Projective Kalman Filter: Multiocular Tracking of 3D Locations Towards Scene Understanding

C.Canton-Ferrer[1], J.R. Casas[1], M.Tekalp[2], and M.Pardàs[1] *

[1] Technical University of Catalonia, Barcelona, Spain,
{ccanton,josep,montse}@gps.tsc.upc.es,
[2] Koc University, Istanbul, Turkey,
mtekalp@ku.edu.tr

Abstract. This paper presents a novel approach to the problem of estimating and tracking 3D locations of multiple targets in a scene using measurements gathered from multiple calibrated cameras. Estimation and tracking is jointly achieved by a newly conceived computational process, the Projective Kalman filter (PKF), allowing the problem to be treated in a single, unified framework. The projective nature of observed data and information redundancy among views is exploited by PKF in order to overcome occlusions and spatial ambiguity. To demonstrate the effectiveness of the proposed algorithm, the authors present tracking results of people in a SmartRoom scenario and compare these results with existing methods as well.

1 Introduction

Estimating the 3D position and velocity of objects in a scene is of interest in a number of applications such as visual surveillance, human-computer interfaces, SmartRoom monitoring and scene understanding. Multiple view geometry has been addressed in [13] from a mathematical viewpoint, but there is still work to be done for the efficient fussion of redundant camera views and its combination with image analysis techniques for object detection and tracking. In this framework, the current paper proposes a novel technique to address the problem of tracking multiple 3D locations based on the data obtained from a set of calibrated cameras.

Many vision based tracking techniques have been developed to deal with sequences from a single perspective [7, 12] but considerably less work has been published on tracking of 3D locations with multiple cameras. One of the main problems within this topic is establishing correspondences among features from different perspectives presents a lot of difficulties for a tracking algorithm [4]. On the other hand, multiple viewpoints allow exploiting spatial redundancy and

* This material is based upon work partially supported by NoE FP6-507609 SIMILAR and IP IST-2004-506909 CHIL of the EU and by TEC2004-01914 project of the Spanish Government.

overcome ambiguities caused by occlusion or segmentation errors and provide 3D position information as well.

The common methodology to this problem in existing approaches is composed by two disjoint successive steps: estimation of the 3D location and Kalman tracking over this estimation. Bayesian networks [5,8], algebraic methods [10,18] or homographies [3] have been employed to establish correspondences among the projections of the 3D tracked points on all views and then perform a Kalman tracking directly on this estimated 3D location. The main drawbacks of these methods are sensitivity to occlusions and spatial ambiguity when resolving the multiple view correspondence problem [4].

In this paper, we present a novel technique that performs a joint estimation and tracking of multiple 3D locations allowing the problem to be posed in a single, unified framework. Projective geometry underlying the image formation process is exploited allowing the definition of our Projective Kalman Filter. Information redundancy among views is taken into account to define a data association process to deal with occlusions and keep a coherent track. The filter has found applicability in a SmartRoom scenario in the fields of body and gesture analysis (see Fig.1) or person tracking.

The outline of this work is as follows. Background topics on projective geometry and Kalman filtering required in forecoming sections are reviewed in Sec.2. Projective Kalman Filter theory is presented on Sec.3. Experimental results are presented in Sec.4. Finally, conclusion and further improvements are given in Sec.5.

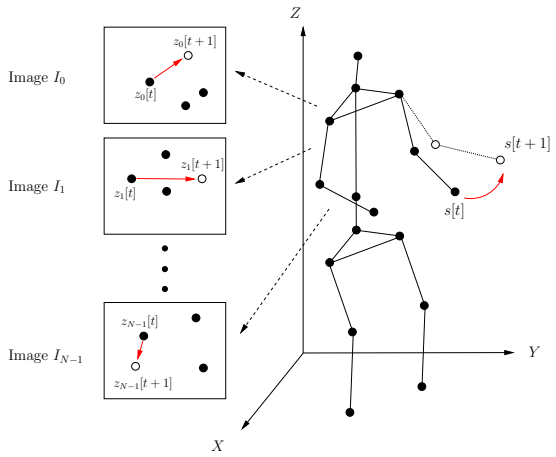

Fig. 1. Example of an application of tracking of 3D locations from its projections within the framework of body analysis (based on [9]). Tracking of the hidden state $s[t]$ among time from its projections $z_k[t]$, $0 \leq k < N$, would allow obtaining the position of body joints.

2 Projective geometry and Kalman Tracking Basics

In order to define a joint estimation-tracking scheme that exploits the underlying projective geometry of a multiple view scenario, some basic concepts are presented. Formation of images formulation and Kalman filtering theory are briefly reviewed but the reader is addressed to [13] and [16] for more references.

2.1 Multiple view systems and projective geometry

Obtaining two-dimensional coordinates (pixel positions) of an image from a three-dimensional magnitude (a 3D location) is a process where a dimension is lost. Formally, projection can be seen as a many-to-one morphism $\psi : \mathbb{R}^3 \rightarrow \mathbb{N}^2$ that transforms 3D Euclidean coordinates in the world reference frame into 2D coordinates in the camera reference frame. The usual mathematical way to model this process passes through projective geometry as an efficient description of the image formation process. Essentially, a camera is regarded as a projective device where an image is the result of the central projection of 3D world points onto the image plane. Specifically, the pinhole camera model is used in this paper. Projective effects due to vanishing points can be easily modeled and understood if we take into consideration projective coordinate systems. Many authors take advantage from projective geometry and homogeneous coordinates when addressing computer vision problems [13].

Projection operation can be fully described in homogeneous coordinates by the linear application $\mathbf{P} : \mathbb{P}^3 \rightarrow \mathbb{P}^2$ denoted as the *projection matrix*[3]. So,

$$\mathbf{x} = \mathbf{P}\mathbf{X}, \qquad \mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}], \qquad \mathbf{x} \in \mathbb{P}^2, \qquad \mathbf{X} \in \mathbb{P}^3, \tag{1}$$

where the *calibration matrix* $\mathbf{K}$ models the intrinsic parameters of the camera (focal length, scaling and projection center) and $\mathbf{R}$ and $\mathbf{t}$ its extrinsic parameters (rotation and translation of the camera).

It must be noted that projection is essentially a non-linear operation when defined by the application $\psi : \mathbb{R}^3 \rightarrow \mathbb{N}^2$. In fact, when adopting the pinhole camera model and the associated projective geometry model, the relation between the image coordinates $\tilde{\mathbf{x}} = [\tilde{x}\ \tilde{y}]^\top \in \mathbb{N}^2$ and the projected coordinates $\mathbf{x} = [x\ y\ z]^\top \in \mathbb{P}^2$ is stated as:

$$\tilde{x} = \left\lfloor \frac{x}{z} \right\rfloor, \qquad \tilde{y} = \left\lfloor \frac{y}{z} \right\rfloor. \tag{2}$$

For the sake of simplicity in the notation, let us re-define $\psi_{\mathbf{P}} : \mathbb{R}^3 \rightarrow \mathbb{N}^2$ as the projection operator from 3D coordinates to image coordinates embedding Eq.1 and Eq.2.

[3] The notation employed in this paper follows the one described by [11, 13].

2.2 Standard Kalman filter data model

The Kalman filter addresses the general problem of estimating the state $\mathbf{s} \in \mathbb{R}^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation:

$$\mathbf{s}[t+1] = \mathbf{F}\ \mathbf{s}[t] + \mathbf{w}[t], \tag{3}$$

with a measurement $\mathbf{z} \in \mathbb{R}^m$ that is

$$\mathbf{z}[t+1] = \mathbf{H}\ \mathbf{s}[t+1] + \mathbf{v}[t+1]. \tag{4}$$

The random variables $\mathbf{w}[t]$ and $\mathbf{v}[t]$ represent the state and measurement noise respectively. The matrix $\mathbf{F}$ in the difference Eq.3 relates the state at the future step $t+1$ to the state at the current step t and the matrix $\mathbf{H}$ in the measurement Eq.4 relates the state to the measurement $\mathbf{z}[t+1]$. Matrices $\mathbf{F}$ and $\mathbf{H}$ might change with each time step despite most of the approximations in Kalman filtering assume they are constant. In order to define a convergent Kalman filter, the random variables $\mathbf{w}[t]$ and $\mathbf{v}[t]$ are assumed to be independent of each other, white and with normal probability distributions

$$p(\mathbf{w}) \sim \mathcal{N}(0, \mathbf{Q}), \tag{5}$$

$$p(\mathbf{v}) \sim \mathcal{N}(0, \mathbf{R}). \tag{6}$$

2.3 Standard Kalman filter evolution

In summary, we have the following situation: starting from an initial estimate $\hat{\mathbf{s}}[0|-1]$, with an initial state covariance matrix denoted as $\mathbf{\Sigma}[-1|-1]$, for each observation $\mathbf{z}[t+1]$, the estimate of the state is updated using the following steps:

1. **State estimate extrapolation:**

$$\hat{\mathbf{s}}[t+1|t] = \mathbf{F}\hat{\mathbf{s}}[t|t] \tag{7}$$

2. **Error covariance extrapolation:**

$$\mathbf{\Sigma}[t+1|t] = \mathbf{F}\mathbf{\Sigma}[t|t]\mathbf{F}^\top + \mathbf{Q} \tag{8}$$

3. **Kalman gain:**

$$\mathbf{K}[t+1] = \mathbf{\Sigma}[t+1|t]\mathbf{H}^\top[t+1]\left(\mathbf{H}[t+1]\mathbf{\Sigma}[t+1|t]\mathbf{H}^\top[t+1] + \mathbf{R}\right)^{-1} \tag{9}$$

4. **State estimate update:**

$$\hat{\mathbf{s}}[t+1|t+1] = \hat{\mathbf{s}}[t+1|t] + \mathbf{K}[t+1]\left(\mathbf{z}[t+1] - \mathbf{H}[t+1]\hat{\mathbf{s}}[t+1|t]\right) \tag{10}$$

5. **Error covariance update:**

$$\mathbf{\Sigma}[t+1|t+1] = \left(\mathbf{I} - \mathbf{K}[t+1]\mathbf{H}[t+1]\right)\mathbf{\Sigma}[t+1|t] \tag{11}$$

3 Projective Kalman Filter (PFK)

Kalman filtering is the optimal strategy when dealing with estimation problems that involve linear relationships between the observed and real state variables and the distorting noise has a normal probability density. In the current analysis scenario, Kalman theory has still applicability and allows defining a joint estimation-tracking scheme exploiting the projective nature of the data gathered from the cameras.

3.1 Multi-camera 3D tracking scenario

Let us define $\tilde{\mathbf{X}}^i[t] = [\tilde{X}^i[t]\ \tilde{Y}^i[t]\ \tilde{Z}^i[t]]^\top$, $0 \leq i < M$, as the M 3D locations, targets, to be tracked along time. The available data of each of the N cameras is noted as $\tilde{\mathbf{x}}_k^i[t] = [\tilde{x}_k^i[t]\ \tilde{y}_k^i[t]]^\top$, $0 \leq i < M$, $0 \leq k < N$ and its formation process can be described as:

$$\tilde{\mathbf{x}}_k^i[t] = \psi_{\mathbf{P}_k}\left(\tilde{\mathbf{X}}^i[t]\right) + \boldsymbol{\xi}_k^i[t], \tag{12}$$

where $\boldsymbol{\xi}_k^i[t]$ is a noise factor present at time t in the projection of the i-th tracked object on the k-th camera and $\psi_{\mathbf{P}_k}$ is the projection operator associated to this camera. The noise factor $\boldsymbol{\xi}_k^i[t]$ is mainly formed by two contributions

$$\boldsymbol{\xi}_k^i[t] = \mathbf{g}_k^i[t] + \mathbf{d}_k^i[t], \tag{13}$$

where $\mathbf{g}_k^i[t]$ is the noise introduced by the inaccuracies of the calibration process, camera resolution, lens distortion,... considered to have a normal probability distribution in virtue of the Central Limit Theorem. On the other hand, $\mathbf{d}_k^i[t]$ is modeled as an impulsive noise result of a bad foreground region detection, occlusions or heavy lens distortion (borders of the image).

3.2 Kalman filtering on multiple projective planes

Defining a scheme embedding estimation and tracking based on a direct application of Kalman equations Eq.3 and Eq.4 is not straightforward. Let us define our state variable $\mathbf{s}[t]$ as the position and velocity that describe the dynamics of the tracked 3D location in homogeneous coordinates:

$$\mathbf{s}[t] = [\mathbf{X}^i[t]\ \ \dot{\mathbf{X}}^i[t]]^\top = [\tilde{X}^i[t]\ \ \tilde{Y}^i[t]\ \ \tilde{Z}^i[t]\ \ 1\ \ \dot{\tilde{X}}^i[t]\ \ \dot{\tilde{Y}}^i[t]\ \ \dot{\tilde{Z}}^i[t]\ \ 0]^\top. \tag{14}$$

The measure process described by Eq.4 must be modelled according to the projective nature of the observations. The data captured by the N cameras, that is the projections of the 3D tracked location given by Eq.12 (pixel positions), forms the observation vector $\mathbf{z}[t]$:

$$\begin{aligned} \mathbf{z}[t] &= [\mathbf{x}_0^i[t]\ \ \mathbf{x}_1^i[t]\ \ \cdots\ \ \mathbf{x}_{N-1}^i[t]]^\top \\ &= [\tilde{x}_0^i[t]\ \ \tilde{y}_0^i[t]\ \ 1\ \ \tilde{x}_1^i[t]\ \ \tilde{y}_1^i[t]\ \ 1\ \ \cdots\ \ \tilde{x}_{N-1}^i[t]\ \ \tilde{y}_{N-1}^i[t]\ \ 1]^\top, \end{aligned} \tag{15}$$

that is the detected projections of $\tilde{\mathbf{X}}^i[t]$ on every view.

It can be seen that the problem of tracking a 3D location (hidden state) from its projections on calibrated cameras (observation) does not fit with the standard Kalman filter formulation. Relations between the real state, $\tilde{\mathbf{X}}^i[t]$, and the observations, $\tilde{\mathbf{x}}_k^i[t]$, are non-linear. Thus, statistical distributions when processed by a projective device, $\psi_{\mathbf{P}_k}$, do not usually keep the same statistical properties. Hence Kalman filtering theory can not be applied directly. Solutions to this problem have arisen as the Extended Kalman Filter (EKF) [17], the Unscented Kalman Filter (UKF) [14] or Particle Filtering [1]. Moreover, normal distribution of the involved random variables is not fulfilled. The random variables modelling the movement of the 3D location to be tracked (position and velocity) are modelled as a normal distribution but the observed variables, that are affected by the noise factor $\boldsymbol{\xi}_k^i[t]$ described by Eq.13, are not. This problem can be coarsely solved by approximating $\boldsymbol{\xi}_k^i[t]$ by a normal distribution however, this solution leads to poor results in presence of occlusions (large values of $\boldsymbol{\xi}_k^i[t]$).

Projective Kalman filter is able to perform a joint estimation and tracking by adding some modifications on the parameters introduced by Eq.3 and Eq.4 in order to deal with the data model defined by Eq.14 and Eq.15. Filter evolution follow the standard Kalman equations defined in Sec.2.3. Regarding the state equation Eq.3:

- **State Transition Matrix:** Matrix $\mathbf{F}$ is set to be constant over time and defined as:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{1}{T} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{T} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{T} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} . \tag{16}$$

- **Process noise:** The statistics of process noise $\mathbf{w}[t]$ are set to be normal. The covariance matrix $\mathbf{Q}$ defining this random variable is learnt from groundtruth data and set invariant through time.

In order to define a Kalman scheme to track 3D positions from multiple camera data, the measure process described by Eq.4 must be modelled accordingly to the projective nature of the observations.

- **Observation Matrix:** The key point of our Kalman filter scheme relies in the definition of the observed data. A first proposal for this matrix would be:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0}_{3\times 4} \\ \vdots & \vdots \\ \mathbf{P}_{N-1} & \mathbf{0}_{3\times 4} \end{bmatrix} . \tag{17}$$

However, this matrix, when applied to the state vector $\mathbf{s}[t]$ would generate coordinates that might not be on the image plane ($z \neq 1$). Hence, the projection non-linearity must be compensated to obtain coordinates fulfilling $z = 1$ in order to have a coherent data model. Our proposal for the adaptive design of the matrix $\mathbf{H}[t+1]$ is as follows:

$$\mathbf{H}[t+1] = \begin{bmatrix} \boldsymbol{\alpha}_0 \cdots & \mathbf{0} \\ \vdots \ \ddots & \vdots \\ \mathbf{0} \ \cdots & \boldsymbol{\alpha}_{N-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{P}_0 & \mathbf{0}_{3\times 4} \\ \vdots & \vdots \\ \mathbf{P}_{N-1} & \mathbf{0}_{3\times 4} \end{bmatrix},$$
$$\boldsymbol{\alpha}_k = \frac{1}{\mathbf{P}_k^3 \cdot \hat{\mathbf{s}}[t+1|t]} \mathbf{I}_{4\times 4}, \tag{18}$$

where $\mathbf{P}_k^3$ is the 3th row of $\mathbf{P}_k$ and $\hat{\mathbf{s}}[t+1|t]$ is the predicted state given by Eq.7. In this way, when computing Eq.10 the observed, $\mathbf{z}[t+1]$, and predicted term, $\mathbf{H}[t+1]\hat{\mathbf{s}}[t+1|t]$, can be compared (both have $z = 1$) leading to a meaningful result. The non-linearity introduced by the projection operator, $\psi_{\mathbf{P}_k}$, is therefore overcome and successfully modelled.

- **Observation noise:** The statistics of the observation noise $\boldsymbol{\xi}_k^i[t]$ can not be modelled as a random variable with normal distribution. Nevertheless, despite Kalman theory would seem not to be applicable, we propose an scheme to design an adaptive covariance matrix $\mathbf{R}[t]$ that will be able to handle occlusions and make Kalman theory fit in our scheme. Covariance matrix $\mathbf{R}[t]$ can be seen as a matrix that controls how reliable is the observed data in order to use it for the estimation of the hidden state $\hat{\mathbf{s}}[t+1|t+1]$. In the observation process, there could be two situations: if there is no occlusion in the projection of $\mathbf{X}^i[t]$ onto the k-th view, then the distorting noise $\boldsymbol{\xi}_k^i[t]$ reduces to be the AWGN $\mathbf{g}_k^i[t]$ part or if there is occlusion and the predominant noise term turns out to be the impulsive $\mathbf{d}_k^i[t]$ factor. Under this model, $\mathbf{R}$ matrix can be defined for every time step as:

$$\mathbf{R}[t] = \begin{bmatrix} \boldsymbol{\beta}_0 \cdots & \mathbf{0} \\ \vdots \ \ddots & \vdots \\ \mathbf{0} \ \cdots & \boldsymbol{\beta}_{N-1} \end{bmatrix}, \tag{19}$$

where

$$\boldsymbol{\beta}_k = \begin{cases} \boldsymbol{\sigma}_k \text{ if there is no occlusion } (\boldsymbol{\xi}_k[t] \approx \mathbf{g}_k[t]) \\ \infty \text{ if there is occlusion } (\boldsymbol{\xi}_k[t] \approx \mathbf{d}_k[t]) \end{cases}. \tag{20}$$

With this scheme, non-informative data coming from occluded views is disregarded when computing the estimation of the hidden state and projections corrupted with AWGN are correctly handled. The algorithm to decide whether a view is occluded or not is described in Sec.3.3.

3.3 Data association problem

In presence of multiple objects, occlusion and noisy measurements, it is important to assign the correct measurement to each tracked object. This is called the

data association problem [2, 6]. The following algorithm describes how to associate data to every tracked object in the scene (inspirated by [18]) and decide whether an occlusion has occurred in some views.

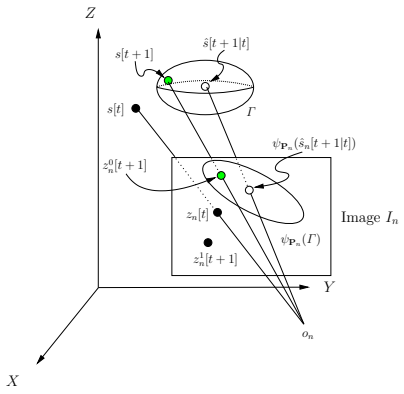

Fig. 2. Data association scenario. State estimation $\hat{\mathbf{s}}[t+1|t]$ and the uncertainty region defined by Γ when projected into image I_n allow associating the correct observation, $\mathbf{z}_n^0[t+1]$, to the interest track dismissing false detections, $\mathbf{z}_n^1[t+1]$.

Data association must determine the spatial correspondence of two projections generated by the same 3D feature at two consecutive time instants in the same image. In this way, when tracking multiple targets, the algorithm will be able to perform properly. Moreover, in the case when a correspondence can not be established probably due to an occlusion, the data association algorithm should modify the $\mathbf{R}[t+1]$ matrix accordingly. The proposed data association procedure is described by the following steps:

1. **State estimate extrapolation:** In order to perform a search for the most likely correspondence on time $t+1$, the algorithm estimates the state at this time through Eq.7 thus obtaining $\hat{\mathbf{s}}[t+1|t]$.
2. **Data bounding:** From the state evolution equation Eq.3, it can be assumed that the uncertainties of the 3D tracked location, the state, are modelled by the process noise described by the covariance matrix $\mathbf{Q}$. Assuming that this matrix has been correctly estimated, it can be inferred that the 3D position, $\mathbf{s}[t+1]$, fulfills the condition:

$$\mathbf{s}[t+1] \in \Gamma, \tag{21}$$

$$\Gamma \; : \; \left\{ \mathbf{X} / \left(\mathbf{X} - \hat{\mathbf{s}}[t+1|t]\right) \; \mathbf{W}^{-1} \; \left(\mathbf{X} - \hat{\mathbf{s}}[t+1|t]\right)^{\top} \leq 0 \right\}. \tag{22}$$

That is, $\mathbf{s}[t+1]$ is inside the ellipsoid Γ in homogeneous coordinates defining an uncertainty region proportional to the state noise covariance. The conic

matrix $\mathbf{W}$ [13] contents information about the topology of the ellipsoid and we define it from $\mathbf{Q}$ as:

$$\mathbf{W} = \begin{bmatrix} & & & 0 \\ & \gamma\mathbf{Q} & & 0 \\ & & & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \gamma\sigma_x & 0 & 0 & 0 \\ 0 & \gamma\sigma_y & 0 & 0 \\ 0 & 0 & \gamma\sigma_z & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}. \tag{23}$$

In our experiments, a value $\gamma = 6$ has provided effective results.

3. **Data Association:** The geometric property defined in Eq.21 and Eq.22 must be also fulfilled when dealing with a projection of this 3D scenario as depicted in Fig.2. A process to associate the most likely projection at time $t+1$ with respect to t can be defined straightforward. Since our input data are pixels detected on the projected images we could associate the pixel that minimizes a given criteria related to the projection of Γ, $\psi_{\mathbf{P}_k}(\Gamma)$, to the i-th track. Generally, the perspective projection of an ellipsoid is an ellipse defined by the matrix $\mathbf{V}$ fulfilling the following condition [13]:

$$\mathbf{V} \propto \left(\mathbf{P}_k \mathbf{W}^{-1} \mathbf{P}_k^\top\right)^{-1}. \tag{24}$$

Then, a proposal to establish the best association between the i-th track at the time $t+1$ with the input data $\mathbf{z}_n^l[t+1]$, $0 \leq l < L$ (there could be uncountable input data coming from the real tracks, false detections,...) can be done through the Mahalanobis distance:

$$\mathbf{z}_n^i[t+1] = \tag{25}$$
$$\min_{\mathbf{z}_n^l[t+1]} \sqrt{\left(\mathbf{z}_n^l[t+1] - \psi_{\mathbf{P}_k}\left(\hat{\mathbf{s}}[t+1|t]\right)\right) \mathbf{V} \left(\mathbf{z}_n^l[t+1] - \psi_{\mathbf{P}_k}\left(\hat{\mathbf{s}}[t+1|t]\right)\right)^\top}.$$

4. **Occlusion detection:** In the case when the condition related to the i-th track association

$$\sqrt{\left(\mathbf{z}_n^i[t+1] - \psi_{\mathbf{P}_k}\left(\hat{\mathbf{s}}[t+1|t]\right)\right) \mathbf{V} \left(\mathbf{z}_n^i[t+1] - \psi_{\mathbf{P}_k}\left(\hat{\mathbf{s}}[t+1|t]\right)\right)^\top} > \delta, \tag{26}$$

is fulfilled, being δ a threshold, we can say that there is an occlusion or the data is too corrupted to be taken into account in next steps of the Kalman filter. Hence, a criterium to set the parameter $\boldsymbol{\beta}_k$ from Eq.20 is defined. For our experiments, we took $\delta = 0.2$

4 Results

In order to evaluate the performance of the proposed tracking method, two experiments were carried out. We applied the described algorithm to both synthetic and real data to demonstrate the efficiency of our solution and compare it to the performance of the existing approaches to this problem within a SmartRoom framework [10, 18]. The scenario where this algorithm was applied (in both synthetic and real data) was the SmartRoom at UPC provided with 5 fully calibrated wide angle lense cameras with a resolution of 768x576 pixels at 25 fps.

Experiment 1: Synthetic data

A synthetic path was created simulating the movement of a single person walking inside a SmartRoom. For this scenario two possibilities of the noise factor $\boldsymbol{\xi}_k$ were studied: only Gaussian noise or Gaussian noise and occlusions added in the projected views. For the first case, different Gaussian noise levels were added in the projected views according to the measurement equation Eq.4. For the second case, occlusions were simulated by adding high amplitude bursts of a duration of 10 frames with $P_{\text{occlusion}} = 0.3$. For these input data, PKF and the standard KF [10,18] algorithms were applied to test and compare the performance of our joint estimation-tracking scheme. Fig.3(a) and 3(b) depict the error curves for different levels of noise in the two situations. Fig.3(c) shows the zenital view of the grountruth and PKF and KF estimated paths. Finally, Table 1 shows some quantitative results comparing PKF and KF performances.

Table 1. Mean and standard deviation of the error for tracks with different levels of Gaussian noise for PKF and KF. (Values in mm)

Gaussian Noise	PKF		KF	
σ^2	μ	σ	μ	σ
50	7.93	3.90	9.48	4.38
100	10.31	5.09	13.13	6.17
150	11.90	5.90	15.87	7.54
200	13.11	6.55	18.15	8.68
250	14.01	7.10	20.12	9.66
300	14.90	7.58	21.83	10.54

Experiment 2: Real data

In order to test our system, a sequence of 400 frames with two people spontaneously interacting with each other was recorded. Foreground regions were segmented and the top of each region in every view was taken as the input data in order to track the 3D head of each person. By applying both tracking filters, PKF and KF, we obtained the tracking results depicted in Fig.4. In the case were the foreground regions representing the two people merged in one view, the redundancy in the other views allowed keeping coherent tracks but accuracy of the position estimation decreased. Video results for this sequence can be get at `http://gps-tsc.upc.es/imatge/_Ccanton/pkf.zip`.

5 Conclusions and Future Work

A new approach towards tracking 3D locations from its projections on multiple calibrated cameras has been presented. The proposed scheme performs a joint estimation and tracking by taking advantage of the projective nature of the observations defining the Projective Kalman Filter. Results on synthetic and real

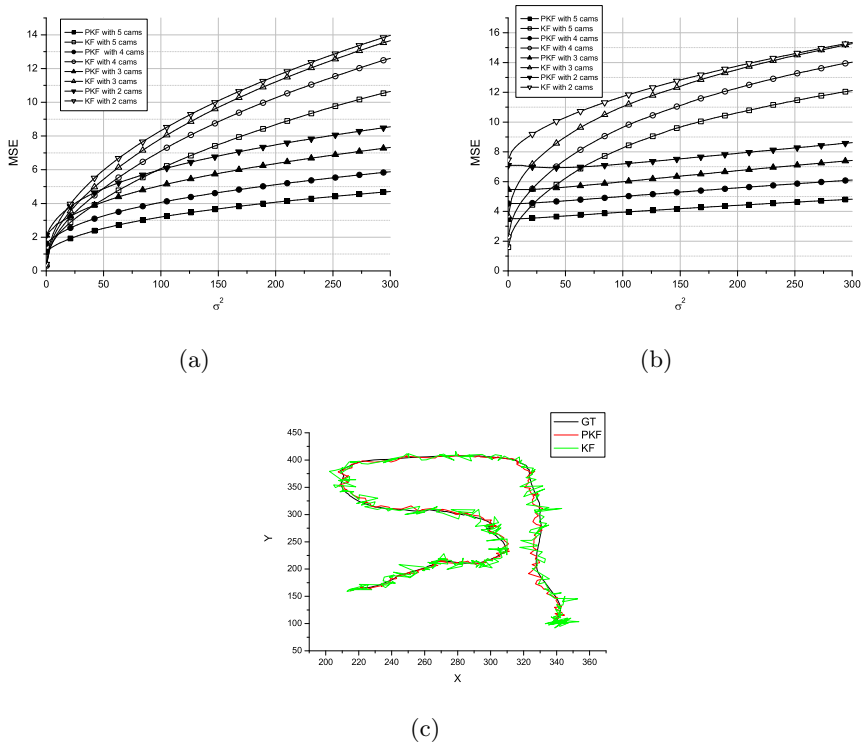

Fig. 3. Results on synthetic data. In (a), the error curves for the PKF and KF for diverse levels of Gaussian noise. In (b), the error curves for the PKF and KF operating in the same noise conditions with a $P_{\text{occlusion}} = 0.3$ and an occlusion length of 10 samples. In (c), the groundtruth trajectory of the location of interest and the results of PKF and KF (zenital view).

data proved this scheme to produce more reliable results in comparison with the standard Kalman approaches to this problem. The accuracy of PKF was good, even though the error in the experiments with real data were conditioned by calibration, foreground segmentation and camera position.

Future research perspectives involve the development of schemes more robust to occlusions, input data inconsistencies and position of the cameras. Applications of this technique to body analysis and joint tracking are under research as well.

References

1. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. on Signal Proc. **50-2** (2002), 174–188.
2. Bar-Shalom, J., Fortmann, T.E.: Tracking and Data Association. Academic Press. 1988.

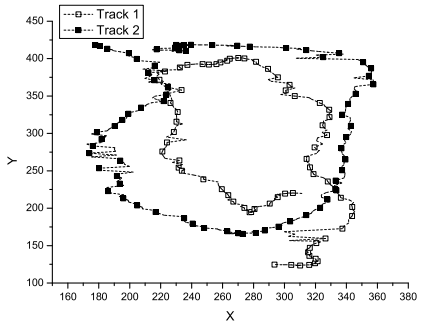

Fig. 4. Results on real data. Zenital plot showing simultaneous tracking of two people.

3. Black, J., Ellis, T.: Multi Camera Image Tracking. Proc. Work. on Motion and Video Computing (2001).
4. Canton-Ferrer, C., Casas, J.R., Pardàs, M.: Towards a Bayesian Approach to Robust Finding Correspondences in Multiple View Geometry Environments. LNCS, **3515:2** (2005), 281–289.
5. Chang, T.H., Gong, S.: Tracking Multiple People with a Multi-Camera System. Proc. IEEE Work. on Multi-Object Tracking (2001).
6. Cox, I.J.: A Review of Statistical Data Association Techniques for Motion Correspondence. Int. J. of Computer Vision, **10-1** (1993), 53–56.
7. Darrell, T., Gordon, G., Harville, M.: Integrated person tracking using stereo, color and pattern detection. Int. J. of Computer Vision, **37-2** (2000), 175–185.
8. Dockstader, S.L., Tekalp, A.M.: Multiple camera tracking of interacting and occluded human motion. Proc. of IEEE **89-10** (2001) 1441–1455.
9. Dockstader, S.L., Berg, M.J., Tekalp, A.M.: Stochastic Kinematic Modeling and Feature Extraction for Gait Analysis. IEEE Trans. on Imag. Proc. **12-8** (2003) 962–976.
10. Focken, D., Stiefelhagen, R.: Towards vision-based 3D people tracking in a smart room. Proc. IEEE Int. Conf. on Multimodal Interfaces (2002) 400–405.
11. Garcia, O.: Mapping 2D images and 3D world objects in a multicamera system. Ms. Thesis. Technical University of Catalonia. 2004.
12. Haritaoglu, I., Harwood, D., David, L.: W^4: Who?When?Where?What?A real time system for detecting and tracking people. LNCS, **1406** (1998), 877–892.
13. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. 2nd Edition. Cambridge University Press. 2004.
14. Julier, S.J., Uhlmann, J.K.: A New Extension of the Kalman Filter to Nonliner Systems. Proc. of AeroSense: The 11th Int.Symp. on Aerospace/Defence Sensing, Simulation and Controls (1997).
15. Jung, S.K., Wohn, K.Y.: 3D Tracking and Motion Estimation using Hierarchical Kalman Filter. Proc. IEE Visual Image Signal Process. **144-5** (1997)
16. Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. Trans. of the ASME–J. of Basic Engineering, **82-D** (1960) 35–45.
17. Lewis, F.L.: Optimal Estimation. John Wiley and Sons, New York. 1986.
18. Mikic, I., Santini, S., Jain, R.: Tracking Objects in 3D using Multiple Camera Views. Proc. Asian Conf. on Computer Vision (2000) 234–239.