GEOMETRIC AND STOCHASTIC ERROR MINIMISATION IN MOTION TRACKING

Karteek Alahari, Sujit Kuthirummal, C. V. Jawahar, P. J. Narayanan

Centre for Visual Information Technology
International Institute of Information Technology
Gachibowli, Hyderabad 500019. INDIA.
jawahar@iiit.net

ABSTRACT

Tracking has been considered to be a single view problem conventionally, where one is interested in the projections of a particular object in a view over time. Even if many views of the same event are available, tracking often proceeds independently in each view. The geometric information due to the projection of the same object onto multiple image planes is not utilized. In this paper, we couple the stochastic error used by the Kalman filter with a geometric error term derived from multiview geometric constraints to achieve improved tracking in individual views. We present experimental results to evaluate the performance of the algorithm.

1. INTRODUCTION

Visual tracking has been an active area of research in the Computer Vision community over the past decades. Due to its practical applications, tracking has gained importance in both Computer Vision and Computer Graphics [1]. Popular methods of tracking include the Kalman filter [2], the Condensation algorithm [3], etc. The Condensation algorithm uses a factored sampling approach to propagate the probability distribution of the parameters that are to be tracked over time. Kalman Filter has been used not only for tracking, but also for structure from motion computations[4] due to its convenient form for on-line real time processing. It is a set of mathematical equations that implement a predictor-corrector type estimator, and minimizes the error covariance under certain assumptions. Kalman Filter is shown to be optimal in the linear Gaussian environment. It is popular for tracking robots [5], radars [1], and for many other practical applications. The condensation algorithm, though applicable to non-linear densities, requires the prior density to be predetermined.

The problem of tracking has inherent challenges associated with it, most important one being occlusion of the feature point. This has been taken care of in various ways using the Kalman filter [6] in single views. Due to various reasons errors get introduced into the process of measuring the observations. Hence, it is desirable for the tracker to be robust enough to handle this noise. Kalman filter assumes the noise in the measurement model to be Gaussian.

In the presence of occlusion or noise, additional information from other viewing angles will be of immense help. Tracking, which has been traditionally assumed to be a single view problem, has not taken advantage of the recent developments in multiview analysis. The algebraic relations among the projections of a point onto multiple cameras have been studied extensively in projective, affine and Euclidean frameworks [7]. It has been recognised that additional views can provide information which helps to solve the problem of extracting meaning and structure from images of the scene. Using a plurality of projections, the lost information about the third dimension can be recovered. Applications of the multiview constraints is not limited to the reconstruction of the third dimension from a set of projections. They have been applied to view generation, object recognition, video stabilization, etc. [7].

In the past few years, multiple view tracking has gained considerable importance [8, 9, 10]. In spite of its importance, the geometric information available in multiple views has seldom been used [11]. Tracking has been performed on different views independently without utilizing their inter-relationships. We present a solution to the tracking problem by modifying the standard Kalman filter formulation in this paper. Based on the multiview relations among different views of the same scene, a new error term – called the *geometric error* – is used in the tracking process. The stochastic error minimised in the conventional filter has been coupled with this geometric error to get better tracking. Tracking proceeds by estimating the state (position) of the object at each time instant. The estimate obtained at each instant is corrected according to the error at that instant. This dynamic correction of the error makes tracking real-time and helps achieving faster results. Due to the incorporation of the multiview information at each stage, there is more than one measurement with which we correct our estimate. We present the formulation for two views in this paper; it can be extended to any number of views easily.

In section 2 we present the problem formulation for two-view tracking. The standard Kalman filter is briefly introduced along with the notations used, motion models, etc. The underlying assumptions are also discussed. Section 3 deals with Kalman Filtering in two views. Tracking in image space is discussed in Section 4 along with a discussion on occlusion and generalizations. Numerical results to validate the claims are presented in Section 5. Section 6 presents the conclusions.

2. TRACKING WITH GEOMETRIC CONSTRAINTS

The Kalman filter estimates the state in a two-stage manner by using feedback control. Based on the feedback obtained through the noisy measurements it updates the estimates of the state. It is a predictor-corrector algorithm where the prediction is done by the time update equations and the correction is done by the measurement update equations [12]. The time update equations are responsible for obtaining the state and error covariance for the next time step. The measurement update equations are responsible for updating the estimates using the noisy measurements. The Kalman filter algorithm is summarized below.

1. Initialize estimates of state and error covariance.
2. Predict the next state and the error covariance based on the best estimate till the present time instant.
3. Take the measurement at that time instant.
4. Correct the state and error covariance estimates based on the measurement made.
5. Repeat steps 2 - 4 for the next time instant.

Now we discuss the multiview tracking problem using the Kalman filter based approach. Let $\mathbf{x}_k = [X_k\ Y_k\ Z_k\ 1]^T$ be the 3D position of a point at time instant k. Let the point move as per a linear motion model as

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w_k} \tag{1}$$

$\mathbf{w_k}$ represents the process noise that is assumed to be independent, with normal probability distribution $p(w) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$. $\mathbf{A}$ is a 4×4 matrix which defines the motion of the point.

Mathematically, any camera can be represented as a 3×4 matrix, known as the camera projection matrix [7]. The relation between a world point $\mathbf{x}_k$ being imaged and the camera matrix is given by

$$\mathbf{I}_{l,k} = \mathbf{M}^l \mathbf{x}_k \tag{2}$$

where $\mathbf{I}_{l,k} = [u_k^l\ v_k^l\ 1]^T$ denotes the image point in view l at time k represented in the homogeneous notation [7] and $\mathbf{M}^l$ denotes the 3×4 camera matrix for view l. We assume the projection model to be affine [7] wherein the Gaussian noise in the world remains Gaussian in the image space. $\tilde{\mathbf{I}}_{l,k}$ represents the noisy measurement of the projection at time k, $\hat{\mathbf{I}}_{l,k}$ is the corrected estimate of the projection at time k, while $\hat{\mathbf{I}}_{l,k}^-$ is the predicted estimate of the projection at time k. We construct a measurement vector by concatenating the measurements in each view, $\mathbf{z}_k = [u_k^1\ v_k^1\ 1\ u_k^2\ v_k^2\ 1]^T$. $\tilde{\mathbf{z}}_k$ is the noisy measurement, the best estimate is denoted by $\hat{\mathbf{z}}_k$, while the predicted estimate is denoted by $\hat{\mathbf{z}}_k^-$. $\hat{\mathbf{x}}_k^-$ is the *a priori* estimate of the location of the world point at time k given knowledge of the process till time $k-1$, and $\hat{\mathbf{x}}_k$ is the *a posteriori* estimate of the location of the world point at time k given measurement $\tilde{\mathbf{z}}_k$. The measurement model is given by

$$\tilde{\mathbf{z}}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v_k} \tag{3}$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{M}^1 \\ \mathbf{M}^2 \end{bmatrix}$ is a 6×4 matrix. The measurement noise $\mathbf{v_k}$ is assumed to be independent, with normal probability distribution $p(v) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$.

In the conventional Kalman filter, the *a priori* estimate error $\mathbf{e_k^-}$ is defined as the difference between the *a priori* estimate of the world location, $\hat{\mathbf{x}}_k^-$ and the actual world location $\mathbf{x}_k$. Similarly, the *a posteriori* estimate error $\mathbf{e_k}$ is defined as difference between the *a posteriori* estimate of the world location, $\hat{\mathbf{x}}_k$ and the actual world location. The stochastic error in measurements is characterized by these errors.

The *a priori* estimate error covariance is the expectation of the outer product of the *a priori* estimate error $\mathbf{e}_k^-$ with itself given by

$$\mathbf{P}_k^- = \mathcal{E}[\mathbf{e}_k^- \mathbf{e}_k^{-T}], \tag{4}$$

and the *a posteriori* estimate error covariance is the expectation of the outer product of the *a posteriori* estimate error $\mathbf{e}_k$ with itself given by

$$\mathbf{P}_k = \mathcal{E}[\mathbf{e}_k \mathbf{e}_k^T]. \tag{5}$$

Kalman filter minimizes the *a posteriori* estimate error covariance given by equation 5.

2.1. Geometric Error

To include geometric information into the tracker, we define a *geometric error* in addition to the stochastic error employed in the conventional Kalman filter. Let $\mathbf{F}^{ij}$; $i, j = 1, 2$ and $i \neq j$ be the 3×3 fundamental matrix [7] between view i and view j. A row of zeros is added to the fundamental matrices to get 4×3 matrices $\mathbf{f}^{ij} = \begin{bmatrix} \mathbf{F}^{ij} \\ \mathbf{0}^T \end{bmatrix}$. The *a priori* geometric error $\mathbf{g}_k^-$ is defined as

$$\mathbf{g}_k^- = (\mathbf{f}^{ij}\hat{\mathbf{I}}_{j,k}^- - \mathbf{f}^{ij}\mathbf{I}_{j,k}) + (\mathbf{f}^{ji}\hat{\mathbf{I}}_{i,k}^- - \mathbf{f}^{ji}\mathbf{I}_{i,k}). \tag{6}$$

The term $(\mathbf{f}^{ij}\hat{\mathbf{I}}^-_{j,k} - \mathbf{f}^{ij}\mathbf{I}_{j,k})$ represents the geometric error in view j and the term $(\mathbf{f}^{ji}\hat{\mathbf{I}}^-_{i,k} - \mathbf{f}^{ji}\mathbf{I}_{i,k})$ represents the geometric error in view i. This error in each view can be visualized as the vector difference between the actual and the predicted epipolar lines [7]. It is shown pictorially in Figure 1. The *a posteriori* geometric error $\mathbf{g}_k$ is also defined on similar lines. For simplicity, we used two views of a scene in the formulation. It can be extended to more views by considering geometric errors corresponding to all the views.

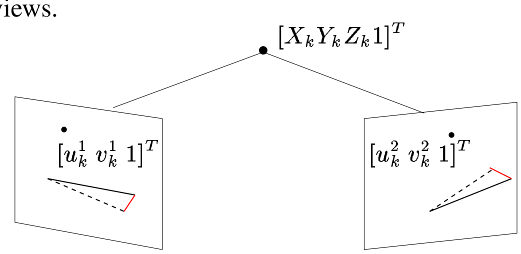

Fig. 1. The dark lines in black indicate the *actual* epipolar line , the dashed lines indicate the predicted epipolar line. The difference between these lines is denoted as the geometric error for that view.

In the dual space, due to the duality of lines and points, the geometric error can be interpreted as the distance between two points, one of which indicates the actual value and the other indicates the predicted value of the epipolar line.

3. KALMAN FILTERING IN TWO VIEWS

The error in the measurements has two components: stochastic and geometric. We can combine these errors to define a new formulation for the Kalman filter. We first derive a set of predictor-corrector equations aimed at estimating the world location of a 3D point from the measurements in two views. Then, we show that this can be easily converted to a form that results in a simple set of predictor-corrector equations for the location of the projections of the world features in the two views without computing the world locations.

The *a priori* geometric error $\mathbf{g}^-_k$ and the stochastic error $[\hat{\mathbf{x}}^-_k - \mathbf{x}_k]$ are combined. With equations 2 and 6, we simplify the new *a priori* estimate error as

$$\begin{aligned} \mathbf{e}^-_k &= [\hat{\mathbf{x}}^-_k - \mathbf{x}_k] + \mathbf{g}^-_k \\ &= (\mathbf{I} + \mathbf{f}^{12}\mathbf{M}^2 + \mathbf{f}^{21}\mathbf{M}^1)(\hat{\mathbf{x}}^-_k - \mathbf{x}_k) \\ &= \mathbf{D}(\hat{\mathbf{x}}^-_k - \mathbf{x}_k) \end{aligned} \quad (7)$$

where, $\mathbf{I}$ is a 4×4 identity matrix and $\mathbf{D} = (\mathbf{I} + \mathbf{f}^{12}\mathbf{M}^2 + \mathbf{f}^{21}\mathbf{M}^1)$. This matrix $\mathbf{D}$ embeds the geometric information obtained from the two views. Similarly, the *a posteriori* error is given by

$$\mathbf{e}_k = \mathbf{D}(\hat{\mathbf{x}}_k - \mathbf{x}_k) \quad (8)$$

Given the *a priori* $\hat{\mathbf{x}}^-_k$ estimate and the measurement $\tilde{\mathbf{z}}_k$ we would like to correct our prediction to get an *a posteriori* estimate $\hat{\mathbf{x}}_k$. This correction is done by the linear combination of the *a priori* estimate $\hat{\mathbf{x}}^-_k$ and a weighted difference between the measurement $\tilde{\mathbf{z}}_k$ and the predicted measurement $\mathbf{H}\hat{\mathbf{x}}^-_k$ as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}^-_k + \mathbf{G}_k(\tilde{\mathbf{z}}_k - \mathbf{H}\hat{\mathbf{x}}^-_k). \quad (9)$$

The 4×6 matrix $\mathbf{G}_k$, the gain or blending factor minimizes the *a posteriori* error covariance $\mathbf{P}_k$, i.e. we choose a gain factor $\mathbf{G}_k$ such that the *a posteriori* error covariance is minimized, as a consequence of which our estimate would coincide with the actual value.

Using equation 7 along with the *a priori* error covariance $\mathbf{P}^-_k$ in equation 4, we get

$$\begin{aligned} \mathbf{P}^-_k &= \mathcal{E}[\mathbf{D}(\hat{\mathbf{x}}^-_k - \mathbf{x}_k)(\hat{\mathbf{x}}^-_k - \mathbf{x}_k)^T\mathbf{D}^T] \\ &= \mathbf{D}\mathbf{B}_k\mathbf{D}^T \end{aligned}$$

where $\mathbf{B}_k = \mathcal{E}[(\hat{\mathbf{x}}^-_k - \mathbf{x}_k)(\hat{\mathbf{x}}^-_k - \mathbf{x}_k)^T]$, implying

$$\mathbf{B}_k = \mathbf{D}^{-1}\mathbf{P}^-_k\mathbf{D}^{-T} \quad (10)$$

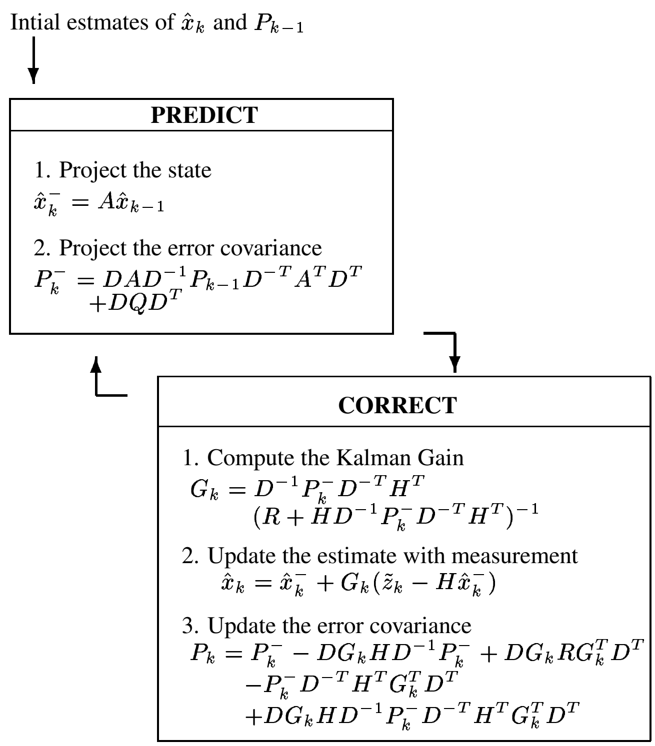

Fig. 2. Time update and measurement update equations for Two View Kalman Filter.

The location of the world point is predicted by making use of the knowledge of the motion model (Equation 1) as

$$\hat{\mathbf{x}}^-_k = \mathbf{A}\hat{\mathbf{x}}_{k-1}. \quad (11)$$

When this prediction equation (Equation 11) and the motion model equation (Equation 1) are used, the *a priori* error covariance $\mathbf{P}_k^-$ (Equation 4) can be expressed as

$$\begin{aligned} \mathbf{P}_k^- &= \mathcal{E}[\mathbf{D}(\hat{\mathbf{x}}_k^- - \mathbf{x}_k)(\hat{\mathbf{x}}_k^- - \mathbf{x}_k)^T\mathbf{D}^T] \\ &= \mathbf{D}\mathcal{E}[(\mathbf{A}\hat{\mathbf{x}}_{k-1} - \mathbf{A}\mathbf{x}_{k-1} - \mathbf{w}_k) \\ &\quad (\mathbf{A}\hat{\mathbf{x}}_{k-1} - \mathbf{A}\mathbf{x}_{k-1} - \mathbf{w}_k)^T]\mathbf{D}^T \end{aligned}$$

By grouping the terms appropriately we get

$$\begin{aligned} \mathbf{P}_k^- &= \mathbf{D}\mathbf{A}\mathcal{E}[(\hat{\mathbf{x}}_{k-1} - \mathbf{x}_{k-1})(\hat{\mathbf{x}}_{k-1} - \mathbf{x}_{k-1})^T]\mathbf{A}^T\mathbf{D}^T \\ &\quad +\mathbf{D}\mathbf{Q}\mathbf{D}^T \\ &= \mathbf{D}\mathbf{A}\mathbf{D}^{-1}\mathbf{P}_{k-1}\mathbf{D}^{-T}\mathbf{A}^T\mathbf{D}^T + \mathbf{D}\mathbf{Q}\mathbf{D}^T \end{aligned} \tag{12}$$

This simplification has been done to eliminate the parameters related to the world space. It provides relation between the *a priori* error covariance at time instant k and the *a posteriori* error covariance at time instant $k-1$.

A similar simplification for the *a posteriori* error covariance can be performed by manipulating equations 5, 8 and 9 as

$$\begin{aligned} \mathbf{P}_k &= \mathcal{E}[\mathbf{e}_k\mathbf{e}_k^T] \\ &= \mathbf{D}(\mathbf{B}_k - \mathbf{G}_k\mathbf{H}\mathbf{B}_k + \mathbf{G}_k\mathbf{R}\mathbf{G}_k^T - \mathbf{B}_k\mathbf{H}^T\mathbf{G}_k^T \\ &\quad +\mathbf{G}_k\mathbf{H}\mathbf{B}_k\mathbf{H}^T\mathbf{G}_k^T)\mathbf{D}^T \\ &= \mathbf{P}_k^- - \mathbf{D}\mathbf{G}_k\mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^- + \mathbf{D}\mathbf{G}_k\mathbf{R}\mathbf{G}_k^T\mathbf{D}^T \\ &\quad -\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T\mathbf{G}_k^T\mathbf{D}^T + \\ &\quad \mathbf{D}\mathbf{G}_k\mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T\mathbf{G}_k^T\mathbf{D}^T \end{aligned} \tag{13}$$

We want to choose the *gain* factor, $\mathbf{G}_k$, that minimizes the trace of $\mathbf{P}_k$ (Equation 13), which is equivalent to minimizing sum of squared errors. Thus, the stochastic and geometric errors are minimized together. For this we differentiate the trace with respect to $\mathbf{G}_k$ and set it to zero, which yields the required $\mathbf{G}_k$ as

$$\mathbf{G}_k = \mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T)^{-1} \tag{14}$$

The process of tracking a world point using the above formulation is stated here. We begin by assigning values to the initial estimates. Then, the next state and the error covariance are predicted by using equations 11 and 12. We then correct our estimates using equations 9, 13 and 14. The various steps involved in using the two view Kalman tracker have been summarized in Figure 2.

4. TRACKING IN IMAGE SPACE

In the previous section, we outlined a mechanism for utilizing the Kalman filter for predicting the 3D location of a world point undergoing linear motion by modeling stochastic and geometric errors. Prediction can also be done in the image space without explicitly taking measurements in the world space. This type of independence between the world and image spaces gives an added advantage. It eliminates the difficult process of measuring 3D world locations [8]. We show how the time and measurement update equations for the world point can be transformed to arrive at time and measurement update equations for the projections of the world point. To achieve this we take projections of the *a priori* ($\mathbf{H}\hat{\mathbf{x}}_k^-$) and *a posteriori* ($\mathbf{H}\hat{\mathbf{x}}_k$) estimates of the location of the world point given by equations in Figure 2 using the projection equation 2. The error covariance and the *gain* factor equations remain the same. The equations for the image space are

Time Update Equations

$$\hat{\mathbf{z}}_k^- = \mathbf{H}\mathbf{A}\mathbf{H}^{-1}\hat{\mathbf{z}}_{k-1} \tag{15}$$
$$\mathbf{P}_k^- = \mathbf{D}\mathbf{A}\mathbf{D}^{-1}\mathbf{P}_{k-1}\mathbf{D}^{-T}\mathbf{A}^T\mathbf{D}^T + \mathbf{D}\mathbf{Q}\mathbf{D}^T$$

Measurement Update Equations

$$\begin{aligned} \hat{\mathbf{z}}_k &= \hat{\mathbf{z}}_k^- + \mathbf{H}\mathbf{G}_k(\tilde{\mathbf{z}}_k - \hat{\mathbf{z}}_k^-) \qquad (16) \\ \mathbf{G}_k &= \mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T)^{-1} \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{D}\mathbf{G}_k\mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^- + \mathbf{D}\mathbf{G}_k\mathbf{R}\mathbf{G}_k^T\mathbf{D}^T \\ &\quad -\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T\mathbf{G}_k^T\mathbf{D}^T + \\ &\quad \mathbf{D}\mathbf{G}_k\mathbf{H}\mathbf{D}^{-1}\mathbf{P}_k^-\mathbf{D}^{-T}\mathbf{H}^T\mathbf{G}_k^T\mathbf{D}^T \end{aligned}$$

Discussion: We formulated the tracking problem for an affine model, which is linear in nature. Extensions for the projective model can be worked out using a non-linear filtering mechanism. Tracking of multiple objects is another possible extension. These are out of scope of this paper.

Occlusion can be handled as follows. Consider a situation where we have three views $1, 2$ and 3 of the object in known cameras. The pair-wise fundamental matrices $\mathbf{f}^{12}$, $\mathbf{f}^{23}$ and $\mathbf{f}^{31}$ (refer section 2) between these views is also known. The object is tracked in all the three views simultaneously. When it is occluded in only one of the views, say view 2, for some duration, measurements in that view cannot be obtained. Since the object is seen in the other views, we can use the corresponding trackers to predict the location of the object in them. From the two predicted locations in those views ($\hat{\mathbf{I}}_{1,k}$ and $\hat{\mathbf{I}}_{3,k}$) we can find the epipolar lines corresponding to these points in the third view as $\mathbf{f}^{12}\hat{\mathbf{I}}_{1,k}$ and $\mathbf{f}^{32}\hat{\mathbf{I}}_{3,k}$. The point of intersection of these epipolar lines gives the predicted location of the object in the third view. This can be used as the measurements when the point is occluded.

5. RESULTS

We tested the Kalman filter based tracker using geometric and stochastic errors by performing various experiments on synthetic as well as image data. The performance of the one-view and two-view trackers is compared in the rest of the section. The two-view tracker performs better in terms of convergence, accuracy, robustness to noise.

5.1. Faster convergence

Kalman filter works by minimizing the trace of the *a posteriori* error covariance $\mathbf{P}_k$. This gives a good measure of the error in prediction. Ideally, this error should approach zero in a few iterations. Figure 3 shows the value of the trace in one-view and two-view cases for the first 25 iterations of the algorithm. As can be seen from the graph, the error in two-view case approaches zero at a faster rate than it does in the one-view case. Real image data was used for this

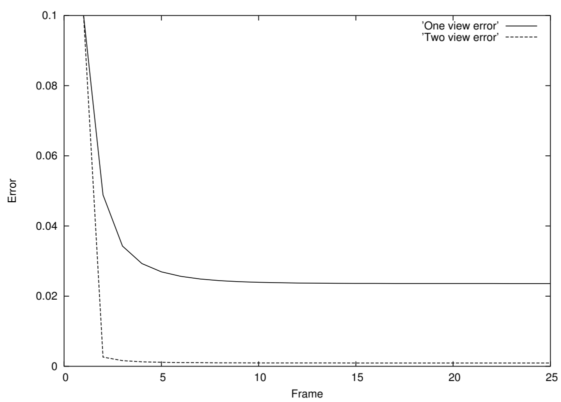

Fig. 3. Graph showing the decay of trace of *a posteriori* error covariance in the case of one-view and two-view Kalman filters.

experiment. It consisted of 25 frames each of two views of an event, shown in Figure 4. The left hand finger tip of the person shown was tracked using the one-view and two-view Kalman trackers (Sections 2 and 3). We calculate trace of the *a posteriori* error covariance matrices for both the trackers, normalize it by their corresponding initial trace values and plot it against the frame number. The error decreases at a faster rate in the two-view Kalman tracker. This is due to the fact that two measurements obtained from the two views are being used to update the *a priori* estimate that we have at each time instant. This in a way corrects the error at a faster rate compared to the standard formulation.

Similar results were observed for synthetic data. A 3D model of a car moving in a linear fashion was used. The motion of the car was captured using two affine cameras. Noisy measurements of the center of mass of the car were used to track it.

5.2. Robustness to noise

The additional information we have in terms of two views of the same scene can be used to make the Kalman filter more robust to noise in addition to making it converge faster. The robustness of the two-view Kalman filter has

(a) (b)

Fig. 4. Observing an event in two cameras (Courtesy - Keck Laboratory, University of Maryland, College Park)

been tested with different values of measurement and process noise. Synthetic data similar to that used in the above experiment was used so to provide flexibility in changing the noise values. We calculated the trace of the *a posteriori* error covariance matrix $\mathbf{P}_k$ for different pairs of measurement and process noise. The resulted are shown in Table 1. The noise values are indicated by the covariances of the error distribution function as $\mathbf{Q} = \sigma_1 \mathbf{I}$ and $\mathbf{R} = \sigma_2 \mathbf{I}$. The trackers exhibit similar behaviour even in the case when different covariance values are used for each dimension.

Measurement Noise σ_1	Process Noise σ_2	One-view Trace($\mathbf{P}_k$)	Two-view Trace($\mathbf{P}_k$)
0.15	0.05	0.283957	0.093196
0.15	0.15	0.523717	0.112909
0.50	0.35	0.845190	0.279587
0.70	1.00	1.123014	0.372782
3.00	2.00	2.318295	0.699715
5.50	3.50	4.647049	1.726031

Table 1. Comparison of the one-view and two-view error values for different measurement and process noise values after 25 iterations.

From Table 1 it can be observed that two-view Kalman filter produces lower error values compared to the one-view case, thereby confirming our claim that two-view tracking is more robust to noise.

Frame Number	One-view Distance	Two-view Distance
0 - 5	0.40311	0.05126
5 - 10	0.42757	0.05570
10 - 15	0.38262	0.04625
15 - 20	0.35815	0.01817
20 - 25	0.35986	0.00932
25 - 30	0.34290	0.00691

Table 2. RMS errors in 1-view and 2-view formulations

5.3. Better accuracy

It is important to know how accurately the object is being tracked. We quantified the accuracy of the tracker as the root mean square of the distances (RMS distance) of all the predicted points from their corresponding actual non-noisy measurements. Process and measurement noise are inherent in real data and hence cannot be eliminated by any simple process. Thus, to obtain the non-noisy measurements we generated the data synthetically conforming to the equations in the previous sections. (Sections 2, 3).

For this experiment, we performed tracking using both the one-view and two-view formulations and measured the corresponding accuracy values by finding the RMS distance. Ideally, this distance should be zero for a completely accurate tracker. The results are shown in Table 2.

5.4. Occlusion handling

To study the effect of occlusion, we considered three views of a scene along with their pair-wise fundamental matrices. After introducing occlusion in one of the views, no measurement is taken in that view. Performance of the tracking algorithm is given in the Table 3. In the one-view case, no correction term is applied during the occluded period. In the three-view case, the non-occluded views predict the location in this third view using epipolar lines. The distance between the prediction and the ground truth when the object is visible again is shown in columns 2 and 3 of the table for the one-view and three-view cases respectively. We can see that the two-view Kalman filter results are superior for handling occlusion.

6. CONCLUSIONS

We incorporated the additional information that is available from multiple views of a scene, in terms of geometric constraints on the error function, into the Kalman filter framework in this paper. The coupling of stochastic and geometric errors using two views led to better tracking. This formulation can be extended to multiple views by modifying the geometric error appropriately. The two-view based tracker defined in this paper can be used for tracking objects moving in a linear fashion. Further work is being done to extend the tracker for any general motion and any non-linear projection. The possibility of incorporating other multiview constraints such as multilinear tensors in place of fundamental matrix is also being explored.

Occlusion duration	One-view Error	Three-view Error
0 - 5	14.78	14.17
5 - 10	5.70	1.47
10 - 15	5.55	2.59
15 - 20	3.10	0.17
20 - 25	0.09	0.005

Table 3. Performance of the tracker under occlusions

7. REFERENCES

[1] K. V. Ramachandra, *Kalman Filtering Techniques for Radar Tracking*, SciTech Publishing Inc., USA, 2000.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of Basic Engineering*, vol. 82(Series D), pp. 35–45, 1960.

[3] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 28, no. 1, pp. 5–28, 1998.

[4] A. Azarbayejani and A. Pentland, "Recursive estimation of motion, structure, and focal length," *IEEE Trans. PAMI*, vol. 17, no. 6, pp. 562–575, June 1995.

[5] A. Fod, A. Howard, and M. J. Mataric, "Laser-based people tracking," *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3024–3029, 2002.

[6] L. Marcenaro, M. Ferrari, L. Marcbesotti, and C. S. Regazzoni, "Multiple object tracking under heavy occlusions by using kalman filters based on shape matching," *ICIP*, vol. 3, pp. 341–344, 2002.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, 2000.

[8] Y. Li, A. Hilton, and J. Illingworth, "Towards reliable real-time multiview tracking," *IEEE Workshop on Multi-Object Tracking (WOMOT'01)*, pp. 43–52, 2001.

[9] S. L. Dockstader and A. Murat Tekalp, "Multiple camera fusion for multi-object tracking," *IEEE Workshop on Multi-Object Tracking (WOMOT'01)*, pp. 95–102, 2001.

[10] Q. Delamarre and O. Faugeras, "3d articulated models and multiview tracking with physical forces," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 328–357, 2001.

[11] R. Vidal, Y. Ma, S. Hsu, and S. Sastry, "Optimal motion estimation from multiview normalized epipolar constraint," *ICCV*, vol. 1, pp. 34–41, 2001.

[12] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *SIGGRAPH - Course 8*, 2001.