Robotica (2010) volume 28, pp. 517–524. © Cambridge University Press 2009
doi:10.1017/S0263574709005827

A method for stereo-vision-based tracking for robotic applications

Pubudu N. Pathirana†*, Adrian N. Bishop†, Andrey V. Savkin‡, Samitha W. Ekanayake† and Timothy J. Black†

†*School of Engineering and IT, Deakin University, Australia*
‡*School of Electrical Engineering and Telecommunications, University of New South Wales, Australia*

(Received in Final Form: April 27, 2009. First published online: June 9, 2009)

SUMMARY
Vision-based tracking of an object using perspective projection inherently results in non-linear measurement equations in the Cartesian coordinates. The underlying object kinematics can be modelled by a linear system. In this paper we introduce a measurement conversion technique that analytically transforms the non-linear measurement equations obtained from a stereo-vision system into a system of linear measurement equations. We then design a robust linear filter around the converted measurement system. The state estimation error of the proposed filter is bounded and we provide a rigorous theoretical analysis of this result. The performance of the robust filter developed in this paper is demonstrated via computer simulation and via practical experimentation using a robotic manipulator as a target. The proposed filter is shown to outperform the extended Kalman filter (EKF).

KEYWORDS: Linear filtering; Set-estimation; Stereo vision; Robust filtering; Target tracking.

1. Introduction

This paper investigates the problem of tracking the real-world position and velocity coordinates of an object using sequences of images provided by a stereo-vision-based sensor system. Vision-based tracking systems permit cost-effective and passive object tracking applications in numerous areas. Particularly, vision-based state estimation is an important tool in navigation of robots, missiles and Unmanned Aerial Vehicles (UAVs).[1–3] Increasing number of applications take advantage of this form of technology over traditional technologies such as radar or sonar due to hardware limitations and associated complexities. Moreover, with this more popular form of measurement, it is possible to complement such tracking systems with higher level event or object understanding paradigms that is simply not possible with most alternative technologies. For example, vision sensors can more readily permit target identification and/or classification. As such, vision-based tracking systems have found application in a number of diverse application areas; e.g. see refs. [4–11]. Active stereo vision has also been widely used in visual servo control.[12] In implementing the designed controller in these approaches, Kalman filter is usually employed as a state estimator.[13] Vision has also been employed as an effective means of calibrating robotic manipulators for increased precision.[14] This is due to the fact that most robotic manipulators lack accuracy in comparison to repeatability particulary in industrial settings.

Specifically, we are focused on estimating the real-world trajectory of an object point in question. There are two fundamentally distinct approaches to the problem of target trajectory estimation using a sequence of image frames generated by a vision sensor. The first approach is feature based, where the target's image plane position is measured on each frame in the sequence (i.e. video) and subsequently used as the basis for a recursive real-world coordinate tracking filter. In ref. [15], for instance, a recursive target tracking filter is developed that uses the non-linear perspective projection measurements as the input to an iterated extended Kalman filter (IEKF). A number of papers have examined the problem along similar lines. Different camera models lead to different measurement systems. The second approach is based on so-called optical flow where the motion in the image plane is represented by a sampled velocity field. Again, this approach has been used in conjunction with dynamic modelling of the targets motion for the purpose of parameter estimation.[16] In ref. [16] a target tracking filter is developed that includes both a perspective projection measurement system and an optical flow based measurement system. This essentially results in more non-linear equations that include measures of the targets velocity as well as position, hence the use of the extended Kalman filter (EKF) in ref. [16]. In ref. [17], a robust version of the EKF (REKF) is used to estimate the heading of a vehicle in an automotive setting using fusion ideas in vision and sonar sensing.

In this paper, we employ a perspective projection measurement system and the corresponding time-derivative measurements in a stereo-vision-based tracking system. This modified version of the measurement conversion method specifically aimed at stereo-vision-based target tracking with both perspective projection and image-velocity-based measurements. It is well known that under the assumption of time independent image intensity, image velocity corresponds to the flow field motion[7,18,19] which can be directly measured from the image sequence. Therefore our measurement space essentially consists of not only the projected locations on each image plane but also the image velocities. The novelty of our approach comes in the form of a measurement conversion based linear robust filter (RF) algorithm that we derive. Essentially, we analytically

* Corresponding author. E-mail: pubudu@deakin.edu.au

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

convert the non-linear measurement equations into linear measurement equations and apply an RF. Hence, we solve the estimation problem strictly within the linear domain since we also consider linear state equations.[20] Traditionally, it has been common to use non-linear estimators such as the EKF which employ some form of numerical approximation (e.g. Taylor-series). Thus, very few results exist which give analytical analysis of convergence properties or estimation errors bounds for some non-linear observers.[21,22] It is well known that initial conditions are critical to the stability and convergence of the EKF. Furthermore, the errors introduced during linearization result in bias and filter inconsistency[23] often leading to divergence. In ref. [16] it is stated that after a detailed comparison of the EKF, IEKF and an iterated linear filter smoother in ref. [24] similar performances are observed for the problem of vision-based target tracking.

Alternatively, measurement conversion methods have been explored particularly for target tracking with radar measurements.[20] The basic idea of these techniques is to transform non-linear measurement equations in to a linear combination of the Cartesian coordinates, estimate the bias and covariance of the converted measurement noise, and then use the standard linear Kalman filter.[25] The measurement conversion methods have proved to be superior to the EKF in performance.[20,25]

The measurement system considered in this paper is similar to that proposed in ref. [16] where the EKF is employed as the state estimator. However, due to the measurement conversion based approach introduced in this paper, we can actually employ a robust version of the standard linear Kalman filter.[26–28] This linear robust filtering approach has proved to be an effective tool for many robust control and state estimation problems; see e.g. refs. [26–30]. Unlike most Taylor-series based algorithms, we can give a mathematically rigorous proof that the state estimation error is bounded with a certain probability while there is no mathematically rigorous analysis results concerning the EKF-based visual surveillance algorithms.

The remainder of this paper is organized as follows. In Section 2 we introduce the state-space target dynamic model considered in this paper. The approach considered in this paper permits a large class of linear dynamic (uncertain) system models. Section 3 contains the main results of this paper where we introduce the measurement transformation algorithm and the robust filtering algorithm. Moreover, in Section 3 we prove the estimation error is bounded. In Section 4 we present computer simulations as well as practical experiments that demonstrate the performance of the proposed algorithm. In the practical experiment we consider a the problem of tracking the real-world coordinates of an object being moved by a robot manipulator (arm). A conclusion is given in Section 5.

2. Object-Camera Dynamic Model

In the kinematic modelling of an object (target) and a tracker (camera) in a cartesian coordinate system the resulting dynamic system equation is linear. A comprehensive survey of dual body kinematic modelling is presented in ref. [20] and a basic principal approach is given in ref. [31] where only the translational kinematics were considered. Based on requirements of the specific application, rotational motion has been considered and the resultant, non-linear dynamic models have been used[15] in the kinematic parameter estimation. For the case of vision-based tracking, it is suffice to consider only the translational effects and the subsequent linear model[20] as no camera motion is engaged. Let the position of the target in each of the traditionally denoted x, y and z directions, and with respect to the camera-based coordinate system be $[x_1, x_2, x_3]' \in \mathbb{R}^3$ with $'$ denoting transposition. Let the velocity component in each direction be given by $[x_4, x_5, x_6]' \in \mathbb{R}^3$ and let the acceleration in each traditionally denoted x, y and z direction be given by $[x_7, x_8, x_9]' \in \mathbb{R}^3$. Hence, we can define $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]' \in \mathbb{R}^9$ such that the state evolves according to

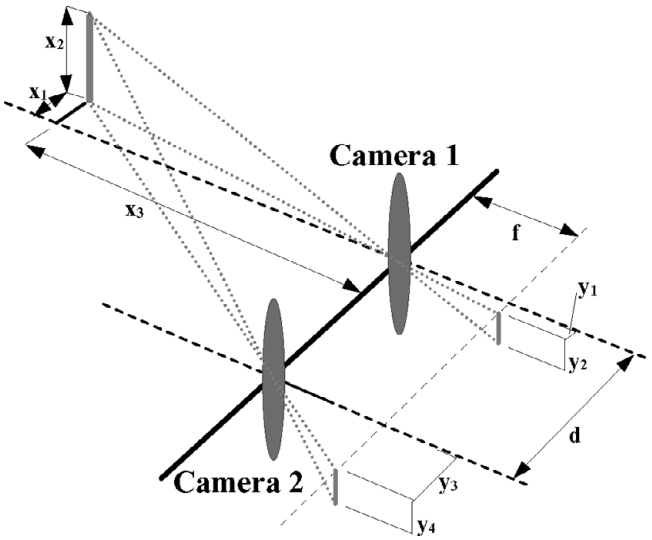

Fig. 1. The principle of perspective projection.

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}\mathbf{w}(k), \tag{1}$$

where $\mathbf{A}$ and $\mathbf{B}$ are suitably defined transition matrices[20] given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_3 & k_s\mathbf{I}_3 & \frac{k_s^2}{2}\mathbf{I}_3 \\ \mathbf{O}_3 & \mathbf{I}_3 & k_s\mathbf{I}_3 \\ \mathbf{O}_3 & \mathbf{O}_3 & \mathbf{I}_3 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{k_s^2}{2}\mathbf{I}_3 \\ k_s\mathbf{I}_3 \\ \mathbf{I}_3 \end{bmatrix} \tag{2}$$

and $\mathbf{w}(k) \in \mathbb{R}^3$ is an uncertainty parameter that encompasses the target's maneuvers and k_s is the sampling time. Our filtering algorithm is derived quite generally and permits a large class of linear dynamic models to be employed. If a point target is considered, then the target's position in $\mathbb{R}^3$ is projected onto the image plane of a suitably defined sensor via the principle of perspective projection. In Fig. 1 we can observe how a target's position is mapped from real $\mathbb{R}^3$ space onto the $\mathbb{R}^2$ image plane.

The principle of perspective projection provides a system of non-linear measurement equations that serve as the basis for the work in this paper. We also work with measurements of the velocity of those projected image plane points. Two sensors are used in a stereo-vision-based system with sensor

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

1 located at the origin of the global coordinate system and sensor 2 located a distance $d > 0$ away on the positive x_1-axis of sensor 1. We require the two sensor's image planes to be orientated in the same direction such that the two local x_3 directions (defined by each camera's local coordinate system; e.g. see Fig. 1) are parallel.

Remark 1. *The choice of coordinate basis is determined first by locating sensor* 1 *at the origin, second, by locating sensor* 2 *a distance* $d > 0$ *away on the positive* x_1*-axis of sensor 1 and such that the two positive* x_3 *directions of each camera's local coordinate basis are parallel. Then, we can define the direction of the horizontal, or equivalently, we choose orientations for the* x_1*-axis and* x_2*-axis, which are only determined up to a rotation by sensors* 1 *and* 2.

In traditional target tracking, the dynamics of a moving target are typically modelled in Cartesian coordinates and the resulting dynamic equations are linear; e.g. see ref. [20]. In refs. [15, 16] the targets are modelled via a dual translational/rotational motion model. The translational and rotational velocities are assumed to be constant. The resulting model is non-linear and adds to the complexity of the filter required. Here we consider a point target (or a number N of point features) that obey a linear dynamic model such as those described in ref. [20]. Any arbitrary number of point targets can be included in this model and object rigidity is not required since each point is tracked independently. However, the data association problem (also known as the feature point association problem)[32] exists in practice for tracking multiple point targets.

3. Linear Robust Filtering with Non-Linear Vision Measurements

In this section we outline the measurement model and the subsequent measurement conversion technique along with the robust linear filter which we derive as the state estimator. Throughout this paper we let $f > 0$ denote the focal length of the two cameras which is assumed to be the same and we let $d > 0$ denote the separation distance of the two cameras on the positive x-axis.

Let $[y_1(k)\ y_2(k)]'$ and $[y_3(k)\ y_4(k)]'$ denote the true values of the measured coordinates of the target point in the image plane of camera 1 and camera 2, respectively. That is, we have

$$\mathbf{y}_1(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ y_3(k) \\ y_4(k) \end{bmatrix} = f \begin{bmatrix} \dfrac{x_1(k)}{x_3(k)} \\ \dfrac{x_2(k)}{x_3(k)} \\ \dfrac{(x_1(k)-d)}{x_3(k)} \\ \dfrac{x_2(k)}{x_3(k)} \end{bmatrix}, \tag{3}$$

where $\mathbf{y}_1$ is simply the true values of the non-linear perspective projection based coordinates in the image planes of camera 1 and 2. Moreover, let $\widehat{\mathbf{y}}_1(k) = \mathbf{y}_1(k) + \mathbf{v}_1$ denote the noisy (actual) measured image coordinates of the target point where $\mathbf{v}_1 = [v_1, v_2, v_3, v_4]'$ are the corresponding measurement errors. Note that $y_2(k) = y_4(k)$ but that in general $\widehat{y}_2(k) \neq \widehat{y}_4(k)$. Hence, for notational simplicity let us define a new (noisy) measurement vector

$$\widehat{\boldsymbol{\psi}}_1(k) = \begin{bmatrix} \widehat{\psi}_1(k) \\ \widehat{\psi}_2(k) \\ \widehat{\psi}_3(k) \end{bmatrix} = \begin{bmatrix} \widehat{y}_1(k) \\ \dfrac{\widehat{y}_2(k) + \widehat{y}_4(k)}{2} \\ \widehat{y}_3(k) \end{bmatrix}, \tag{4}$$

where the true measured values of y_i, $\forall i \in \{1, 2, 3, 4\}$ are defined as before, i.e. in Eq. (3). Note that the error in $\widehat{\psi}_2(k) = \frac{\widehat{y}_2(k)+\widehat{y}_4(k)}{2}$ is now given by $\frac{v_2+v_4}{2}$.

Moreover, let $[y_5(k)\ y_6(k)]'$ and $[y_7(k)\ y_8(k)]'$ be the true values of the image coordinate velocities (between successive frames) in the planes of camera 1 and camera 2, respectively. Then we get the following measurement model

$$\mathbf{y}_2(k) = \begin{bmatrix} y_5(k) \\ y_6(k) \\ y_7(k) \\ y_8(k) \end{bmatrix} = f \begin{bmatrix} \dfrac{x_4(k)}{x_3(k)} - \dfrac{x_1(k)x_6(k)}{x_3(k)^2} \\ \dfrac{x_5(k)}{x_3(k)} - \dfrac{x_2(k)x_6(k)}{x_3(k)^2} \\ \dfrac{x_4(k)}{x_3(k)} - \dfrac{x_1(k)x_6(k)}{x_3(k)^2} + \dfrac{dx_6(k)}{x_3(k)^2} \\ \dfrac{x_5(k)}{x_3(k)} - \dfrac{x_2(k)x_6(k)}{x_3(k)^2} \end{bmatrix}, \tag{5}$$

where $\mathbf{y}_2$ is thus the true values of the time derivatives of the image plane coordinates given in Eq. (3) by $\mathbf{y}_1$. Again, we let $\widehat{\mathbf{y}}_2(k) = \mathbf{y}_2(k) + \mathbf{v}_2$ denote the noisy (actual) measured values where $\mathbf{v}_2 = [v_5, v_6, v_7, v_8]'$ are the corresponding measurement errors. Note that $y_6(k) = y_8(k)$ but that in general $\widehat{y}_6(k) \neq \widehat{y}_8(k)$. Hence, for notational simplicity let us define a new (noisy) measurement vector

$$\widehat{\boldsymbol{\psi}}_2(k) = \begin{bmatrix} \widehat{\psi}_4(k) \\ \widehat{\psi}_5(k) \\ \widehat{\psi}_6(k) \end{bmatrix} = \begin{bmatrix} \widehat{y}_5(k) \\ \dfrac{\widehat{y}_6(k) + \widehat{y}_8(k)}{2} \\ \widehat{y}_7(k) \end{bmatrix}, \tag{6}$$

where the true measured values of y_i, $\forall i \in \{5, 6, 7, 8\}$ are defined as before, i.e. in Eq. (5) and the error in $\widehat{\psi}_5(k) = \frac{\widehat{y}_6(k)+\widehat{y}_8(k)}{2}$ is now given by $\frac{v_6+v_8}{2}$.

Let $\widehat{\boldsymbol{\psi}}(k) = [\widehat{\boldsymbol{\psi}}_1(k), \widehat{\boldsymbol{\psi}}_2(k)]'$ such that in a noiseless environment it is clear that the true value of $\widehat{\boldsymbol{\psi}}(k)$ denoted by $\boldsymbol{\psi}(k)$ is simply a re-organization of the independent measurements in both Eqs. (3) and (5). This is because $y_2(k) = y_4(k)$ and $y_6(k) = y_8(k)$ implies that one of the true values from each pair offers no additional *information* when the values are error-free. However, in a noisy environment we find that $\widehat{y}_2(k) \neq \widehat{y}_4(k)$ and $\widehat{y}_6(k) \neq \widehat{y}_8(k)$ which means that $\widehat{\boldsymbol{\psi}}(k) = [\widehat{\boldsymbol{\psi}}_1(k), \widehat{\boldsymbol{\psi}}_2(k)]'$ provides a well-defined system of measurement equations (i.e. an equal number of equations as there is unknowns) with the added redundancy and noise tolerance of the additional measurements. Now assume that the target motion is described by (1) where the matrix $\mathbf{A}$ is

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

non-singular. Let $0 < p_0 \leq 1$ be a given constant and suppose that the system initial condition $\mathbf{x}(0)$, noise $\mathbf{w}(k)$ and the actual measurement noises $v_i(k)$, $\forall i \in \{1, \ldots, 8\}$ satisfy the following assumption.

Assumption 1. *The following inequalities with probability p_0 simultaneously hold:*

$$|v_i| \leq \epsilon |y_i| \ \forall i \in \{1, \ldots, 4\}, \quad |v_i| \leq \delta |y_i| \ \forall i \in \{5, \ldots, 8\}, \tag{7}$$

$$(\mathbf{x}(0) - \mathbf{x}_0)' \mathbf{N} (\mathbf{x}(0) - \mathbf{x}_0) + \sum_{0}^{T-1} \mathbf{w}(k)' \mathbf{Q}(k) \mathbf{w}(k) \leq \tau. \tag{8}$$

Here $\mathbf{x}_0$ is a given initial state estimate vector, $\mathbf{N} = \mathbf{N}'$ and $\mathbf{Q} = \mathbf{Q}'$ are given positive definite weighting matrices, $\tau > 0$ is a given constant associated with the system, and $T > 0$ is a given time.

The weighting matrices $\mathbf{N}$ and $\mathbf{Q}$ can be adjusted in order to compensate appropriately for the relative uncertainties. For example, lesser the initial state uncertainty, smaller the values of the weighting matrix $\mathbf{N}$ should be.

Using the preceding noisy measurement model $\widehat{\boldsymbol{\psi}}(k) = [\widehat{\boldsymbol{\psi}}_1(k), \widehat{\boldsymbol{\psi}}_2(k)]'$ we can define the converted measurement system as

$$\begin{bmatrix} \tilde{x}_1(k) \\ \tilde{x}_2(k) \\ \tilde{x}_3(k) \\ \tilde{x}_4(k) \\ \tilde{x}_5(k) \\ \tilde{x}_6(k) \end{bmatrix} = \begin{bmatrix} \dfrac{d\widehat{\psi}_1(k)}{\widehat{\psi}_1(k) - \widehat{\psi}_3(k)} \\ \dfrac{d\widehat{\psi}_2(k)}{\widehat{\psi}_1(k) - \widehat{\psi}_3(k)} \\ \dfrac{df}{\widehat{\psi}_1(k) - \widehat{\psi}_3(k)} \\ \dfrac{d(\widehat{\psi}_6(k)\widehat{\psi}_1(k) - \widehat{\psi}_4(k)\widehat{\psi}_3(k))}{(\widehat{\psi}_1(k) - \widehat{\psi}_3(k))^2} \\ \dfrac{d(\widehat{\psi}_2(k)(\widehat{\psi}_6(k) - \widehat{\psi}_4(k)) + \widehat{\psi}_5(k)(\widehat{\psi}_1(k)) - \widehat{\psi}_3(k))}{(\widehat{\psi}_1(k) - \widehat{\psi}_3(k))^2} \\ \dfrac{df(\widehat{\psi}_6(k) - \widehat{\psi}_4(k))}{(\widehat{\psi}_1(k) - \widehat{\psi}_3(k))^2} \end{bmatrix}, \tag{9}$$

where the $\tilde{x}_i(k)$ are the converted noisy measurements of the state components $x_i(k)$, $\forall i \in \{1, \ldots, 6\}$ which will be applied to the linearly formulated estimation algorithm to be derived. We denote the converted measurement vector as $\mathbf{m} = [\tilde{x}_1(k) \ \ \tilde{x}_2(k) \ \ \tilde{x}_3(k) \ \ \tilde{x}_4(k) \ \ \tilde{x}_5(k) \ \ \tilde{x}_6(k)]$. Immediately, we notice that we have not employed any Taylor-series based approximations in determining Eq. (9). We have in some regards transformed the non-linearities into the measurement errors that are ultimately associated with each of the $\tilde{x}_i(k)$. Essentially, each of the $\tilde{x}_i(k)$ are found by solving the equations in $\widehat{\boldsymbol{\psi}}(k) = [\widehat{\boldsymbol{\psi}}_1(k), \widehat{\boldsymbol{\psi}}_2(k)]'$ for the x_i as if they were noiseless. Of course, since they are not noiseless we find that $\tilde{x}_i(k)$ are corrupted by a non-additive and state-dependent error which we will subsequently try and correct.

Note that $\widehat{y}_i = y_i + v_i$ with $|v_i| \leq \epsilon |y_i|$ implies $\widehat{\psi}_i \leq \psi_i + \epsilon |\psi_i| \ \ \forall i \in \{1, \ldots, 4\}$ and $|v_i| \leq \delta |y_i|$ implies $\widehat{\psi}_i \leq \psi_i + \delta |\psi_i|$ for $\forall i \in \{5, \ldots, 8\}$. Indeed, these relationships are straightforward for $\widehat{\psi}_i$ with $i \in \{1, 3, 4, 6\}$. The error in $\widehat{\psi}_2(k) = \frac{\widehat{y}_2(k) + \widehat{y}_4(k)}{2}$ is given by $\frac{v_2 + v_4}{2}$ which clearly obeys either $|\frac{v_2 + v_4}{2}| \leq \epsilon |y_2|$ or $|\frac{v_2 + v_4}{2}| \leq \epsilon |y_4|$ since $y_2 = y_4$. Moreover, the error in $\widehat{\psi}_5(k) = \frac{\widehat{y}_6(k) + \widehat{y}_8(k)}{2}$ is given by $\frac{v_6 + v_8}{2}$ which clearly obeys either $|\frac{v_6 + v_8}{2}| \leq \delta |y_6|$ or $|\frac{v_6 + v_8}{2}| \leq \delta |y_8|$ since $y_6 = y_8$. Finally, we can easily deduce that $\widehat{y}_i - \widehat{y}_j$ for any $i, j \in \{1, 2, 3, 4\}$ implies that $v_i - v_j \leq \epsilon |y_i - y_j|$ and, similarly, $\widehat{y}_i - \widehat{y}_j$ for any $i, j \in \{5, 6, 7, 8\}$ implies that $v_i - v_j \leq \delta |y_i - y_j|$.

Our solution to the state estimation problem involves the following Riccati difference equation:

$$\begin{aligned} \mathbf{F}(k+1) &= \left[\hat{\mathbf{B}}' \mathbf{S}(k) \hat{\mathbf{B}} + \mathbf{I}\right]^{-1} \hat{\mathbf{B}}' \mathbf{S}(k) \hat{\mathbf{A}}, \\ \mathbf{S}(k+1) &= \hat{\mathbf{A}} \mathbf{S}(k) \left[\hat{\mathbf{A}} - \hat{\mathbf{B}} \mathbf{F}(k+1)\right] + \mathbf{C}\mathbf{C}' - \mathbf{K}'\mathbf{K}, \\ \mathbf{S}(0) &= \mathbf{N}. \end{aligned} \tag{10}$$

where $\hat{\mathbf{A}} \triangleq \mathbf{A}^{-1}$ and $\hat{\mathbf{B}} \triangleq \mathbf{A}^{-1}\mathbf{B}$. We also define

$$\mathbf{C} \triangleq \begin{bmatrix} \beta_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_6 & 0 & 0 & 0 \end{bmatrix}, \tag{11}$$

$$\mathbf{K} \triangleq \begin{bmatrix} \tilde{\alpha}_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{\alpha}_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\alpha}_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\alpha}_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{\alpha}_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \tilde{\alpha}_6 & 0 & 0 & 0 \end{bmatrix}. \tag{12}$$

where

$$\begin{aligned} \beta_1 &= \frac{1+\epsilon}{2(1-\epsilon)} + \frac{1-\epsilon}{2(1+\epsilon)}, \beta_3 = \frac{1}{2(1+\epsilon)} + \frac{1}{2(1-\epsilon)}, \\ \beta_2 &= \beta_1, \beta_4 = \frac{(1+\delta)(1+\epsilon)}{2(1-\epsilon)^2} + \frac{(1-\delta)(1-\epsilon)}{2(1+\epsilon)^2}, \\ \beta_5 &= \beta_4, \beta_6 = \frac{(1+\delta)}{2(1-\epsilon)^2} + \frac{(1-\delta)}{2(1+\epsilon)^2}, \end{aligned} \tag{13}$$

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

and

$$\tilde{\alpha}_1 = \frac{1+\epsilon}{2(1-\epsilon)} - \frac{1-\epsilon}{2(1+\epsilon)},\ \ \tilde{\alpha}_3 = \frac{1}{2(1+\epsilon)} - \frac{1}{2(1-\epsilon)},$$
$$\tilde{\alpha}_2 = \tilde{\alpha}_1,\ \ \tilde{\alpha}_4 = \frac{(1+\delta)(1+\epsilon)}{2(1-\epsilon)^2} - \frac{(1-\delta)(1-\epsilon)}{2(1+\epsilon)^2},$$
$$\tilde{\alpha}_5 = \tilde{\alpha}_4, \tilde{\alpha}_6 = \frac{(1+\delta)}{2(1-\epsilon)^2} - \frac{(1-\delta)}{2(1+\epsilon)^2}. \tag{14}$$

We now consider a set of state equations of the form

$$\begin{aligned}
\boldsymbol{\eta}(k+1) &= \left[\hat{\mathbf{A}} - \mathbf{F}(k+1)\right]' \boldsymbol{\eta}(k) + \mathbf{C}'\mathbf{m}(k+1),\\
\boldsymbol{\eta}(0) &= \mathbf{N}\mathbf{x}_0,\\
g(k+1) &= g(k) + \mathbf{m}(k+1)'\mathbf{m}(k+1) -\\
&\qquad \boldsymbol{\eta}(k)'\hat{\mathbf{B}}\left[\hat{\mathbf{B}}'\mathbf{S}(k)\hat{\mathbf{B}} + \mathbf{Q}(k)\right]^{-1}\hat{\mathbf{B}}'\boldsymbol{\eta}(k),\\
g(0) &= \mathbf{x}_0'\mathbf{N}\mathbf{x}_0.
\end{aligned} \tag{15}$$

The above state Eq. (15) and Riccati Eq. (11) can simply be thought of as a robust implementation of the standard linear Kalman Filter[33] for uncertainties obeying Assumption 3.1, e.g. see refs. [28, 33, 34]. Now we are in a position to present the main result of this section.

Theorem 1. *Let $0 < p_0 \leq 1$ be given, and suppose that Assumption 1 holds. Then the state $\mathbf{x}(T)$ of the system (1) with probability p_0 belongs to the ellipsoid*

$$E_T \stackrel{\triangle}{=} \left\{ \begin{array}{l} \mathbf{x}_T \in \mathbf{R}^n : \\ \|(\mathbf{S}(T)^{\frac{1}{2}}\mathbf{x}_T - \mathbf{S}(T)^{-\frac{1}{2}}\boldsymbol{\eta}(T))\|^2 \\ \leq \rho + \tau \end{array} \right\}, \tag{16}$$

where

$$\rho \stackrel{\triangle}{=} \boldsymbol{\eta}(T)'S(T)^{-1}\boldsymbol{\eta}(T) - g(T),$$

and $\boldsymbol{\eta}(T)$ and $g(T)$ are defined by the Eq. (15). Also, we require $\rho + \tau \geq 0$.

Proof 1. *It follows from Eqs. (9) and (7) that*

$$\tilde{x}_i(k) = \beta_i x_i(k) + n_i(k), \tag{17}$$

$\forall i \in \{1, \ldots, 6\}$ and where $x_i(k)$ is the i^{th} component of the state vector $\mathbf{x}(k)$ of the system (1) and the inequalities

$$|n_i(k)| \leq \tilde{\alpha}_i |x_i(k)|, \tag{18}$$

hold together with Eq. (8) with probability p_0. Therefore, Eq.(18) implies that

$$\mathbf{m}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{n}(k), \tag{19}$$

where $\mathbf{n}(k) \stackrel{\triangle}{=} [n_1(k)\ n_2(k)\ n_3(k)\ n_4(k)\ n_5(k)\ n_6(k)]'$ and the condition

$$\|\mathbf{n}(k)\|^2 \leq \|\mathbf{K}\mathbf{x}\|^2, \tag{20}$$

holds together with Eq. (8) with probability p_0. From Eqs. (8) and (20) we obtain that the following sum quadratic constraint is satisfied

$$\begin{aligned}
&(\mathbf{x}(0) - \mathbf{x}_0)'\mathbf{N}(\mathbf{x}(0) - \mathbf{x}_0) +\\
&\sum_0^{T-1} \left(\mathbf{w}(k)'\mathbf{Q}(k)\mathbf{w}(k) + \|\mathbf{n}(k+1)\|^2\right)\\
&\leq \tau + \sum_0^{T-1} \|\mathbf{K}\mathbf{x}\|^2,
\end{aligned} \tag{21}$$

with probability p_0. Now it follows from Theorem 5.3.1 of ref. [28], p. 75 (see also ref. [34]) that the state $\mathbf{x}(T)$ of the system (1), Eq. (19) belongs to the ellipsoid (16) with probability p_0.

A point value state estimate can be obtained from the bounded ellipsoidal set's center and is given by $\widehat{\mathbf{x}} = \mathbf{S}(k)^{-1}\boldsymbol{\eta}(k)$.

We have therefore proved our algorithm's estimation errors are bounded in a probabilistic sense when the relevant uncertainties obey Assumption 1. The sum quadratic constraint given in Assumption 1 accommodates a large class of non-linear and dynamic process noise characteristics. As the Gaussian noise is bounded within the first standard deviation with a probability $p_0 \approx 0.68$ and within two standard deviations with probability $p_0 \approx 0.95$ etc, we lose no generality by considering uncertainties satisfying Assumption 1. That is, Gaussian measurement, process and initial condition errors form special cases of Assumption 1 which defines a larger class of uncertainties. We solve the problem in the linear domain and our algorithm permits very large initial errors. No similar proofs exist for the EKF or the majority of other approaches that employ some form of Taylor-series based approximation. Indeed, the fact that we can prove bounded tracking performance with arbitrarily large initial condition errors is a novel contribution.

4. Illustrative Examples

In this section a number of examples are examined using both simulated data and real physical experiments.

4.1. Fictional simulation data

In the first case we consider fictitious simulation data of a single mobile target. Consider a camera that is located at the origin of the global coordinate system and a second camera that is located a distance d on the positive x_1-axis of the first camera. We assume the background is stationary and the target stays within both camera's fields of view which are determined by the focal length and frame size of both cameras. In all cases the focal length f for both cameras is assumed to be unitary. We consider a maneuverable target that obeys the following dynamic model for simplicity

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}\mathbf{w}(k), \tag{22}$$

where the transition matrices $\mathbf{A}$ and $\mathbf{B}$ are based on the well-known Wiener-sequence acceleration model of ref. [20].

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

Table I. Simulation parameters.

Input	Value	Comments
$\mathbf{w}$	$\sigma_w = 0.001$	Gaussian accel. Input
$v_i, i \in 1, 2, 3, 4$	$\sigma_{vi} = 0.005$	Gaussian meas. Noise 1
$v_i, i \in 5, 6, 7, 8$	$\sigma_{vi} = 0.005$	Gaussian meas. Noise 2
ϵ	$\epsilon = 0.002$	Assumed meas. Bound 1
δ	$\delta = 0.002$	Assumed meas. Bound 2
$[\mathbf{N}_R, \ \mathbf{Q}_R]$	$[1, \ 1] \cdot \mathbf{I}$	Robust filter parameters
$[\mathbf{R}_E, \ \mathbf{Q}_E]$	$[0.005, \ 0.001^2] \cdot \mathbf{I}$	EKF uncertainty weightings
$T \ @ \ t_s$	120 s @ 1 s	Track duration and periodicity

The process noise $\mathbf{w}(k)$ is a white Gaussian random vector representing the targets acceleration input.

For each simulation we analyse the results of 1000 simulation runs and compare the performance of the robust filtering technique and the EKF. We plot the root mean squared (RMS) position and velocity errors

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left[\Phi_i^T \Phi_i\right]},$$

where M is the total number of simulation runs, i indicates the ith run and $\Phi = [x_1 - \widehat{x}_1 \ \ x_2 - \widehat{x}_2 \ \ x_3 - \widehat{x}_3]'$ or $\Phi = [x_4 - \widehat{x}_4 \ \ x_5 - \widehat{x}_5 \ \ x_6 - \widehat{x}_6]'$. The measurement noise in this paper is consistent with the simulations and discussion in ref. [16].

The simulation parameters are given in Table I. Both filter's initializations are Gaussian distributed about the true initial state with $\sigma = 5\%$ of the true initial values. The EKF is known to be potentially unstable without correct initialization. The true initial state is $[-100 \ 250 \ 1 \ -1 \ 0.0001 \ 0.0001]'$.

The EKF parameters, i.e. $\mathbf{Q}_E$ and $\mathbf{R}_E$, were tuned fairly accurately. The initial covariance of the EKF is also tuned assuming the initial error statistics are known to the tracking system. That is, for the EKF parameters we assumed perfect knowledge of all the relevant error statistics and tuned around these true values in order to get the best performance. On the other hand, for the RF we simply used the identity matrix for both the initial and process noise weighting. For the RF, α_1 and α_2 are taken as two times the first standard deviations of the Gaussian measurement noise.

It may be reasonable to assume (at least partial) knowledge of the video measurement's uncertainty statistics due to the routine sensor testing and calibration operations performed. However, it is dangerous to assume knowledge of the target's uncertainty statistics which characterize it's maneuvers.

We plot the surface of RMS position and velocity errors over 1000 simulation runs and for camera separation distances $d \in [3, 10]$ units for both the robust linear filter and the EKF. The RMS position error for the robust approach is given in Fig. 2 and the EKF approach is given in Fig. 3.

From Fig. 2 and 3 it is clear that the robust filtering approach outperforms the EKF significantly in tracking the target position. The linear robust approach is potentially sensitive to the separation distance d between the two

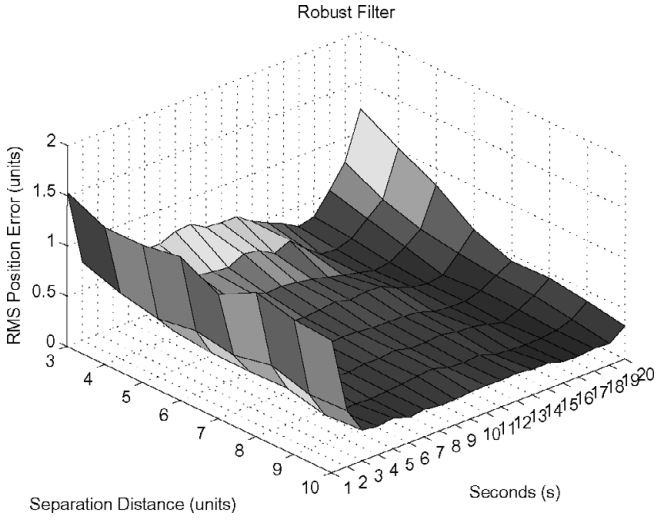

Fig. 2. The RMS position error surface for the RF.

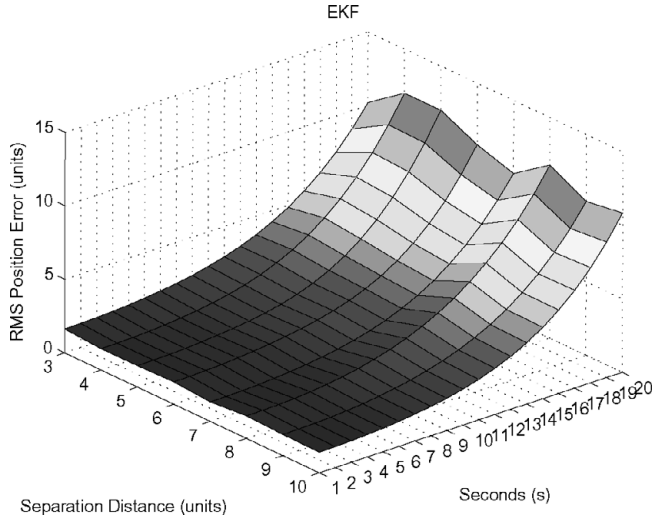

Fig. 3. The RMS position error surface for the EKF.

cameras and we see from Fig. 2 that as the separation distance increases, the RF's performance improves slightly. The EKF position estimate slowly diverges when tracking the maneuverable target.

The RMS velocity error for the robust approach is given in Fig. 4 and the EKF approach is given in Fig. 5.

From Figs. 4 and 5 it is similarly clear that the robust filtering approach outperforms the EKF significantly in tracking the target velocity as the separation distance d increases. It can be seen that for the proposed robust linear filter the estimation accuracy increases as the separation distance increases. The EKF velocity estimate slowly diverges when tracking the maneuverable target regardless of the separation distance. The proposed linear RF uses the measurement conversion technique which is, in fact, computation of 3D coordinates of a target or a closed-form solution. Then, our robust estimator improves the accuracy. In contrast, the EKF does not contain any computation of 3D coordinates of a target and is based on linearization and Taylor series approximations. It causes the accumulation of errors resulting in divergence for large uncertainties.

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

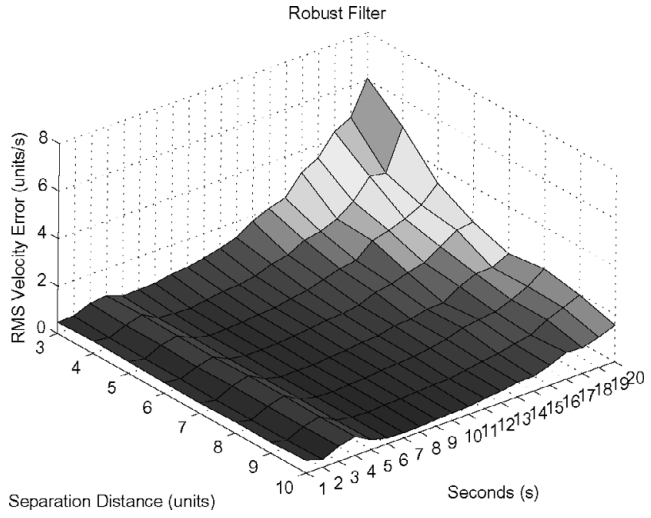

Fig. 4. The RMS velocity error surface for the RF.

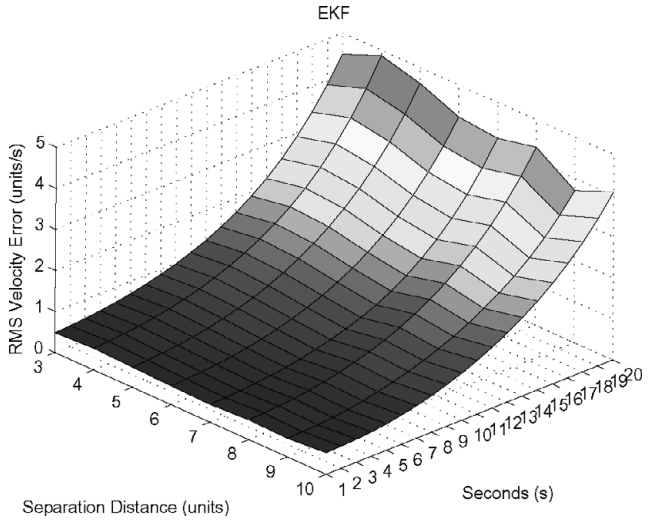

Fig. 5. The RMS velocity error surface for the EKF.

4.2. Real Experimental Data

In this second simulation sub-section we examine a simple practical vision-based tracking problem in order to illustrate that our algorithm is feasible using real vision sensors and real moving objects.

Figure 6 shows the actual locations of the end effector of the robotic arm. Figure 7 shows the estimated path using the stereo vision ideas. Using the object location in the image plane($[y_1(k)\ y_2(k)]'$ and $[y_3(k)\ y_4(k)]'$ measurements) of the two camera's as recursive measurements, we deduce the converted 3D locations(Eq. 9) as well as the estimated position from RF and EKF. It is quite evident that the RF out performs the extended Kalman filter. In fact the extended Kalman filter diverges. Further the converted measurements are improved due to the robustness of the filter. Figure 8 shows the error in comparison to the actual path data obtained using robot co-ordinate readings. Although there seems a noticeable inaccuracy due to alignments etc in the converted measurements the RF improves this measurements. We have used 20 pixel/mm and identical cameras with focal length of 50 mm. The distance between the two cameras, $d = 360$ mm

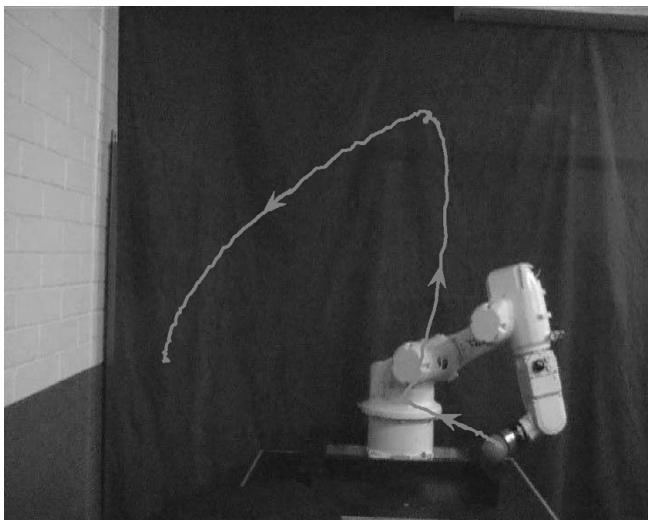

Fig. 6. End effector measured path from the video sequence.

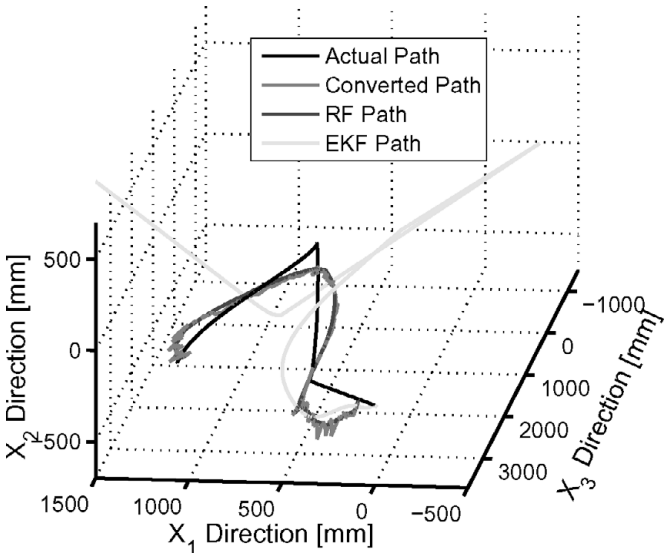

Fig. 7. Estimated and actual paths.

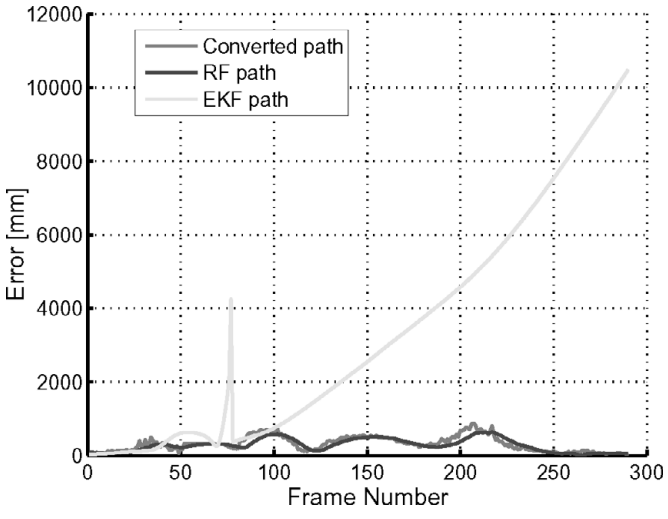

Fig. 8. Errors in RF and EKF.

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press

and the duration of the experiment was 36 s with a frame rate of 23.5.

5. Conclusions

In this paper we derived a linear state estimator with provable performance limits for vision-based surveillance and object tracking using non-linear perspective projection and image velocity measurements. We use a novel measurement conversion approach that does not use Taylor-series approximation and allows us to derive a completely linear algorithm. A significant contribution of this technique is the mathematically rigorous proof of the boundedness of the filtering error. No such results are known for the extended Kalman filter.

Acknowledgments

This work was supported by the Australian Research Council (ARC).

References

1. J. M. Hespanha, O. A. Yakimenko, I. I. Kaminer and A. M. Pascoal, "Linear parametrically varying systems with brief instabilities: An application to vision/inertial navigation," *IEEE Trans. Aerosp. Electron. Syst*. **40**(3), 889–902 (2004).
2. E. M. P. Low, I. R. Manchester and A. V. Savkin, "A biologically inspired method for vision based docking of mobile robots," *Robot. Auton. Syst.* **55**(10), 769–784 (Oct. 2007).
3. I. R. Manchester, A. V. Savkin and F. A. Faruqi, "A method for optical-based precision missile guidance," *IEEE Trans. Aerosp. Electron. Syst*. **44**(3), 835–851 (2008).
4. I. Horswill, "Visual Collision Avoidance by Segmentation," *IEEE International Conference on Intelligent Robots and Systems*, Munich, Germany (1994) pp. 902–909.
5. D. Pomerleau and T. Jochem, "Rapidly adapting machine vision for automated vehicle steering," *IEEE Expert* **11**(2), 19–27 (1996).
6. M. Nashman and H. Schneiderman, "Real-time visual processing for autonomous driving," *IEEE Symp. Intell. Veh.* (1993) pp. 373–378.
7. B. K. Horn, *Robot Vision* (McGraw-Hill Higher Education, New York, NY, 1986).
8. D. A. Pomerleau, *Neural Network Perception for Mobile Robot Guidance* (Kluwer Academic, Boston, MA, 1994).
9. P. Gurfil, "Robust guidance for electro-optical missiles," *IEEE Trans. Aerosp. Electron. Syst.* **39**(2) (Apr. 2003).
10. V. Malyavej, I. R. Manchester and A. V. Savkin, "Precision missile guidance using radar/multiple-video sensor fusion via communication channels with bit-rate constraints," *Automatica* **42**(5), 763–769 (2006).
11. G. L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino and T. Ellis, "Active video-based surveillance system," *IEEE Signal Process. Mag.* **22**(2), 25–37 (Mar. 2005).
12. M. Shibata and N. Kobayashi, Image-based visual tracking for moving targets with active stereo vision robot. *Proceedings of the SICE-ICASE International Joint Conference 2006*, Bexco, Busan, Korea (Oct. 18–21, 2006) pp. 5329–5334.
13. P. I. Corke and M. C. Good, "Dynamic effects in visual closed-loop systems," *IEEE Trans. Robot. Autom.* **12**(5), 671–683 (1996).
14. M. R. Driels and U. S. Pathre, "Vision-based automatic theodolite for robot calibration," *IEEE Trans. Robot. Autom.* **7**(3), 351–360 (1991).
15. T. J. Broida, S. Chandrashekhar and R. Chellappa, "Recursive 3-D motion estimation from a monocular image sequence," *IEEE Trans. Aerosp. Electron. Syst.* **26**(4), 639–656 (Jul. 1990).
16. S. D. Blostein, L. Zhao and R. M. Chann, "Three-dimensional trajectory estimation from image position and velocity," *IEEE Trans. Aerosp. Electron. Syst.* **36**(4), 1075–1089 (Oct. 2000).
17. P. N. Pathirana, A. Lim, A. V. Savkin and P. D. Hodgson, "Robust video/ultrasonic fusion based estimations for automotive applications," *IEEE Trans. Veh. Technol.* **56**(4), 1631–1639 (2005).
18. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *DARPA Image Understanding Workshop, DARPA* (1981) pp. 121–130. Washington, DC, USA.
19. T. A. Murat, *Digital Video Processing* (Prentice Hall PTR, Upper Saddle River, NJ, 1995).
20. X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking part I: Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.* **39**(4), 1333–1364 (2003).
21. W. Respondek, A. Pogromsky and H. Nijmeijer, "Time scaling for observer design with linearizable dynamics," *Automatica* **40**(2), 277–285 (2004).
22. S. Huan and G. Dissanayake, "Convergence and consistency analysis for extended Kalman filter based SLAM," *IEEE Trans. Robot.* **23**(5), 1036–1049 (2007).
23. S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE* **92**(3) (Mar. 2004).
24. R. M. Chann, Recursive Estimation of 3-D Motion and Structure in Image Sequences Based on Measurement Transformations *Master's Thesis* (Kingstone, Ontario, Canada: Queen's University, 1994).
25. Z. Zhao, X. R. Li, and V. P. Jilkov, "Best linear unbiased filtering with nonlinear measurements for target tracking," *IEEE Trans. Aerosp. Electron. Syst.* **40**(4), 1324–1336 (2004).
26. A. V. Savkin and I. R. Petersen, "Recursive state estimation for uncertain systems with an integral quadratic constraint," *IEEE Trans. Autom. Control* **40**(6), 1080 (1995).
27. A. V. Savkin and I. R. Petersen, "Model validation for robust control of uncertain systems with an integral quadratic constraint," *Automatica* **32**(4), 603–606 (1996).
28. I. R. Petersen and A. V. Savkin, *Robust Kalman Filtering for Signals and Systems with Large Uncertainties* (Birkhauser, Boston, MA, 1999).
29. I. R. Petersen, V. A. Ugrinovskii and A. V. Savkin, *Robust Control Design Using H^∞ Methods* (Springer-Verlag, London, UK, 2000).
30. A. V. Savkin and R. J. Evans, *Hybrid Dynamical Systems. Controller and Sensor Switching Problems*. (Birkhauser, Boston, MA, 2002).
31. A. V. Savkin, P. N. Pathirana and F. A. Faruqi, "The problem of precision missile guidance : LQR and H^∞ control framework," *IEEE Trans. Aerosp. Electron. Syst.* **39**(3), 901–910 (2003).
32. Y. Bar-Shalom and X. R. Li, *Estimation and Tracking Principles, Techniques and Software*. (Artech, Norwood, MA, 1993).
33. B. D. O Anderson and J. B. Moore, *Optimal Filtering*. (Prentice Hall, Englewood Cliffs, NJ, 1979).
34. A. V. Savkin and I. R. Petersen, "Robust state estimation and model validation for discrete-time uncertain systems with a deterministic description of noise and uncertainty," *Automatica* **34**(2), 271–274 (1998).

https://doi.org/10.1017/S0263574709005827 Published online by Cambridge University Press