = \mathbf{B}_i^{-1}(\mathbf{K}_g X_e - \dot{X}_i^t), \tag{13}$$

where $\dot{X}_i^t = \mathbf{J}_t V_t$ is the target image velocity defined in (4), matrices $\mathbf{B}_i$ and $\mathbf{J}_t$ are the robot and target image Jacobian defined in (2) and (3), respectively, X_e is the system error state defined in (5), and $\mathbf{K}_g$ is a 3-by-3 diagonal gain matrix defined in (9).

Proof. Consider the closed-loop visual tracking system (6). We first define a positive-definite Lyapunov function associated with the system error state

$$V(x_e,y_e,d_e) = \frac{1}{2}(x_e^2 + y_e^2 + d_e^2). \tag{14}$$

Taking the derivative of (14) yields

$$\dot{V} = X_e^{\mathrm{T}}\dot{X}_e = -(X_e^{\mathrm{T}}\mathbf{J}_t V_t + X_e^{\mathrm{T}}\mathbf{B}_i u) = -X_e^{\mathrm{T}}(\dot{X}_i^t + \mathbf{B}_i u) \equiv -f(u), \tag{15}$$

where $f(u) = X_e^{\mathrm{T}}(\dot{X}_i^t + \mathbf{B}_i u)$. In view of Lyapunov theory [29], expression (15) tells us that if $f(u) > 0$ then the equilibrium point of (6) is asymptotically stable. Substituting the control law (13) into $f(u)$, we then have

$$f(u) = X_e^{\mathrm{T}}\mathbf{K}_g X_e, \tag{16}$$

where $\mathbf{K}_g = \mathrm{diag}(\alpha_1,\alpha_2,\alpha_3)$, and $(\alpha_1,\alpha_2,\alpha_3)$ are three positive constants. Since $\mathbf{K}_g$ is a symmetric positive definite (SPD) matrix, the following inequality holds:

$$f(u) \geq \lambda_{\min}(\mathbf{K}_g)\,\|X_e\|^2 = \min(\alpha_1,\alpha_2,\alpha_3)\,\|X_e\|^2 > 0, \tag{17}$$

where $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of matrix $\mathbf{A}$. Expression (17) concludes that the closed-loop visual tracking system (6) is asymptotically stable and hence completes the proof. ■

3.2. Singularity analysis

The feedback linearization control law (13) poses a singularity problem of matrix $\mathbf{B}_i$. In [25], the singularity condition of matrix $\mathbf{B}_i$ is derived such that

$$f_y = (y_i + Sd_x)\tan\phi, \tag{18}$$

where $S = (f_y\delta y)/(f_x W)$ is a fixed scalar factor. Let (x_c,y_c,z_c) represents the related position between robot and target in the camera coordinate frame. Because of $f_y = k_y z_c$, $y_i = k_y y_c$, and $d_x = k_x W$ (see (A.2) in Appendix), singularity condition (18) can be rewritten such that

$$z_c = (y_c + \delta y)\tan\phi, \tag{19}$$

where z_c is the distance from the camera to the target. Fig. 4 shows the particular configuration of the singular condition (19). In Fig. 4, the distance z_ϕ equals:

$$z_\phi = (y_c + \delta y)\tan\phi. \tag{20}$$

From (19) and (20), it is clear that the physical meaning of the singularity condition (19) is that the distance between the camera and the target equals to the distance z_ϕ. In general, this situation usually will not happen unless the target directly locates on Y_m-axis (the Y-axis of the mobile coordinate frame F_m), which means that the target is directly above or directly below the robot. Under such circumstances, the robot will be unable to approach the target in any way due to deficient degree-of-freedom, and the robot should stop tracking temporarily.

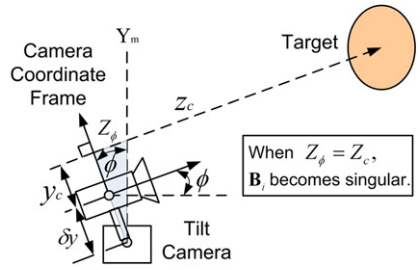

Fig. 4. Particular configuration of the singularity condition (10).

Remark 1. The proposed control law (13) is associated with the target's velocity in the working space or target's image velocity in the image plane. If the 3D velocity of the target V_t is known, the target image velocity $\dot{X}_i^t$ can be calculated by using target image Jacobian $\mathbf{J}_t$ without the estimation process. However, in practical applications, it is difficult to measure the 3D velocity of the target when using only one camera in real-time operations. In this situation, an estimation process is required to estimate the target image velocity $\dot{X}_i^t$ in the image plane directly. In Section 4, a VSE will be proposed to accomplish this task. This design will facilitate more general performance of the proposed tracking control scheme in the image plane.

3.3. Robustness analysis

In this subsection, we investigate the robustness of the proposed VTC (13) against model uncertainties on camera parameters (f_x,f_y), robot parameters (θ_f^m,ϕ), and target parameters $(W,\dot{x}_i^t,\dot{y}_i^t,\dot{d}_x^t)$. Consider the following closed-loop visual tracking system with parametric uncertainties:

$$\dot{X}_e = -\dot{\bar{X}}_i^t - \bar{\mathbf{B}}_i u = -(\dot{X}_i^t + \delta\dot{X}_i^t) - (\mathbf{B}_i + \delta\mathbf{B}_i)u, \tag{21}$$

where $\delta\dot{X}_i^t$ and $\delta\mathbf{B}_i$ are unknown bounded disturbances. Recall the positive-definite Lyapunov function defined in (14), the derivative of (14) with parametric uncertainties becomes

$$\dot{\bar{V}} = -X_e^{\mathrm{T}}(\dot{\bar{X}}_i^t + \bar{\mathbf{B}}_i u) = -[f(u) + \delta f(u)] \equiv -\bar{f}(u), \tag{22}$$

where $\delta f(u) = X_e^{\mathrm{T}}(\delta\dot{X}_i^t + \delta\mathbf{B}_i u)$ is unknown. Assume that $\delta f(u)$ is bounded and there exists a SPD matrix $\tilde{\mathbf{Q}}$ such that

$$\delta f(u) < \|X_e\|_{\tilde{\mathbf{Q}}}^2, \tag{23}$$

where $\|X\|_{\mathbf{A}} = (X^{\mathrm{T}}\mathbf{A}X)^{1/2}$ denotes a weighted vector norm with a SPD matrix $\mathbf{A}$. Now, the main result is presented as follows.

Theorem 1. *Consider the dual-Jacobian visual interaction system (4) with unknown bounded parametric uncertainties $\delta\dot{X}_i^t$ and $\delta\mathbf{B}_i$ defined in (21). Let $\tilde{\mathbf{Q}} > 0$ be a SPD matrix defined in (23). Choose the controller u as given in the expression (13) with a constant SPD matrix $\mathbf{K}_g = \mathrm{diag}(\alpha_1,\alpha_2,\alpha_3) > 0$. Then, the closed-loop visual tracking system (6) is asymptotically stable for all $\alpha_i > \lambda_{\max}(\tilde{\mathbf{Q}})$, $i = 1,2,3$.*

Proof. From (22) and (23), it follows that

$$f(u) = \bar{f}(u) - \delta f(u) > \bar{f}(u) - \|X_e\|_{\tilde{\mathbf{Q}}}^2. \tag{24}$$

Expression (24) implies that $\bar{f}(u) - \|X_e\|_{\tilde{\mathbf{Q}}}^2$ is a lower-bound of $f(u)$. If $\bar{f}(u) - \|X_e\|_{\tilde{\mathbf{Q}}}^2 > 0$ can be guaranteed, then $f(u) > 0$ is satisfied and thus the system has the robust property w.r.t. the parametric uncertainties.

Choose the controller u as in (13) with parametric uncertainties such that

$$u = \bar{\mathbf{B}}_i^{-1}(\mathbf{K}_g X_e - \dot{\bar{X}}_i^t), \tag{25}$$

where $\dot{\bar{X}}_i^t = \bar{\mathbf{J}}_t V_t$. Substituting (25) into (22) yields

$$\dot{\bar{V}} = -\bar{f}(u) = -\|X_e\|^2_{\mathbf{K}_g}, \tag{26}$$

where $\mathbf{K}_g = \text{diag}(\alpha_1, \alpha_2, \alpha_3) > 0$ is a constant SPD matrix. From (24) and (26), it is clear that

$$f(u) > \|X_e\|^2_{\mathbf{K}_g} - \|X_e\|^2_{\tilde{\mathbf{Q}}} \geq [\lambda_{\min}(\mathbf{K}_g) - \lambda_{\max}(\tilde{\mathbf{Q}})]\,\|X_e\|^2, \tag{27}$$

where $\lambda_{\min}(\mathbf{A})$ is defined in (17), and $\lambda_{\max}(\mathbf{A})$ denotes the maximum eigenvalue of matrix $\mathbf{A}$. Expression (27) tells us that if $\lambda_{\min}(\mathbf{K}_g) - \lambda_{\max}(\tilde{\mathbf{Q}})$ is positive, then $f(u) > 0$ is satisfied and thus the equilibrium point of (6) is asymptotically stable. This means that the proposed VTC (13) is robust against the unknown parametric uncertainties and hence completes the proof. ■

Remark 2. In realization of the control schemes, it is worth noting that the quantization error in velocity commands degrade the performance of the controller and might make the system unstable. In order to overcome this problem, the proposed VTC (13) will be combined with a robust control law presented in [25] for improving the robustness of the visual tracking control system. Interested readers refer to [25] for more technical details.

4. Visual state estimator (VSE) design

As implied by Proposition 1, the VTC (13) requires the information of target image velocity $\dot{X}_i^t$. This requirement poses two questions: first, how the target image velocity can be estimated in the image plane directly; second, what estimation methods can be used. In this section, we first formulate the problem of target image velocity estimation in the image space. A VSE will then be proposed in order to compute the optimal estimates of target status X_i and target image velocity $\dot{X}_i^t$ in the weighted least squared error sense for later used by the VTC.

4.1. Problem formulation

Since actual image processing is discrete, the first step of VSE design is to discretize the system model (4) into a corresponding discrete form. By the definition $\dot{x}(t) = \lim_{T\to 0}[x(t) - x(t-T)]/T$, T denotes the sampling time of the digital system, one can approximate the system model (4) as

$$X_i[n] = X_i[n-1] + T\dot{X}_i^t[n-1] + T\mathbf{B}_i u_{n-1}, \quad \text{for } n = 1, 2, \ldots \tag{28}$$

where $u_n = \begin{bmatrix} v_f^m & w_f^m & w_t^m \end{bmatrix}^{\mathrm{T}}$ is the discrete-time control signal at sample instant n. Suppose that the target motion can be approximated by a smooth motion during a sampling period. Then the target image velocity has the following relationship between two consecutive sampling instants

$$\dot{X}_i^t[n] = \dot{X}_i^t[n-1]. \tag{29}$$

Based on (28) and (29), the propagation model of the visual state estimator is given by

$$X_n = \begin{bmatrix} \mathbf{I}_3 & T\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} X_{n-1} + \begin{bmatrix} T\mathbf{B}_i \\ \mathbf{0}_3 \end{bmatrix} u_{n-1} \equiv \mathbf{A}_{est} X_{n-1} + \mathbf{B}_{est} u_{n-1}, \tag{30}$$

where $X_n^{\mathrm{T}} = \begin{bmatrix} (X_i[n])^{\mathrm{T}} & (\dot{X}_i^t[n])^{\mathrm{T}} \end{bmatrix}$ is the vector of system estimates at instant n, $\mathbf{I}_3$ is a 3-by-3 identity matrix, and $\mathbf{0}_3$ is a 3-by-3 zero matrix. Next, because the observed image only contains the information of target status X_i in each instant, the observation model of VSE is given by

$$Z_n = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix} X_n \equiv \mathbf{H}_{est} X_n. \tag{31}$$

Based on the propagation model (30) and observation model (31), the estimation problem in the image plane becomes to find the state estimate X_n^* that minimizes the weighted least square distance:

$$X_n^* = \arg\min_X [(X_n - X)^{\mathrm{T}} \mathbf{P}_n^{-1} (X_n - X) + (Z_n - \mathbf{H}_{est} X)^{\mathrm{T}} \mathbf{R}_n^{-1} (Z_n - \mathbf{H}_{est} X)], \tag{32}$$

where $\mathbf{P}_n = \mathbf{A}_{est}\mathbf{P}_{n-1}\mathbf{A}_{est}^{\mathrm{T}}$ is the covariance matrix of propagation model at instant n, and $\mathbf{R}_n$ is the covariance matrix of observation model at instant n.

4.2. Self-tuning Kalman filter

Because the propagation model (30) and observation model (31) are both linear equations, a Kalman filter will provide the optimal estimate X_n^* based on the performance criterion (32) when (30) and (31) have Gaussian uncertainties [30]:

$$\text{Propagation: } X_n = \mathbf{A}_{est} X_{n-1}^* + \mathbf{B}_{est} u_{n-1} + \delta X_{n-1} \tag{33a}$$

$$\text{Propagation Covariance Matrix: } \mathbf{P}_n = \mathbf{A}_{est}\mathbf{P}_{n-1}^*\mathbf{A}_{est}^{\mathrm{T}} + \mathbf{Q}_{n-1} \tag{33b}$$

$$\text{Observation: } Z_n = \mathbf{H}_{est} X_n + \delta Z_n \tag{33c}$$

where $(X_n, \mathbf{P}_n)$ are the propagation state and the corresponding covariance matrix at instant n; $(X_{n-1}^*, \mathbf{P}_{n-1}^*)$ are the optimal estimate and the corresponding covariance matrix at instant n-1; $\delta X_n^{\mathrm{T}} = [(\delta X_i[n])^{\mathrm{T}} \quad (\delta \dot{X}_i^t[n])^{\mathrm{T}}] \sim N(0, \mathbf{Q}_n)$ represents Gaussian propagation uncertainty with zero mean and covariance matrix $\mathbf{Q}_n$ at instant n; and $\delta Z_n \sim N(0, \mathbf{R}_n)$ denotes Gaussian observation uncertainty with zero mean and covariance matrix $\mathbf{R}_n$ at instant n. Based on Eqs. (33a)–(33c), the local minimum solution of performance criterion (32) and the corresponding covariance matrix at instant n are given by

$$X_n^* = X_n^p + \mathbf{K}_n(Z_n - \mathbf{H}_{est} X_n^p) \quad \text{and} \quad \mathbf{P}_n^* = (\mathbf{I}_6 - \mathbf{K}_n\mathbf{H}_{est})\mathbf{P}_n, \tag{34}$$

where $X_n^p = \mathbf{A}_{est} X_{n-1}^* + \mathbf{B}_{est} u_{n-1}$ is the ideal propagation state, $\mathbf{K}_n = \mathbf{P}_n\mathbf{H}_{est}^{\mathrm{T}}(\mathbf{H}_{est}\mathbf{P}_n\mathbf{H}_{est}^{\mathrm{T}} + \mathbf{R}_n)^{-1}$ is the Kalman gain matrix, and $\mathbf{I}_6$ is a 6-by-6 identity matrix.

Although expression (34) provides the best linear estimates at each instant, the filter performance still depends on the covariance matrices $\mathbf{Q}_n$ and $\mathbf{R}_n$. Thus, a difficult problem in Kalman filter applications is to determine the values of matrices $\mathbf{Q}_n$ and $\mathbf{R}_n$ for computing Kalman gain matrix $\mathbf{K}_n$ [31]. Moreover, the observation uncertainty usually varies with the conditions of target motion (such as orientation and rotation of the human face) and working environment (such as light variation and occlusion); the corresponding covariance matrix $\mathbf{R}_n$ would be time-varying for various operating conditions. These problems motivate us to combine a self-tuning algorithm with a Kalman filter to choose a suitable observation covariance matrix $\mathbf{R}_n$ in varying environmental conditions. On the other hand, because the propagation uncertainty and the corresponding covariance matrix $\mathbf{Q}_n$ are difficult to estimate online, the propagation covariance matrix $\mathbf{Q}_n$ will be fixed at initialization without updating in this design.

The proposed self-tuning algorithm attempts to estimate the minimum variance of a set of observation data recorded over time. To do so, a linear-least-squares regression method is adopted to analyze the observed time series data [32]. The typical linear regression model for a discrete time series is given by

$$y_n = an + b + \varepsilon_n, \tag{35}$$

where the residual ε_n is a random variable with zero mean, (a, b) are the parameters to be determined by minimizing the variance of residuals. Fig. 5 shows the concept of the linear-least-squares regression, in which the solid line is the observed time series, and the dotted line indicates the best linear fitting $\hat{y}_n = an + b$ with minimum residual variance σ^2.

Fig. 5. Concept of time series linear-least-squares regression.

Let k denote the length of the observed time series. Based on the linear regression model (35), the observed time series can be modeled as

$$Y = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ \vdots & \vdots \\ k & 1 \end{bmatrix} \theta + \varepsilon \equiv \mathbf{A}_{st}\theta + \varepsilon, \tag{36}$$

where $Y = [y_1 \quad y_2 \quad \cdots \quad y_k]^{\mathrm{T}}$ is the vector of observed data over time, and $\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_k]^{\mathrm{T}}$ represents the corresponding residuals. $\theta = [a \quad b]^{\mathrm{T}}$ is the parameter vector to be detected such that

$$\theta^* = \min_{\theta} \mathrm{var}(\varepsilon) = \min_{\theta} \|\varepsilon\| = \min_{\theta} \|Y - \mathbf{A}_{st}\theta\|, \tag{37}$$

where $\mathrm{var}(x)$ is the variance of vector x, and $\|x\|$ is the norm of vector x. The optimal solution of (37) will be the least-squares solution such that

$$\theta^* = \mathbf{A}_{st}^{+} Y, \tag{38}$$

where $\mathbf{A}_{st}^{+} = (\mathbf{A}_{st}^{\mathrm{T}}\mathbf{A}_{st})^{-1}\mathbf{A}_{st}^{\mathrm{T}}$ denotes the pseudo-inverse matrix of $\mathbf{A}_{st}$. Substituting (38) into (36), the residual vector with minimum variance can be obtained by

$$\varepsilon^* = Y - \mathbf{A}_{st}\mathbf{A}_{st}^{+}Y = (\mathbf{I}_k - \mathbf{A}_{st}(\mathbf{A}_{st}^{\mathrm{T}}\mathbf{A}_{st})^{-1}\mathbf{A}_{st}^{\mathrm{T}})Y \equiv \mathbf{T}_{st}Y, \tag{39}$$

where $\mathbf{T}_{st} = \mathbf{I}_k - \mathbf{A}_{st}(\mathbf{A}_{st}^{\mathrm{T}}\mathbf{A}_{st})^{-1}\mathbf{A}_{st}^{\mathrm{T}}$ is a fixed k-by-k coefficient matrix, and $\mathbf{I}_k$ is a k-by-k identity matrix. Expression (39) tells us that the minimum variance residual vector ε^* is the linear transformation of observed data vector Y through a fixed transformation matrix $\mathbf{T}_{st}$. This observation provides us an efficient method for detecting the minimum variance of an observed data sequence in real-time. For instance, let X_1^k, Y_1^k, and D_1^k denote, respectively, the observed data sequence of x_i, y_i, and d_x over time steps 1 to k. Using (39), the minimum variances of X_1^k, Y_1^k, and D_1^k are given by

$$\begin{aligned} &\sigma_x^2 = \mathrm{var}(\mathbf{T}_{st}X_1^k), \qquad \sigma_y^2 = \mathrm{var}(\mathbf{T}_{st}Y_1^k), \quad \text{and} \\ &\sigma_d^2 = \mathrm{var}(\mathbf{T}_{st}D_1^k). \end{aligned} \tag{40}$$

Based on (40), the covariance matrix $\mathbf{R}_n$ is updated as

$$\mathbf{R}_n = \mathbf{R}_0 + \mathrm{diag}((\sigma_x^2)^2, (\sigma_y^2)^2, (\sigma_d^2)^2), \tag{41}$$

where $\mathbf{R}_0$ is the initial covariance matrix of $\mathbf{R}_n$. Combining the self-tuning equations (40)–(41) with Kalman filter equations (33)–(34), the implemented self-tuning Kalman filter is summarized in Fig. 6. The processing steps are listed as follows:

(1) Choose two initial covariance matrices $\mathbf{Q}_0$ and $\mathbf{R}_0$, usually by a trial-and-error procedure.
(2) Assume that the initial position of target locates 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) Store current observed measurement in a shift register with length k. If the length of storage data is equal to k, then compute the variance of the observed data sequences by (40) and update covariance matrix $\mathbf{R}_n$ by (41); else set $\mathbf{R}_n = \mathbf{R}_0$; go to step (4).
(4) Compute the ideal propagated state X_n^p defined in (32) and the corresponding propagation covariance matrix $\mathbf{P}_n$ using (33b).
(5) If the target is detected in the observed image, then compute the Kalman gain matrix $\mathbf{K}_n$ and update the estimated state vector X_n^* with the corresponding covariance matrix $\mathbf{P}_n^*$ using (34); else set $X_n^* = X_n^p$ and $\mathbf{P}_n^* = \mathbf{P}_n$; go to step (6).
(6) 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

Computer simulations and several interesting experiments have been performed to validate the robustness and tracking performance of the proposed control system. In the computer

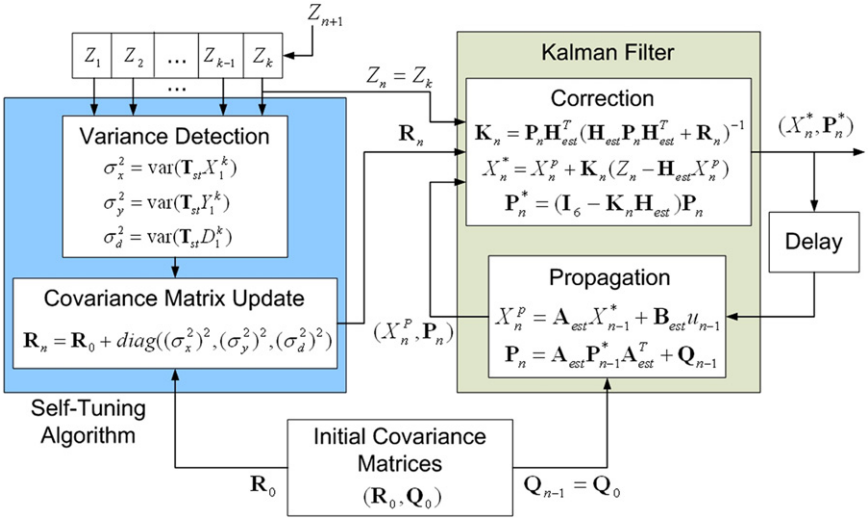

Fig. 6. Architecture of the proposed self-tuning Kalman filter.