Trajectory Association across Non-overlapping Moving Cameras in Planar Scenes

Yaser Sheikh
Robotics Institute
Carnegie Mellon University
yaser@cs.cmu.edu

Xin Li
Mathematics Department
University of Central Florida
xli@math.ucf.edu

Mubarak Shah
SEECS
University of Central Florida
shah@cs.ucf.edu

Abstract

The ability to associate objects across multiple views allows co-operative use of an ensemble cameras for scene understanding. In this paper, we present a principled solution to object association where both the scene and the object motion are modeled. By making the motion model of each object with respect to time explicit, we are able to solve the trajectory association problem in a unified framework for overlapping or non-overlapping cameras. We recover the assignment of associations while simultaneously computing the maximum likelihood estimates of the inter-camera homographies and the trajectory parameters using the Expectation Maximization algorithm. Quantitative results on simulations are reported along with several results on real data.

1. Introduction

To understand an expansive scene, rather than deciphering the observations of a single camera, it is useful to deploy an ensemble of distributed cameras. In addition to providing richer scene information and resistance to object occlusions, wider areas can be observed by multiple cameras. However, a key problem that needs to be addressed before the cameras can be used co-operatively is the association of objects *across* the cameras. Once association has been established for all objects, problems such as object counting, co-operative object tracking and occlusion resolution are simpler to tackle. A large and growing body of work addresses various problems in object association. Constraints on the motion of the objects across non-overlapping cameras were first proposed by Kettnaker and Zabih, [10], where positions, object velocities and transition times across cameras were used in a setup of known path topology and transition probabilities. In [2], Collins *et al*. used a system of calibrated cameras with an environment model to track objects across multiple views. The method proposed by Javed *et al*., in [7], did not assume a site model or explicit calibration of cameras, instead they learnt the inter-camera illumination and transition properties during a training phase, which were then used to track objects across the cameras. In [16], Stauffer and Tieu tracked across multiple cameras with both overlapping and non-overlapping fields of view, building a correspondence model for the entire set of cameras. They made an assumption of scene planarity and recovered the inter-camera homographies. Recently, Shan *et al*. in [14] proposed a collection of edge-based measure to solve a 'same-different' variation of the object association problem. Some work has been published for recovering the pose and/or tracks between cameras with non-overlapping fields of view. Fisher, in [4], showed that given a set of randomly placed cameras, recovering pose was tractable using distant moving features and nearby linearly moving features. In [11], Makris *et al*. also extracted the topology of a number of cameras based on the co-occurrence of entries and exits. Rahimi *et al.*, in [13], presented an approach that reconstructed the trajectory of a target and the external calibration parameters of the cameras, given the location and velocity of each object. Recently, Tieu *et al*. in [18] proposed an approach to recover the topology of an ensemble of cameras.

These approaches have assumed the camera remained stationary, with overlapping and non-overlapping FOVs. The collective field of view of the sensors can be further increased if motion is allowed in sensors. A limited type of camera motion has been examined in previous work: motion of the camera about the camera center, i.e. pan-tilt-zoom (PTZ) motion. One such work is [12], where Matsuyama and Ukita present a system based approach using active cameras, developing a fixed point PTZ-enabled camera for wide area imaging. In [9] Kang *et al.* proposed a method that involved multiple fixed and PTZ-enabled cameras. It was assumed that the scene was planar and that the homographies *between* cameras were known. Using these transformations, a common coordinate frame was established and objects were tracked across the cameras using

1-4244-1180-7/07/$25.00 ©2007 IEEE

color and motion characteristics.

In this paper we present a unified framework for the association of multiple objects across multiple cameras in planar scenes. This approach makes additional assumptions on the object kinematics but is able to recover object associations, inter-camera transformations and canonical trajectories across cameras irrespective of whether the cameras are stationary or moving, or whether the fields of view are overlapping or not as long as the kinematic model is valid. The intuition used to solve this problem is that association across cameras with spatiotemporally non-overlapping fields of view can be achieved by explicitly modeling the motion of objects, providing constraints for the estimation of inter-camera homographies. We use polynomial kinematic models for the motion of objects and under this model an Expectation Maximization algorithm is formulated to estimate the inter-camera homographies and motion parameters. Within the taxonomy of [15] we address the most general instance of the object association problem where no spatiotemporal overlap can be assumed.

Most existing approaches to estimating inter-camera homographies from curves, such as conics, perform the matching given the parameters of the curves. The general theory is covered in [8]. A separate portion of literature cover the problem of fitting curves to points - a survey for conics can be found in [5]. In this paper, we fuse the two problems, of estimating curve parameters and the homographies simultaneously. The benefit of this is two-fold. First, it is difficult to directly characterize an error model for curve coefficients, since they are not usually directly measurable. On the other hand, it is reasonable to assume an error model for point detection, and then develop statistically meaningful estimation algorithms for estimating homographies between views. Second, since only a portion of the curve is observed in each view, it is likely that the curve may be erroneously fit in each view. This is due to the fact that samples from the curve are localized in small intervals for each view (partial occlusion). By estimating curve parameters and homographies simultaneously, recovery is possible from local over-fitting (per camera).

There are two principal applications where the algorithms in this paper can be used. First, where multiple aerial cameras at high altitudes, observing objects such as vehicles or people move along the ground, and the problem is to recover the association of the objects across cameras and estimate the inter-camera transformations. Second, for a single camera in this setting if, due to the motion of the camera, an object exits and then re-enters the field of view of one camera, the problem of reassociation can also be solved in this context. We would like to make it clear at the outset that the single camera tracking problem is assumed to be solved (although this can also be simultaneously solved within the proposed framework as described in the future work section). We also assume that the data has been time-stamped or has been temporally aligned. Finally, we concentrate on best associating *trajectories* across cameras based on their motion characteristics. Appearance based matching constraints as used in [7] and [14] are a strong cue for matching and can also be used along with motion cues during matching. Unfortunately, it is still unclear how best to use appearance constraints, and that problem is beyond the scope of this paper.

The rest of the paper is organized as follows. In Section 2 we introduce the notation used in this paper and describe the scene and data model (including all approximations and assumptions) used in the rest of the paper. In Section 3 we describe the application of our scene model, and the maximum likelihood estimate of the motion parameters, inter-camera homographies and the object associations. Results of experimentation are presented in Section 4. Finally, conclusions and future directions are presented in Section 5.

2. Data Model

The scene is modeled as a plane in 3-space, $\mathbf{\Pi}$, with K moving objects. The k-th object[1], O_k, moves along a trajectory on $\mathbf{\Pi}$, represented by a time-ordered set of points, $\mathbf{x}_k(t) = \big(x_k(t), y_k(t)\big) \in \mathbb{R}^2$, where $x_k(t)$ and $y_k(t)$ evolve according to some spatial algebraic curve such as a line, a quadratic or a cubic. The finite temporal support is denoted by Δt. The scene is observed by N perspective cameras, each observing some subset of the entire scene motion, due to a spatially limited field of view and temporally limited window of observation (due to camera motion). The imaged trajectory observed by the n-th camera for O_k is $\mathbf{x}_k^n(t)$. We assume that within each sequence frame-to-frame motion within camera has been compensated so $\mathbf{x}_k^n(t)$ is in a single reference coordinate. The measured image positions of objects, $\bar{\mathbf{x}}_k^n$ are described in terms of the canonical image positions, $\mathbf{x}_k^n$, with independent normally distributed measurement noise, $\mu = \mathbf{0}$ and covariance matrix $\mathbf{R}_k^n$, that is

$$\bar{\mathbf{x}}_k^n(t) = \mathbf{x}_k^n(t) + \epsilon, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k^n). \tag{1}$$

The imaged trajectory is related to $\mathbf{x}_k(t)$ by a projective transformation denoted by an invertible 3×3 matrix, $\mathbf{H}^n$. The homogeneous representation of a point $\mathbf{x}_k^n(t)$ is $\mathcal{X}_k^n(t) = (\lambda x_k^n(t), \lambda y_k^n(t), \lambda) \in \mathcal{P}^2$. Thus, we have,

$$\mathcal{X}_k^n(t) = \mathbf{H}^n \mathcal{X}_k(t).$$

Finally, we introduce the association or correspondence variables $\mathbf{C} = \{c_k^n\}_K^N$, where $c_j^i = m$ represents the hypothesis that O_j^i is the image of O_m, where $p(c)$ is the

[1]The abstraction of each object is as a point, such as the centroid. It should be noted, however, that since the centroid is not preserved under general perspective transformations using the centroid will introduce bias.

probability of association c. Since the association of an imaged trajectory with different scene trajectories are mutually exclusive and exhaustive, $\sum_{l=1}^{K} p(c_k^n = l) = 1$. A term $p(c_k^n = 0)$ may be included to model the probability of spurious trajectories but we do not consider this in the remainder of this work (i.e., we assume $p(c_k^n = 0) = 0$).

2.1. Kinematic Polynomial Models

The position $\mathbf{x}_j(t)$ of an Object O_j is modeled as an $d-th$ order polynomial in time,

$$\mathbf{x}_j(t) = \sum_{i=0}^{d} \mathbf{p}_i t^i, \tag{2}$$

where $\mathbf{p}_i$ are the coefficients of the polynomial. In matrix form,

$$\mathbf{x}_j(t) = \mathbf{P}_j \mathbf{t}^{(d)} = \begin{bmatrix} p_{x,0} & p_{x,1} & \cdots & p_{x,d} \\ p_{y,0} & p_{y,1} & \cdots & p_{y,d} \end{bmatrix} \begin{bmatrix} 1 \\ t \\ \vdots \\ t^d \end{bmatrix}.$$

We omit the dependence of p on j for notational simplicity. Selecting the appropriate order of polynomials is an important consideration. If the order is too low, the polynomial may not correctly reflect the kinematics of the object. On the other hand, if the order is too high, some of the estimated coefficients may not be statistically significant, [3]. This problem is even more important in the situation under study since oftentimes only a segment of the polynomial is observed and over or under-fitting is likely. Thus, numerical considerations while estimating the coefficients of the curve are of paramount important, especially during the optimization routine. Readers are advised to refer to [6] for information on numerical conditioning during estimation. The monograph by Fitzgibbon and Fischer [5] on conic fitting is also informative.

For instance, the number of parameters that need to be estimated when a parametric cubic curve is to be fit to the trajectories is at most $8K + 9N$, since there are K curves which are described by 8 parameters each, with N homographies, each with 9 unknowns. At least four points per object must be observed and just one curve must be observed between a pair of views. The parametrization for a cubic curve is,

$$\mathbf{x}(t) = \mathbf{p}_3 t^3 + \mathbf{p}_2 t^2 + \mathbf{p}_1 t + \mathbf{p}_0. \tag{3}$$

In this case

$$\mathcal{P} = \begin{bmatrix} p_{x,0} & p_{x,1} & p_{x,2} & p_{x,3} \\ p_{y,0} & p_{y,1} & p_{y,2} & p_{y,3} \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Since the scene is modeled as a plane, a point on $\mathbf{\Pi}$ is related to its image in the n-th camera by $\mathbf{H}^n$. Thus a measured point $\mathcal{X}_j^i$ at time t associated with O_m (i.e., $c_j^i = m$) is,

$$\bar{\mathcal{X}}_j^i = \mathbf{H}^i \mathcal{P}_m \mathbf{t}^{(d)} + \tilde{\epsilon}. \tag{4}$$

3. Maximum Likelihood Estimation

The problem statement is as follows: Given the trajectory measurements for each camera $\{\bar{\mathbf{x}}_k^n\}_K^N$, find associations $\mathbf{C}$ of each object across cameras and the Maximum Likelihood Estimate of $\mathbf{\Theta} = (\{\mathbf{P}_k\}_K, \{\mathbf{H}^n\}^N)$, where $\{\mathbf{P}_k\}_K$ are the motion parameters of the K objects, and $\{\mathbf{H}^n\}^N$ are the set of homographies to $\mathbf{\Pi}$. For the remainder of this paper, θ_k^n represents $(\mathbf{P}_k, \mathbf{H}^n)$.

For each individual observed trajectory $\bar{\mathbf{x}}_j^i$ we have,

$$p(\bar{\mathbf{x}}_j^i | c_j^i, \mathbf{\Theta}) = p(\bar{\mathbf{x}}_j^i | \theta_{c_j^i}^i) = \prod_{t=\delta_s(i,j)}^{\delta_e(i,j)} p(\bar{\mathbf{x}}_j^i(t) | \mathbf{x}_j^i(t)), \tag{5}$$

where $\delta_s(i,j)$ and $\delta_e(i,j)$ are the start-time and end-time of O_j^i respectively[2]. Computing $\mathbf{x}_m^n(t)$ requires description of the object kinematic model, which we described in Section 2.1. Using Equation 5 and assuming conditional independence between trajectories we then have,

$$p(\bar{\mathbf{X}}, \mathbf{C} | \mathbf{\Theta}) = \prod_{i=1}^{N} \prod_{j=1}^{z(i)} p(\bar{\mathbf{x}}_j^i | c_j^i, \mathbf{\Theta}) p(c_j^i) = \prod_{i=1}^{N} \prod_{j=1}^{z(i)} \frac{1}{K} p(\bar{\mathbf{x}}_j^i | \theta_{c_j^i}^i), \tag{6}$$

where $z(i)$ denotes the total number of trajectories observed in Camera i. Thus, the complete data log-likelihood,$p(\bar{\mathbf{X}}, \mathbf{C} | \mathbf{\Theta})$ is,

$$\log p(\bar{\mathbf{X}}, \mathbf{C} | \mathbf{\Theta}) = \sum_{i=1}^{N} \sum_{j=1}^{z(i)} \log \frac{1}{K} p(\bar{\mathbf{x}}_j^i | \theta_{c_j^i}^i). \tag{7}$$

The problem, of course, is that we do not have measurements of $\mathbf{C}$ so we cannot use Equation 7 directly. Therefore, we need to find the Maximum Likelihood Estimate of $\mathbf{\Theta}$ given $\bar{\mathbf{X}}$, i.e.

$$\mathbf{\Theta}^* = \arg\max_{\mathbf{\Theta}} p(\bar{\mathbf{X}} | \mathbf{\Theta}). \tag{8}$$

To evaluate the MLE we need to (i) describe how to evaluate $\mathcal{L}(\mathbf{\Theta} | \bar{\mathbf{X}}) = p(\bar{\mathbf{X}} | \mathbf{\Theta})$ and (ii) describe a maximization routine. By marginalizing out the association in Equation 5, $p(\bar{\mathbf{x}}_k^n | \mathbf{\Theta})$ can be expressed as a mixture model,

$$p(\bar{\mathbf{x}}_j^i | \mathbf{\Theta}) = \frac{1}{K} \sum_{m=1}^{K} p(\bar{\mathbf{x}}_j^i | \theta_m^i). \tag{9}$$

[2]Evaluating $p(\bar{\mathbf{x}}_j^i(t) | \mathbf{x}_j^i(t))$ requires a measurement error model to be defined, e.g. normally distributed in which case $p(\bar{\mathbf{x}}_j^i(t) | \mathbf{x}_j^i(t)) = \mathcal{N}(\bar{\mathbf{x}}_j^i(t) | \mathbf{x}_j^i(t), \mathbf{R}_j^i)$.

Then, the incomplete data log-likelihood from the data is given by,

$$\begin{aligned}\log \mathcal{L}(\mathbf{\Theta}|\bar{\mathbf{X}}) &= \log \prod_{i=1}^{N}\prod_{j=1}^{z(i)} p(\bar{\mathbf{x}}_j^i|\mathbf{\Theta}) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{z(i)} \log \frac{1}{K}\sum_{m=1}^{K} p(\bar{\mathbf{x}}_j^i|\theta_m^i).\end{aligned}$$

This function is difficult to maximize since it involves the logarithm of a large summation. The Expectation-Maximization Algorithm provides a means to maximize $p(\bar{\mathbf{X}}|\mathbf{\Theta})$, by iteratively maximizing a lower bound,

$$\begin{aligned}\mathbf{\Theta}^+ &= \arg\max_{\mathbf{\Theta}} \mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^-) \\ &= \arg\max_{\mathbf{\Theta}} \sum_{\mathbf{C}\in\mathcal{C}} p(\mathbf{C}|\bar{\mathbf{X}}, \mathbf{\Theta}^-) \log p(\bar{\mathbf{X}}, \mathbf{C}|\mathbf{\Theta}),\end{aligned}$$

where $\mathbf{\Theta}^-$ and $\mathbf{\Theta}^+$ are the current and the new estimates of $\mathbf{\Theta}$, respectively, and $\mathcal{C}$ is the space of configurations that $\mathbf{C}$ can assume. To evaluate this expression we have,

$$p(\mathbf{C}|\bar{\mathbf{X}}, \mathbf{\Theta}^-) = \prod_{i=1}^{N}\prod_{j=1}^{z(i)} p(c_j^i|\bar{\mathbf{x}}_j^i, \mathbf{\Theta}^-), \tag{10}$$

where by Bayes Theorem and Equation 9,

$$p(c_j^i|\bar{\mathbf{x}}_j^i, \mathbf{\Theta}^-) = \frac{p(\bar{\mathbf{x}}_j^i|c_j^i, \mathbf{\Theta}^-)p(c_j^i)}{p(\bar{\mathbf{x}}_j^i|\mathbf{\Theta}^-)} = \frac{\frac{1}{K}p(\bar{\mathbf{x}}_j^i|\theta^{i-}_{c_j^i})}{\sum_{j=1}^{K}\frac{1}{K}p(\bar{\mathbf{x}}_j^i|\theta^{i-}_{c_j^i})}. \tag{11}$$

After manipulation (see [1]), we get an expression for $\mathbf{\Theta}$,

$$\begin{aligned}\mathcal{Q}(\mathbf{\Theta}, \mathbf{\Theta}^-) &= \sum_{\mathbf{C}\in\mathcal{C}} p(\mathbf{C}|\bar{\mathbf{X}}, \mathbf{\Theta}^-) \log p(\bar{\mathbf{X}}, \mathbf{C}|\mathbf{\Theta}) \\ = \sum_{m=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{z(i)} p(c_j^i = m|\bar{\mathbf{x}}_j^i, \theta_m^{i-}) &\log \frac{1}{K} p(\bar{\mathbf{x}}_j^i|\theta_m^i).\end{aligned} \tag{12}$$

In order to derive the update terms for $\mathbf{H}$ and $\mathbf{P}$, we need to make explicit the algebraic curve we are using to model the object trajectory and the measurement noise model.

If noise is normally distributed,

$$p(\bar{\mathbf{x}}_k^n|\theta_m^n) = \prod_{t=\delta_s(n,k)}^{\delta_e(n,k)} \frac{1}{(2\pi\|\mathbf{R}_m^n\|)^{\frac{1}{2}}} e^{-\frac{1}{2}d(\bar{\mathbf{x}}_k^n(t), \mathbf{x}_m^n(t))}, \tag{13}$$

where $d(\cdot)$ is the Mahalanobis distance. The probability $p(\bar{\mathbf{X}}|\mathbf{C}, \mathbf{\Theta})$ can be evaluated as follows,

$$p(\bar{\mathbf{X}}|\mathbf{C}, \mathbf{\Theta}) = \prod_{n=1}^{N}\prod_{k=1}^{z(n)}\prod_{t=\delta_s(n,k)}^{\delta_e(n,k)} \frac{1}{(2\pi\|\mathbf{R}_{c_k^n}^n\|)^{\frac{1}{2}}} e^{-\frac{1}{2}d(\bar{\mathbf{x}}_k^n(t), \mathbf{x}_{c_k^n}^n(t))}, \tag{14}$$

where

$$d(\bar{\mathbf{x}}_k^n(t), \mathbf{x}_{c_k^n}^n(t)) = (\bar{\mathbf{x}}_k^n(t) - \mathbf{x}_{c_k^n}^n(t))^T (\mathbf{R}_{c_k^n}^n)^{-1} (\bar{\mathbf{x}}_k^n(t) - \mathbf{x}_{c_k^n}^n(t)),$$

and $\mathbf{x}_{c_k^n}(t)$ is the corresponding point that lies *exactly* on the curve described by $\mathbf{P}_{c_k^n}$, and is transformed to the coordinate system of Camera n using $\mathbf{H}^n$. Explicitly,

$$[\lambda x_{c_k^n}^n(t)\ \lambda y_{c_k^n}^n(t) \lambda]^T = \mathbf{H}^n [x_{c_k^n}(t)\ y_{c_k^n}(t)\ 1]^T. \tag{15}$$

It is instructive to note that unlike the Maximum Likelihood term for independent point detections defined in terms of the reprojection error in [17], where the parameters of reprojection error function include 'error free' data points, the curve model fit on the points allows the error function to be written compactly in terms of the parameters of the curve and a scalar value denoting the position along the curve (taken here to be the time index t). This drastically reduces the number of parameters that need to be estimated.

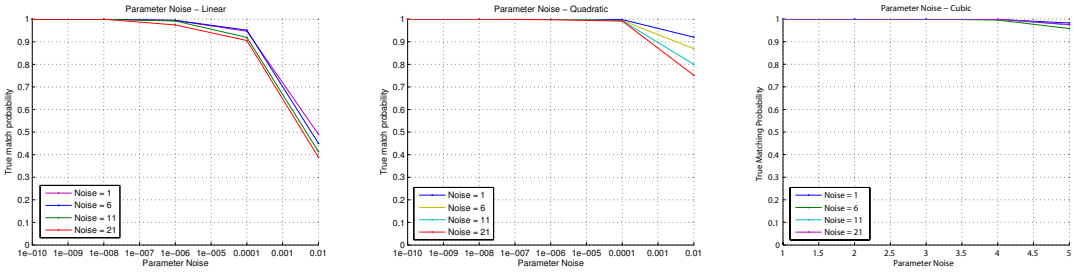
We need an analytical expression for $\log \frac{1}{K} p(\bar{\mathbf{x}}_j^i|\theta_m)$, which will then be maximized in the 'M-step'. Taking the partial derivatives, with respect to the homography and curve parameters, $\{\frac{df}{dh_1^i}, \ldots, \frac{df}{dh_9^i}, \frac{df}{dp_1^i}, \ldots, \frac{df}{dp_4^i}\}$, for each of the cameras (except the reference camera) and all the world objects, we arrive at the updating formulae. The Jacobian can then be created to guide minimization algorithms (such as the Levenberg-Marquardt algorithm).

3.1. Initialization

Good initialization of $\mathbf{\Theta}$ is an important requirement of the EM algorithm. There are several initialization methods that can be used. Ideally, for the inter-frame homographies, telemetry information, which is usually noisy, can be used for initialization. Alternatively, initial association can be computed using appearance values and initial estimates of homographies and curve coefficients can be estimated using robust methods. For the second application, i.e. reacquisition of objects in single views, the initialization is simpler: estimate of the initial homography can be computed using the frame-to-frame homography estimation, and the curve coefficients can be initialized by estimating them w.r.t to the original trajectories (before exit).

4. Experimentation and Results

We performed quantitative analysis through simulations to test the behavior of the proposed approach to noise. In addition, we show qualitative results on a number of real sequences, recovering the true underlying scene geometry and object kinematics. For the real sequences the video was collected by cameras mounted on aerial vehicles. Frame to frame registration was performed using robust direct registration methods and object detection/tracking were performed partially using an automated tracking system and partly through manual tracking.



(a) (b) (c)

Figure 1. Performance with respect to noise (a) Linear model, (b) Quadratic model, (c) Cubic model

4.1. Simulations

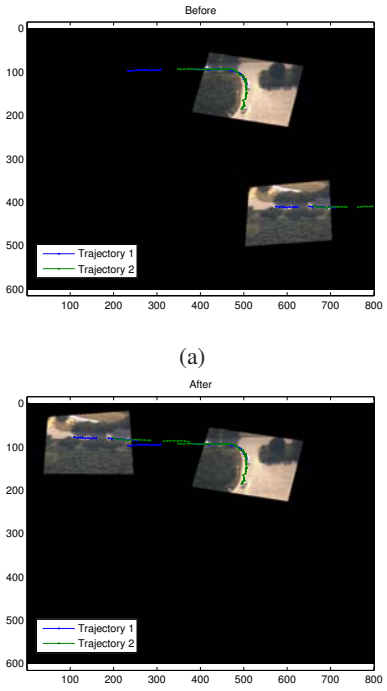
In this set of experiments we generated random trajectories fitting a prescribed model. The variable scene descriptors included number of objects, number of cameras and number of frames (observations). For each camera there was a separate probability of observation of an object, and for each object a duration of observation was randomly selected. In this way, spatio-temporal overlap was not guaranteed during data generation. A noise parameter was set for introducing errors into the true parameter values (camera parameters and curve coefficients), which were then treated as initial estimates. The homographies subtended by the camera parameters were calculated and used to project each curve onto the image, depending on its probability of observation and its duration of observation. Zero mean noise was then added to the projected points.

We tested the sensitivity of the algorithm with respect to corruption of the curve coefficients by white noise and with respect to measurement error. For these experiments five object trajectories were randomly generated according to linear, quadratic and cubic models, and two homographies (two cameras) were generated. The probability of observation was set to 1 so that both cameras were guaranteed to see both object (but not necessarily at the same time). Only 10 frames were observed, and 10 iterations of the EM algorithm were run. Four measurement noise levels were tested: 1, 6, 11 and 21, against five coefficient noise levels of $1 \times 10^{-10}, 1 \times 10^{-8}, 1 \times 10^{-6}, 1 \times 10^{-4}$ and 1×10^{-2} and each configuration was repeated 25 times (to generate statistics). This experiment demonstrates that although higher order models have a larger number of parameters to estimate, they are less susceptible to noise. This follows intuition since more information on the underlying homography is placed by each object.

4.2. Real Sequences

In this set of experiments, we study the association of objects across multiple sequences in real videos. We tested the proposed approach on three sequences. In the first sequence, several cars were moving in succession along a road, shown in Figure 2 (a) and (b). From the space-time plot it is clear that one of the objects is moving quicker than the rest of the objects (indicated by the angle with the horizontal plane). The linear (constant velocity) model was used for this experiment. Within six iterations the correct associations are discerned and, as shown in Figure 2 (c) and (d), the trajectories are correctly aligned. It should be noted that in this case the lines were almost parallel they constitute the degenerate case. However, the correct association was still found, and the alignment was also reasonable.

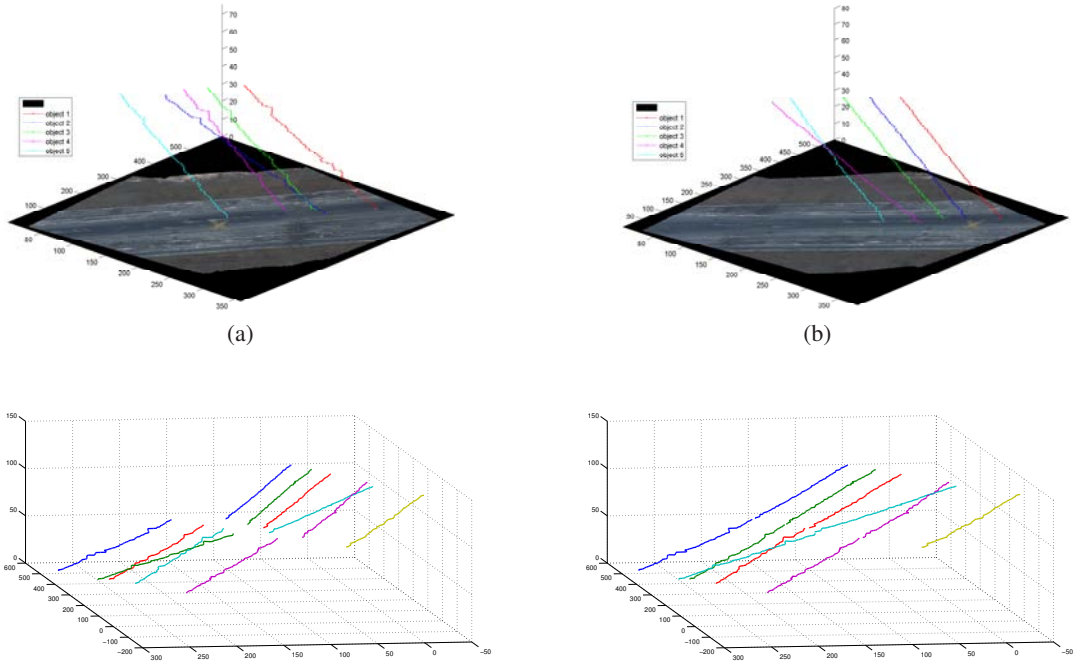
In the second experiment a quadratic kinematic model was used during experimentation in two sequences. Figure 3 shows the relative positions of the first set of sequences



(a)

(b)

Figure 3. Object Association across multiple non-overlapping cameras - Quadratic curve. (a) Initialization, (b) Converged Solution.

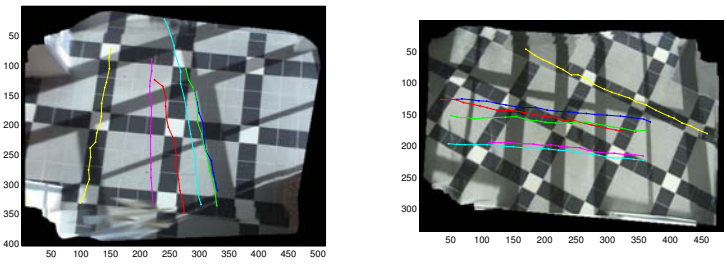


(a) (b)

(c) (d)

Figure 2. Experiment 1 - Reacquisition of objects. (a) Trajectories overlayed on the first segment mosaic, (b) Trajectories overlayed on the second segment mosaic (c) Space time plot of trajectories show that object 2 is moving faster than the rest of the objects, (d) Space time plot of trajectories of segment 2.

before (a) and after (b) running the proposed approach. can be observed that the initial misalignment was almo 400-500 pixels. It took 27 iterations of the algorithm converge. For the second set of videos, Figure 4 shows tl objects (b) before and (c) after running the proposed alg rithm. In this case the initial estimate of the homograpl was good (within 50 pixels), but the initial estimate of tl curve parameters was poor. The final alignment of the s quences is shown in Figure 4 (a). The algorithm took only iterations to converge. Finally, in Figure 5 we illustrate performance on video taken from two overhead cameras looking at people walking. The color-code of each trajectory shows the association across views recovered by the algorithm. Due to the large rotation present between the views the algorithm took a large number of iterations were executed (39 iterations).



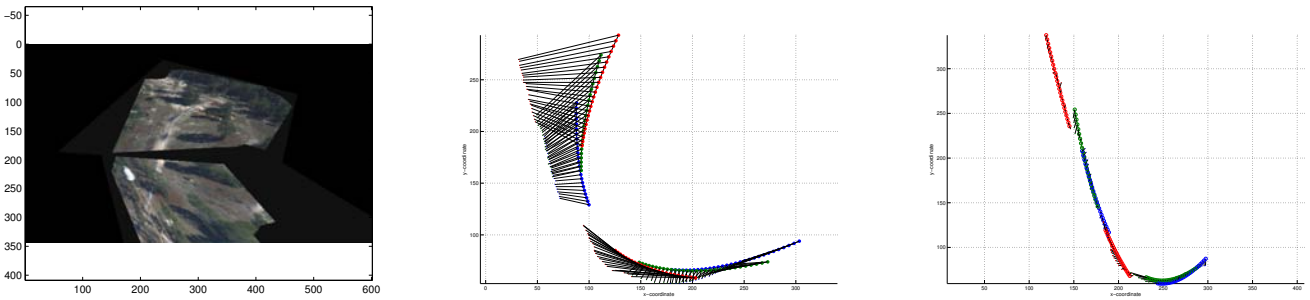
(a) (b)

Figure 5. Overhead view of people walking. (a) Shows the color-coded trajectories viewed from the first camera, (b) shows the same trajectories from the second camera.

5. Conclusion and Future Work

In this paper, we solve the trajectory association problem across multiple non-overlapping views by explicitly modeling the motion of objects. Instead of a two-step approach of fitting motion models and then associating objects, we present an algorithm that simultaneously estimates the parameters of the motion and the inter-camera homographies. The associations are treated as hidden variables and the Expectation-Maximization algorithm is used to compute the maximum likelihood estimates of the unknown parameters of object motion and inter-camera transformation. We report both quantitative results on simulated data and qualitative results on real data.

The main theme in this paper has been the recovery of a coherent understanding of the world (homographies of cameras and canonical trajectories) given imaged data at each camera. To that end, we have investigated better models



(a) (b) (c)

Figure 4. Object Association across multiple non-overlapping cameras - Quadratic curve. (a) Initialization, (b) Converged Solution.

for the scene for association across multiple cameras where spatiotemporal overlap cannot be assumed. This theme has led us to pose a model that reflect the geometry of the scene while capturing the uncertainty and incompleteness of data at each camera. An interesting addition that would fit seamlessly into the proposed framework is co-operative tracking of objects multiple cameras. Currently, we assume that the single camera tracking problem is solved, but this problem can also be solved simultaneously. Instead of associating trajectories, we would associate points in the same manner. The result would be a genuinely co-operative ensemble of cameras where the information obtained by each camera would improve the estimate of object positions at a certain point and this paper lays the framework for just such a system.

Acknowledgements

This research was funded in part by the US Government VACE program.

References

[1] J. Bilmes. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. In *Technical Report, University of Berkeley*, 1997.

[2] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade. Algorithms for cooperative multisensor surveillance. In *Proceedings of the IEEE*, 2001.

[3] Y. B.-S. (Editor). *Multitarget-Multisensor Tracking: Advanced Applications*. Artech House, 1990.

[4] R. Fisher. Self-organization of randomly placed sensors. In *Proceedings of the European Conference on Computer Vision*, 2002.

[5] A. Fitzgibbon and R. Fischer. A buyer's guide to conic fitting. In *British Machine Vision Conference*, 1995.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, September 2000.

[7] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.

[8] J. Kaminski and A. Shashua. Multiple view geometry of general algebraic curves. In *International Journal of Computer Vision*, 2004.

[9] S. Kang and K. Ikeuchi. Toward automatic robot instruction from perception – mapping human grasps to manipulator grasps. In *IEEE Trans. on Robotics and Automation*, volume 12, Dec. 1996.

[10] V. Kettnaker and R. Zabih. Bayesian multi-camera surveillance. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[11] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[12] T. Matsuyama and N. Ukita. Real-time multitarget tracking by a cooperative distributed vision system. In *Proceedings of the IEEE*, 2002.

[13] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[14] Y. Shan, H. Sawhney, and R. Kumar. Unsupervised learning of discriminative edge measures for vehicle matching between non-overlapping cameras. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.

[15] Y. Sheikh and M. Shah. Object tracking across multiple independently moving airborne cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, 2005.

[16] C. Stauffer and K. Tieu. Automated multi-camera planar tracking correspondence modelling. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[17] P. Sturm. Vision 3d non calibrée - contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage. In *PhD Thesis*, 1997.

[18] K. Tieu, G. Dalley, and E. Grimson. Inference of non-overlapping camera network topology by measuring statistical dependence. In *Proceedings of the IEEE International Conference on Computer Vision*, 2005.