Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008



# 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

***Abstract*— Vision based tracking of an object using the ideas of perspective projection inherently consists of nonlinearly modelled measurements although the underlying dynamic system that encompasses the object and the vision sensors can be linear. Based on a necessary stereo vision setting, we introduce an appropriate measurement conversion techniques which subsequently facilitate using a linear filter. Linear filter together with the aforementioned measurement conversion approach conforms a robust linear filter that is based on the set values state estimation ideas; a particularly rich area in the robust control literature. We provide a rigorously theoretical analysis to ensure bounded state estimation errors formulated in terms of an ellipsoidal set in which the actual state is guaranteed to be included to an arbitrary high probability.**

**Using computer simulations as well as a practical implementation consisting of a robotic manipulator, we demonstrate our linear robust filter significantly outperforms the traditionally used extended Kalman filter under this stereo vision scenario.**

## I. 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. 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. For these range of reasons, vision-based tracking systems have found application in a number of diverse application areas, e.g. see [1]–[8].

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 [9] for instance, a recursive target tracking filter is developed that uses the nonlinear perspective projection measurements as the input to an iterated EKF (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 [10]. In [10] 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 nonlinear equations that include measures of the targets velocity as well as position, hence the use of the EKF in [10]. In [11], a robust version of the extended kalman filter (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 [4], [12], [13] 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 algorithm that we derive. Essentially, we analytically convert the nonlinear measurement equations into linear measurements and apply a robust filter. Hence, we solve the estimation problem strictly within the linear domain since we also consider linear state equations [14]. Traditionally, it has been common to use non-linear estimators such as the extended Kalman filter (EKF) which employ some form of numerical approximation (e.g. Taylor-series). Thus, very few results exist which give analytical analysis of convergence properties and estimation error bounds. 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 [15] often leading to divergence. In [10] it is stated that after a detailed comparison of the EKF, IEKF, and an iterated linear filter smoother in [16] similar performances are observed for the problem of vision-based target tracking.

This work was supported by Australian Reseach Council(ARC)

P.N. Pathirana, A.N. Bishop, S.W. Ekanayake and T.J. Black are with School of Engineering and IT, Deakin University, Victoria 3217, Australia. pubudu@deakin.edu.au

A.V. Savkin is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney 2052, Australia.

978-1-4244-3124-3/08/$25.00 ©2008 IEEE 

Alternatively, measurement conversion methods have been explored particularly for target tracking with radar measurements [14], [17], [18]. The basic idea of these techniques is to transform nonlinear measurements 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 [19]. The measurement conversion methods have proved to be superior to EKF in performance [14], [17]–[19].

The overall measurement system is somewhat similar to that proposed in [10] where the EKF is employed as the state estimator. However, due to the measurement conversion based approach we employ a robust version of the standard linear Kalman filter from [20]–[22]. This linear robust filtering approach has proved to be an effective tool for many robust control and state estimation problems; see e.g. [20]–[24] while the measurement conversion ideas have proven to outperform EKF [17], [25]. 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 any mathematically rigorous analysis results on the EKF nor on the existing measurement conversion based approaches.

The underlying state estimation ideas we present are based on bounded deterministic measurement noise as opposed to Gaussian noise models which is an integral part in Kalman filtering. Less restrictive bounded assumptions coupled with more effective linear filtering due to measurement conversion provide a stronger conceptual and practice basis for state estimation particularly in a more uncertain vision based setting. In addition, well known tuning difficulties and requirement of accurate initial estimates associated with Kalman filtering is directly addressed via this robust approach. In particular, the initial condition uncertainty-omnipresent in vision based estimation, is modelled in the uncertainty constraint included in the worst case design paradigm. Further, this formulation encapsulate the state estimation to a ellipsoidal bounding set based on the assumed overall uncertainty and guarantees to consists the true state to an arbitrary high probability [22]. Although the resulting robust state estimator is suboptimal essentially due to relaxed noise model assumptions, the knowledge of the bounded uncertainty allows easy tuning the set values to a higher degree of precision.

The remainder of this paper is organized as follows : In section II we introduce the state space formulation which provides the basis for the non-linear measurement model derived from the well known perspective projection ideas. Section III, provides main results of this paper concerning the boundedness of the filter error together with the converted measurement of the non-linear measurements deriving filter parameters for the robust filter. Underlying parameters are presented in the form of standard uncertainty conditions in a set value estimation setting. In section IV, we present computer simulations as well as state estimation of an object being moved by a robotic manipulator in a stereo vision setting to demonstrate the effectiveness of the robust filter while the concluding remarks are given in section V.

## II. 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 [17] and a basic principal approach is given in [26] 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 [9] 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 [17] 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 $[x_1, x_2, x_3]' \in \mathbb{R}^3$. 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

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

where $\mathbf{A}$ and $\mathbf{B}$ are suitably defined transition matrices [14] given by

$$\mathbf{A} = \begin{bmatrix} I_3 & k_s I_3 & \frac{k_s^2}{2} I_3 \\ O_3 & I_3 & k_s I_3 \\ O_3 & O_3 & I_3 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} \frac{k_s^2}{2} I_3 \\ k_s I_3 \\ I_3 \end{bmatrix} \quad (2)$$

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 Figure 1 we can observe how a target's position is mapped from real $\mathbb{R}^3$ space onto the $\mathbb{R}^2$ image plane.

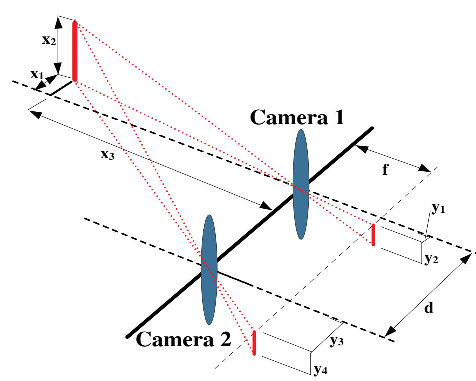


Fig. 1. The principle of perspective projection.

The principle of perspective projection provides a system of nonlinear measurements 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 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. Figure 1) are parallel. This requirement simplifies the subsequent derivations and presentation significantly. However, it is possible to extend the algorithms derived here to arbitrarily oriented camera systems involving multiple $\geq 2$ vision sensors.

*Remark 1:* The choice of coordinate basis is determined first by locating sensor 1 at the origin, secondly 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 modeled in Cartesian coordinates and the resulting dynamic equations are linear, e.g. see [14]. In [9], [10] the targets are modeled via a dual translational/rotational motion model. The translational and rotational velocities are assumed to be constant. The resulting model is nonlinear 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 [14]. 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) [27], [28] exists in practice for tracking multiple point targets.

## III. LINEAR ROBUST FILTERING WITH NONLINEAR 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,

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

where $\mathbf{y}_1$ is simply the true values of the nonlinear 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

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

where the true measured values of $y_i$, $\forall i \in \{1,2,3,4\}$ are defined as before, i.e. in equation (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

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

where $\mathbf{y}_2$ is thus the true values of the time derivatives of the image plane coordinates given in (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

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

where the true measured values of $y_i$, $\forall i \in \{5,6,7,8\}$ are defined as before, i.e. in equation (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 (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 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:

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

$$(\mathbf{x}(0)-\mathbf{x}_0)'\mathbf{N}(\mathbf{x}(0)-\mathbf{x}_0) + \sum_{0}^{T-1} \mathbf{w}(k)'\mathbf{Q}(k)\mathbf{w}(k) \le d. \quad (8)$$

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, $d > 0$ 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, given perfect initial state knowledge then the weighting matrix $\mathbf{N}$ should be given by $\mathbf{N} = \mathbf{I}$.

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

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

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 (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 $|v_i| \le \epsilon|y_i|$, $\forall i \in \{1,\ldots,4\}$ or $|v_i| \le \delta|y_i|$, $\forall i \in \{5,\ldots,8\}$ implies that any $\widehat{\psi}_i \in \widehat{\boldsymbol{\psi}}_1$ obeys $\widehat{\psi}_i \le \psi_i + \epsilon|\psi_i|$ and any $\widehat{\psi}_i \in \widehat{\boldsymbol{\psi}}_2$ obeys $\widehat{\psi}_i \le \psi_i + \delta|\psi_i|$. 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 $|\frac{v_2+v_4}{2}| \le \epsilon|y_2|$ or $|\frac{v_2+v_4}{2}| \le \epsilon|y_4|$ 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 $|\frac{v_6+v_8}{2}| \le \delta|y_6|$ or $|\frac{v_6+v_8}{2}| \le \delta|y_8|$ since $y_6 = y_8$. Finally, we can easily find that $\widehat{y}_i - \widehat{y}_j$ for any $i,j \in \{1,2,3,4\}$ implies that $v_i - v_j \le \epsilon|y_i - y_j|$ and similarly $\widehat{y}_i - \widehat{y}_j$ for any $i,j \in \{5,6,7,8\}$ implies that $v_i - v_j \le \delta|y_i - y_j|$.

Therefore, our solution to the state estimation problem involves the following Riccati difference equation,

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

where $\hat{\mathbf{A}} \triangleq \mathbf{A}^{-1}$ and $\hat{\mathbf{B}} \triangleq \mathbf{A}^{-1}\mathbf{B}$ and where we define

$$\mathbf{C} \triangleq \begin{bmatrix} \beta_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_6 & 0 & 0 & 0 \end{bmatrix} \quad (11)$$

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

Furthermore, let us define

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

and

$$\begin{aligned} \tilde{\alpha}_1 &= \frac{1+\epsilon}{2(1-\epsilon)} - \frac{1-\epsilon}{2(1+\epsilon)}, \quad \tilde{\alpha}_3 = \frac{1}{2(1+\epsilon)} - \frac{1}{2(1-\epsilon)} \\ \tilde{\alpha}_2 &= \tilde{\alpha}_1, \quad \tilde{\alpha}_4 = \frac{(1+\delta)(1+\epsilon)}{2(1-\epsilon)^2} - \frac{(1-\delta)(1-\epsilon)}{2(1+\epsilon)^2} \\ \tilde{\alpha}_5 &= \tilde{\alpha}_4, \tilde{\alpha}_6 = \frac{(1+\delta)}{2(1-\epsilon)^2} - \frac{(1-\delta)}{2(1+\epsilon)^2} \end{aligned} \quad (14)$$

Also, we consider a set of state equations of the form

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

The above state equations (15) and Riccati equations (10) can simply be thought of as a robust implementation of the standard linear Kalman Filter [29] for uncertainties obeying Assumption 1, e.g. see [22], [29], [30]. Now we are in a position to present the main result of this section.

*Theorem 1:* Let $0 < p_0 \le 1$ 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

$$E_T \triangleq \left\{ \begin{array}{l} \mathbf{x}_T \in \mathbf{R}^n : \\ \|(\mathbf{S}(T)^{\frac{1}{2}}\mathbf{x}_T - \mathbf{S}(T)^{-\frac{1}{2}}\boldsymbol{\eta}(T))\|^2 \\ \leq \rho + d \end{array} \right\} \quad (16)$$

where

$$\rho \triangleq \boldsymbol{\eta}(T)' S(T)^{-1} \boldsymbol{\eta}(T) - g(T)$$

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

*Proof:* The proof will be provided in an extended version of the paper ■

*Corollary 1:* A so-called 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. Indeed, Gaussian measurement, process and initial condition errors form special cases of Assumption 1. We solve the problem in the linear domain and our algorithm permits very large initial errors. No similar proofs exist for the extended Kalman filter (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.

*Remark 2:* It is well known that 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. Thus, we lose no generality by considering uncertainties satisfying Assumption 1. Indeed, there are systems in place in many practical frameworks to remove large Gaussian outliers (e.g. gating etc.). Furthermore, in many vision systems, the assumption of a bounded uniformly distributed measurement error is more realistic than a Gaussian error assumption. Finally, we point out that the system uncertainty $\mathbf{w}(k)$ in (1) may be better represented by an unknown deterministic uncertainty as compared to a white Gaussian random variable.

## IV. ILLUSTRATIVE EXAMPLES

In this section we present an example using a real physical experiment.

### A. Real Experimental Data

In this second simulation subsection 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 2 shows the actual locations of the end effector of the robotic arm. Figure 3 shows the estimated path using the stereo vision ideas. It is quite evident that the robust filter 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 4

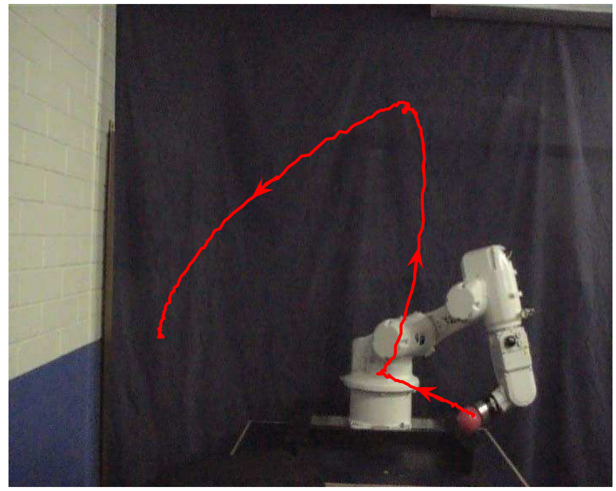

Fig. 2. End effector measured path from the video sequence

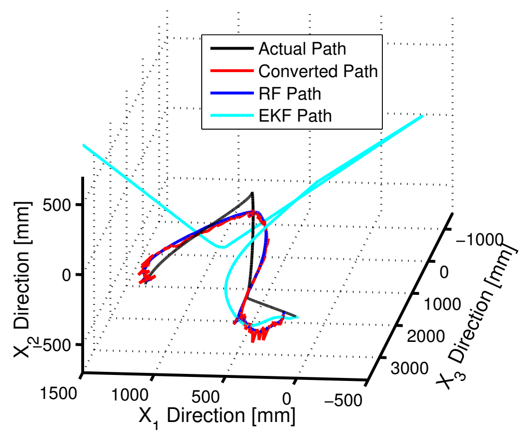


Fig. 3. Estimated and actual paths

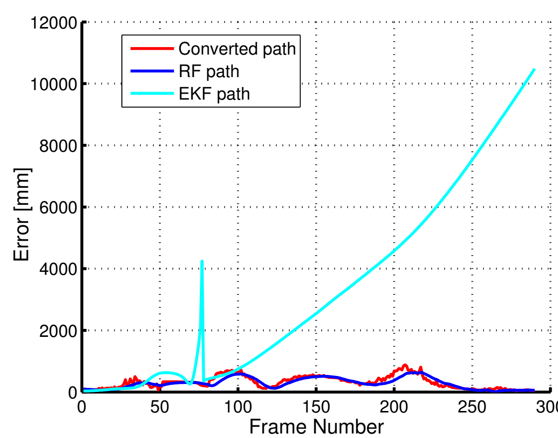


Fig. 4. Errors in RF and EKF

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 robust filter improves this measurements.

We have used 20 pixel/ml and identical cameras with focal length of 50mm. The distance between the two cameras $d = 360mm$ and the duration of the experiment was 36s with a frame rate of 23.5.

## V. CONCLUSIONS

In this paper we derived a linear state estimator with provable performance limits for vision-based surveillance and object tracking using nonlinear 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 approach of the boundedness of the filtering error*. No such results can be easily given for the extended Kalman filter.

## REFERENCES


[1] I. Horswill. Visual collision avoidance by segmentation. In *IEEE International Conference on Intelligent Robots and Systems*, Munich, Germany, 1994.

[2] D. Pomerleau and T. Jochem. Rapidly adapting machine vision for automated vehicle steering. *IEEE Expert*, 11(2):19–27, 1996.

[3] M. Nashman and H. Schneiderman. Real-time visual processing for autonomous driving. In *IEEE Symposium on Intelligent Vehicles*, 1993.

[4] B.K. Horn. *Robot Vision*. McGraw-Hill Higher Education, New York, NY, 1986.

[5] D.A. Pomerleau. *Neural Network Perception for Mobile Robot Guidance*. Kluwer Academic, Boston, 1994.

[6] P. Gurfil. Robust guidance for electro-optical missiles. *IEEE Transactions on Aerospace and Electronic Systems*, 39(2), April 2003.

[7] 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.

[8] G. L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis. Active video-based surveillance system. *IEEE Signal Processing Magazine*, 22(2):25–37, March 2005.

[9] T.J. Broida, S. Chandrashekhar, and R. Chellappa. Recursive 3-D motion estimation from a monocular image sequence. *IEEE Transactions on Aerospace and Electronic Systems*, 26(4):639–656, July 1990.

[10] S.D. Blostein, L. Zhao, and R.M. Chann. Three-dimensional trajectory estimation from image position and velocity. *IEEE Transactions on Aerospace and Electronic Systems*, 36(4):1075–1089, October 2000.

[11] P.N Pathirana, A Lim, A.V. Savkin, and P.D. Hodgson. Robust video/ultrasonic fusion based estimations for automotive applications. *IEEE Transactions on Vehicle Technology*, 56(4):1631–1639, 2005.

[12] B Lucas and T Kanade. An iterative image registration technique with an application to stereo vision. pages 121–130. DARPA Image understanding Workshop, DARPA, 1981.

[13] T.A Murat. *Digital Video Processing*. Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[14] X.R. Li and V.P. Jilkov. Survey of maneuvering target tracking part I: Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1333–1364, 2003.

[15] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3), March 2004.

[16] R.M. Chann. Recursive estimation of 3-D motion and structure in image sequences based on measurement transformations. *Master's Thesis, Queen's University*, 1994.

[17] X.R. Li and V.P. Jilkov. Survey of maneuvering target tracking part III: Measurement models. In *SPIE Conference on Signal and Data Processing of Small Targets*, San Diego, USA, July-August 2001.

[18] D.Lerro and V.P. Jilkov. Tracking with debiased consistent converted measurements versus ekf. *EEE Transactions on Aerospace and Electronic Systems*, 29(3):1015–1022, July 1993.

[19] Z. Zhao, X.R. Li, and V.P. Jilkov. Best linear unbiased filtering with nonlinear measurements for target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 40(4):1324–1336, 2004.

[20] A.V. Savkin and I.R. Petersen. Recursive state estimation for uncertain systems with an integral quadratic constraint. *IEEE Transactions on Automatic Control*, 40(6):1080, 1995.

[21] 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.

[22] I.R. Petersen and A.V. Savkin. *Robust Kalman Filtering for Signals and Systems with Large Uncertainties*. Birkhauser, Boston, 1999.

[23] I.R. Petersen, V.A. Ugrinovskii, and A.V. Savkin. *Robust Control Design Using $H^\infty$ Methods*. Springer-Verlag, London, 2000.

[24] A.V. Savkin and R.J. Evans. *Hybrid Dynamical Systems. Controller and Sensor Switching Problems*. Birkhauser, Boston, 2002.

[25] D. Lerro and Y. Bar-Shalom. Tracking with debiased consistent converted measurements versus EKF. *IEEE Transactions on Aerospace and Electronic Systems*, 29(3):1015–1022, July 1993.

[26] A.V. Savkin, P.N. Pathirana, and F.A.Faruqi. The problem of precision missile guidance : LQR and $H^\infty$ control framework. *IEEE Transactions on Aerospace and Electronic Systems*, 39(3):901 – 910, 2003.

[27] Y. Bar-Shalom. *Tracking and Data Association*. Academic Press Professional, Inc, San Diego, CA, 1987.

[28] Y. Bar-Shalom and X.R. Li. *Estimation and Tracking Principles, Techniques and Software*. Artech, Norwood, MA, 1993.

[29] B.D.O Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, N.J., 1979.

[30] 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.