A Bayesian Approach to Simultaneous Motion Estimation of Multiple Independently Moving Objects [*]

Gang Qian, Rama Chellappa and Qinfen Zheng
Center for Automation Research and
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742-3275
{gqian,rama,qinfen}@cfar.umd.edu

Abstract

In this paper, the problem of simultaneous motion estimation of multiple independently moving objects is addressed. A novel Bayesian approach is designed for solving this problem using the sequential importance sampling (SIS) method. In the proposed algorithm, a balancing step is added into the SIS procedure to preserve samples of low weights so that all objects have enough samples to propagate empirical motion distributions. By using the proposed algorithm, the relative motions of all moving objects with respect to camera can be simultaneously estimated . This algorithm has been tested on both synthetic and real image sequences. Improved results have been achieved.

1. Introduction

Simultaneous estimation of 3D motion between each moving object and camera has remained a big challenge to researchers in computer vision. In this paper, we refer to this problem as the *Multiple Independently Moving Objects Motion Estimation* (MIMOME) problem: *given an image sequence containing two or more independently moving (with respect to the camera) objects, one desires to simultaneously estimate the motion between each moving object and the camera.*

Existing MIMOME algorithms can be roughly categorized into two groups based on the type of tokens used in the method: optical flow or feature correspondences. The pioneering work of MIMOME using optical flow was proposed by Adiv [1]. In [1], with the assumption of piecewise planar scene, the flow field is first partitioned into connected segments using the generalized Hough technique and then the segments belonging to each moving object are grouped such that all segments in one group can be explained by a single rigid motion. Due to the assumption made about the scene structure, this algorithm has difficulties in dealing with image sequences containing complex 3D scene structure with rich higher-order 3D curved surfaces.

Another main stream approach to MIMOME is based on feature correspondences. A multi-body Factorization method [2] was designed for solving the MIMOME problem after the well-known Factorization method for structure from motion (SfM) [3]. The shape interaction matrix is employed to segment out multiple moving objects. In spite of the elegance of this algorithm, the assumption made about the camera projection model is restrictive.

The extended Kalman filter (EKF) has also been used to solve the MIMOME problem. In [4], an EKF-based MIMOME algorithm was proposed based on a recursive SfM algorithm presented in [5]. The minimum description length (MDL) criterion is employed to determine the minimum number of models and their corresponding feature groups such that the models can well interpret the trajectories of the features in the associated groups. However, the recursive SfM algorithm in [5] is not guaranteed to converge to the true SfM solution, hence the MIMOME algorithm based on this method is doubtful. Another MIMOME algorithm using EKF is proposed in [6]. It is designed based on the motion estimation algorithm called *essential filter* [7]. The main problem with this approach is that the essential filter basically computes the relative motion of objects between adjacent image frames. This brings the feature based method to the flow based method in the sense of low signal-to-noise ratio (SNR). Due to inherent ambiguities present in the SfM problem when the observations are noisy, the essential filter might converge to a false solution and yield over-segmentation of moving objects.

Recently the MIMOME problem has been considered under certain restrictions on trajectories of independently

---

moving objects. A MIMOME algorithm was developed in [8] assuming linear motion (constant velocity) of dynamic points with orthographic projection camera model. In another algorithm presented in [9], the dynamic points are also assumed to move along straight-line paths with constant velocity.

In this paper, we present a novel Bayesian approach for solving the MIMOME problem. We use feature correspondences and the perspective projection camera model. We do not make any assumption about either the 3D scene structure or the motion of moving objects except that all objects are required to be rigid.

Relation to previous work In [10], a Bayesian moving object detection algorithm was presented. Given feature trajectories obtained from an image sequence recored using a moving camera, features on moving objects can be separated from those on the background. Moving objects can thus be detected. In this approach, a state space model was used to describe the camera motion and feature segmentation parameters and their dynamics. The sequential importance sampling (SIS) method was used to take care of the non-linear observation equation, which is a result of using perspective projection camera model. During the SIS procedure, state samples and their related weights, which represent the posterior distribution of state parameters, are propagated from one time instant to the next time instant. Ideally, when sufficiently large number of samples are used in the SIS procedure, the moving object detection algorithm proposed in [10] is able to simultaneously estimate the motion parameters for each independently moving object. However, due to the fact that only a finite number of samples can be used in practice, if one SIS procedure is used for all moving objects, samples of different moving objects will compete against each other for the number of offsprings in the resampling step. It is very possible that the samples of one object can not get enough offsprings so that they are eliminated from this competition, i.e. they just disappear from the SIS procedure! In this paper, an additional step called *balancing* is inserted into the original SIS procedure so that each potential moving object candidate can get approximately equal number of offsprings for its samples.

2. MIMOME Using SIS

To solve the MIMOME problem, we need to find optimal estimates for object motions and feature segmentation which can provide the best interpretation of the observations using criteria such as the maximum *a posteriori* probability (MAP). In this paper, $Prob(parameters|observation)$, the posterior distribution of the parameters, is approximated using random sampling methods. In this section, we first formulate the MIMOME problem using a state space model. Then we introduce two essential steps: clustering and balancing. Finally, a complete SIS algorithm for finding ap-
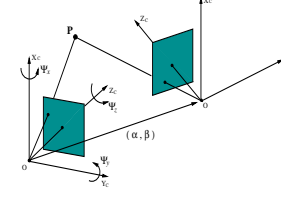

Figure 1. Imaging model of a moving camera

proximations to the posterior distribution of motion and feature segmentation parameters is developed.

2.1 State Space Model for Object Motion and Feature Segmentation

Parameterization of relative motion between objects and camera The relative motion between each object and camera needs to be estimated. During the problem formulation, either the object or the camera can be assumed to be static. In our approach, we assume that the objects are static and the camera motion relative to individual object is estimated. Once the camera motion is obtained, it is very straightforward to compute the object motion relative to the camera if it is desired.

The parameterization of the camera motion is as follows. Two 3D Euclidean coordinate systems are used in the formulation. One coordinate system is attached to the camera and uses the center of projection of the camera as its origin. It is denoted by C. The Z axis of C is along the optical axis of the camera, with the positive half-axis in the camera looking direction. The X-Y plane of C is perpendicular to the Z axis with the X and Y axes parallel to the borders of the image plane. Also, the X-Y-Z axes of C satisfy the right-hand rule. The other coordinate system is a world inertial frame, denoted by I. I is fixed on the ground. The coordinate axes of I are configured in such a way that initially, I and C coincide. When the camera moves, C travels with the camera and I stays at the initial position. At time instant t, five parameters are used to describe camera motion:

$$\mathbf{m}_t = (\psi_x, \psi_y, \psi_z, \alpha, \beta)$$

(ψ_x, ψ_y, ψ_z) are the rotation angles of the camera about the coordinate axes of the inertial frame I and (α, β) are the elevation and azimuth angles of the camera translation direction, measured in the world system I. The unit vector in the translation direction is given by $\mathbf{T}(\alpha, \beta) = (\sin(\alpha)\cos(\beta), \sin(\alpha)\sin(\beta), \cos(\alpha))^T$.

Remark *The motion parameters of the camera at time instant t are all relative to the world inertial coordinate system, which is the camera coordinate system at the first time instant in the current coordinate system configuration.*

Parameterization of feature point segmentation In our approach to solving the MIMOME problem, the input to the

algorithms are trajectories of a set of feature points tracked throughout the given image sequence. These feature points belong to different moving objects. The task of feature point segmentation is to separate this feature set into multiple subsets and ideally, the features on the same object should be on the same subset.

To represent the ownership of the feature points by different moving objects, we use the validity vector $\mathbf{v}$, introduced in [10]. If there are M feature points used, $\mathbf{v}$ is an M-dimensional vector. Each feature point has a corresponding entry in $\mathbf{v}$, which indicates the belonging of the related feature point. The parameters for camera motion and feature segmentation contains both motion parameters and the validity vector:

$$\mathbf{x}_t = (\mathbf{m}_t, \mathbf{v}_t) \tag{1}$$

In the SIS procedure, the posterior distribution of the state parameters is described by samples representing state and related weights. Given a state sample, the motion parameters represent the camera motion relative to a particular moving object. Associations of feature points to this moving object are described by the validity vector in the same state sample. If the value of the entry in the validity vector of a feature point is positive and high, it is very probable that this feature point belongs to this particular moving object; otherwise, it is on some other objects.

State space model Given the above state parameters, a state space model can describe the behavior of a moving camera observing multiple independently moving objects:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{n}_x \tag{2}$$
$$\mathbf{y}_t = Proj(\mathbf{x}_t, \mathcal{S}_t) + \mathbf{n}_y \tag{3}$$

where $\mathbf{x}_t$ is the state vector and $\mathbf{y}_t$ is the observation at time t. $Proj(\cdot)$ denotes the perspective projection, a function of camera motion $\mathbf{x}_t$ and the scene structure $\mathcal{S}_t$. $\mathbf{n}_x$ denotes the dynamic noise in the system, describing the time-varying property of the state vector and it contains both the dynamic noise of the camera motion and the validity vector. Even though the associations of the feature points are fixed, if we assume that their is no object splitting/merging occurred, validity vectors are still time varying during the SIS procedure in this formulation. By allowing the validity vector to be time-varying, the resulting temporal integration of the feature segmentation makes the feature segmentation results more robust to measurement errors.

In order to deal with the non-linear system, the SIS is used to approximate the posterior distributions of the motion parameters and validity vectors. Due to space limitations, we will not introduce SIS in this paper and refer the readers to [11] for a full treatment. In the SIS procedure, given samples at previous time instant, the new state samples at current time instant are drawn using the so-called trial function. Although any trial function can be used, to design effective SIS

algorithms, trial functions that can describe the dynamics of the system are desired. For motion parameters, as the prior knowledge about motion parameters is not available, a random walk is a suitable alternative for modeling the camera position and orientation. Hence, the motion samples at $t+1$ are drawn using

$$\mathbf{m}_{t+1} = \mathbf{m}_t + \mathbf{n}_m \tag{4}$$

where $\mathbf{n}_m$ represents the dynamic noise of the motion parameters. The samples of the validity vector at $t + 1$ are drawn using

$$\mathbf{v}_{t+1} = \gamma \mathbf{v} + \xi(\mathbf{m}_t, \mathbf{y}_t) + \mathbf{n_v} \tag{5}$$

where $\mathbf{n_v}$ is the dynamic noise in the validity vector and γ is an exponential forgetting factor. Both of them represent the possible time-varying nature of the validity vector. $\xi(\cdot)$ is a function used to update the current validity vector. Each element of ξ is given by

$$\xi_i(\mathbf{m}_t, \mathbf{y}_t) = \left(\frac{e_{th}}{e_i + 1}\right)^2 - sign(\lfloor e_i/e_{th} \rfloor)\frac{e_i + 1}{e_{th}} \tag{6}$$

where $e_i = e(\mathbf{m}_t, \mathbf{y}_t^{(i)})$ is the distance from the ith feature point to its associated epipolar line given the motion parameters $\mathbf{m}_t$. e_{th} is a prechosen threshold for this distance according to the feature tracking error level. In our experiments, the values e_{th} are usually twice as large as the standard deviation of the feature tracking errors.

The likelihood function of the observation given the state parameter is obtained as

$$f(\mathbf{y}_t|\mathbf{x}_t) \propto I_{\{\sum \frac{\mathbf{v}_t(i)}{|\mathbf{v}_t(i)|} > 7\}}(\mathbf{v}_t) \sum_{i \in \mathcal{A}} \mathbf{v}_t^+(i) \exp\left\{\frac{-\epsilon}{\sigma_u^2 + \sigma_v^2}\right\} \tag{7}$$

where $\mathcal{A} = \{i | \mathbf{v}_t(i)$ is among the 7 highest value of $\mathbf{v}_t\}$ and

$$\mathbf{v}_t^+(i) = \begin{cases} \mathbf{v}_t(i), & \mathbf{v}_t(i) > 0 \\ 0, & \text{otherwise} \end{cases}, \epsilon = \frac{\sum_{i \in \mathcal{A}} e_t(i)^2 \mathbf{v}_t^+(i)}{\sum_{i \in \mathcal{A}} \mathbf{v}_t^+(i)} \tag{8}$$

$I_{f(\mathbf{v})}(\mathbf{v})$ is an indicator function. It returns one if $f(\mathbf{v})$ is true and zero otherwise.

2.2. Sample Clustering

In the SIS procedure, the motion and validity vector samples and their corresponding weights are propagated such that they are properly weighted with respect to the posterior distribution of the motion parameters associated with the moving objects. However, the samples of motion and validity vectors of moving objects are mixed together. Samples of state parameters (including motion parameters and validity vector) are clustered into multiple groups and each group is related to one potential candidate of the moving object. The clustering method is described as follows.

Sign-based Sample Clustering Algorithm

1. Sample sorting. Let $\mathcal{X} = \{\mathbf{x}^{(i)}\} = \{\mathbf{m}^{(i)}, \mathbf{v}^{(i)}\}$ be the samples of motion and validity vector before the resampling step in the SIS procedure and $\mathcal{W} = \{\mathbf{w}^{(i)}\}$ be the corresponding weights of the above state samples. Sort the samples according to their weights such

that the first sample has the highest weight. Denote the sorted sample set and weight set as $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{W}}$, respectively.

2. Sample cluster center generation. Starting from the first sample in $\tilde{\mathcal{X}}$, candidate cluster centers are generated by looking at the signs of samples in the new sample set. Assume that up to the current time instant, there are K clusters $\{\mathcal{S}_k\}_{k=1}^{K} = \{\mathcal{X}_k \mathcal{W}_k\}_{k=1}^{K}$ and each cluster has a cluster center. Given a sample of the state, first we ignore the entries in the validity vector that have been occupied by any positive entry of the validity vectors belonging to existing candidate centers. Then, we count the number of remaining positive entries. If this number is larger than or equal to P_m, the required number of feature points for motion estimation (e.g. eight points are sufficient when the camera is fully calibrated) the $K + 1$st candidate cluster is generated with the current sample.

3. Sample clustering. The samples that contain at least P_m common positive entries in the validity vector as the new cluster center will be segmented into that cluster. Thus a new sample weight cluster is formed as $\mathcal{S}_{K+1} = \{\mathcal{X}_{K+1} \mathcal{W}_{K+1}\}$. Go to step 2 and look for new candidate cluster centers. Stop when all the samples in $\tilde{\mathcal{X}}$ have been visited by the center generation step.

4. Redistribution. It is possible that a sample is not assigned to any cluster. To preserve these samples, the samples not included in any cluster are grouped into a new cluster. Hence, all the samples in $\tilde{\mathcal{X}}$ are segmented into different clusters. Each cluster corresponds to one candidate moving object.

By using the above clustering algorithm, the samples in $\tilde{\mathcal{X}}$ are segmented into different clusters related to potential moving objects.

2.3. Balancing

The balancing step is used in the SIS procedure so that the numbers of samples belonging to all of the potential moving objects are comparable. Balancing can prevent samples associated with different moving objects from competing for offsprings so that samples with lower weights due to smaller size of the object or higher levels of noise in feature correspondences will not be extinct when a limited number of samples are used. Hence, during SIS, each moving object has its state samples in the whole sample population and the motion/structure parameters of all objects can be obtained simultaneously from their associated samples. To handle the problem due to samples belonging to different objects competing for offsprings during the resampling step, the following balancing step is added in the SIS procedure.

Definition *For a sample weight cluster* $\mathcal{S}_k = \{\mathcal{X}_k, \mathcal{W}_k\}$, W_k, *the* **cumulative weight** *of* $\mathcal{S}_k$, *is the summation of the*

weights in $\mathcal{W}_k = \{w_k^{(j)}\}_{j=1}^{m}$, *i.e.* $W_k = \sum_{j=1}^{m} w_k^{(j)}$

Balancing step

1. Cluster weight evaluation. Assume that after the above sample clustering procedure, totally K sample clusters are formed including the one produced in the finalization step. Since each cluster should be treated equally in the resampling step in the SIS, the cumulative weights of all clusters should be equal to W^* and $W^* = \frac{1}{K}$.

2. Weights modification. The sample weights in all clusters need to be modified such that the cumulative weight of each individual cluster equals to W^*. The weight modification is done as below. For a cluster $\mathcal{S}_k = \{\mathcal{X}_k \mathcal{W}_k\}$, multiply the weights in $\mathcal{W}_k$ by a scalar λ_k and $\lambda_k = \frac{W^*}{W_k}$ where W_k is the cumulative weight of this cluster before the weight modification step. This step is called *weight-balancing*. Denote the weight-balanced cluster as $\hat{\mathcal{S}}_k = \{\mathcal{X}_k \ \hat{\mathcal{W}}_k\} = \{\mathcal{X}_k \lambda_k \mathcal{W}_k\}$.

2.4. SIS Procedure for MIMOME

By adding clustering and balancing steps, we have the following complete SIS procedure for solving the MIMOME problem.

1. Initialization. Draw samples of motion parameters $\{\mathbf{m}_0^{(j)}\}_{j=1}^{m}$ from the initial distribution π_0. π_0 describes the distribution of motion parameters $\mathbf{m}_0$ before camera moves. Although the rotation angles ψ and the translational vector are all zero, the translational angles can be uniformly distributed. Hence, in $\{\mathbf{m}_0^{(j)}\}$, the components of the rotation angles are all set to zero and the samples of α and β are drawn from the uniform distributions in $[0, \pi]$ and $[0, 2\pi]$, respectively. The components in the samples corresponding to the validity vector are set to one. Assign equal weights to above samples.

For $t = 1, \cdots, F$:

2. Samples generation. Draw samples of the motion parameters at time instant t, $\{\mathbf{m}_t^{(j)}\}_{j=1}^{m}$, from the distributions of $\{\mathbf{m}_{t-1}^{(j)}\}_{j=1}^{m} + n_m$. Since video sequences are used, a random walk dynamic model is assumed and the following distribution can be used as a good approximation to that of n_m.

$$\begin{cases} n_{\psi_\iota} \sim \mathcal{N}(0, \sigma_\iota), \iota \in \{x, y, z\} \\ n_\kappa \sim U(-\delta_\kappa, \delta_\kappa), \kappa \in \{\alpha, \beta\} \end{cases} \quad (9)$$

where $\sigma_\iota, \delta_\alpha$ and δ_β can be chosen as some positive numbers. Draw samples of the validity vector $\{\mathbf{v}_t^{(j)}\}_{j=1}^{m}$ via (5).

3. Weight computation. Compute the weights of the samples, $\{w_t^{(j)}\}$, using the observed feature correspondence according to the likelihood equation (7). The

resulting samples and their corresponding weights $(\mathbf{x}_t^{(j)}, w_t^{(j)})$ are properly weighted with respect to $\pi_t(\mathbf{x}_t)$.

4. Sample clustering. Segment the whole samples and weights set into multiple clusters using the above sign-based sample clustering algorithm.

5. Balancing. Modify the weights of each cluster so that the cumulative weights of the clusters are equal.

6. Resampling. Resample the samples in the clusters according to the balanced weights. Thus, the motion/structure and validity vector related to the two moving objects can be simultaneously estimated. Go to step 2 if the feature correspondences from the next frame are available.

3. Experimental Results

The proposed MIMOME algorithm has been tested using both synthetic and real image sequences. Two examples using real image sequences are included. In the first example, an image sequence captured by a moving camera is used. In this image sequence, two persons are walking in a room. One is walking to the right of the scene and the other is moving forward along the optical axis of the camera. Since the camera is also moving, the background scene serves as the third moving object with respect to the camera. 44 feature points are detected and tracked through the sequence. Figure 2 (a) shows feature points (marked by squares) in the last image of the sequence and their trajectories. By using the proposed algorithm, features on different moving objects can be clustered and at the same time, the motion of each moving object can be estimated. Figure 2 (b) (c) and (d) show the segmentation results of the feature points belonging to the first, the second persons and background, respectively. The points segmented out for the object are marked by circles. Posterior distributions of the motion parameters and the empirical means of the validity vector of these three objects are shown in Figure 5. The first five figures at the top of the first column of Figure 5 are the marginal posterior distributions of the motion parameters. The last one in this column shows the empirical means of the validity vector associated with the first person (walking to the right) where the horizontal axis indicates the feature indices. The results related to the second person and the background are shown in the figures in the second and third columns, respectively. Figure 3 shows the empirical means (thick curves) and the standard deviation (thin bars) of the validity vectors using the clustered samples related to the background in the first example. Figures 3 (a), (b) and (c) show these values at the initial, the middle and the last time instants. It can be seen from these figures that as time goes by, the mean of the validity vector gradually evolves to a desired structure that the entries of the features on the background have large positive
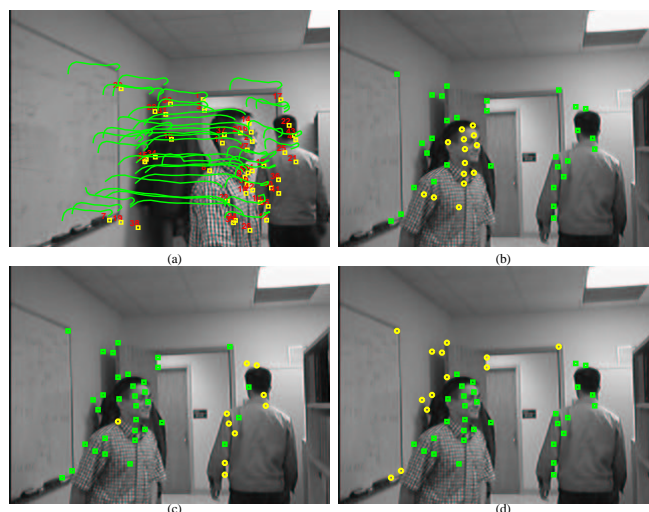

Figure 2. (a) shows feature points (marked by squares) in the last image of the sequence and their trajectories. Feature points belonging to different objects are marked by circles in (b), (c) and (d). (b) shows the features on the first person (walking to the right) and (c) shows the features on the second person (walking along the optical axis) and (d) shows the features on the background scene.

values while the others negative. The standard deviations of the validity vector are also decreasing.

In the second example, we used an image sequence with three walking persons. This sequence was recorded using a static camera with known background. After background subtraction, image regions belonging to moving objects can be found. Feature points on moving object can hence be detected in these regions and tracked throughout the whole sequence. In our experiment, we used 50 features and their positions in the last frame of the sequence are indicated by squares in Figure 4 (a), where the curves are the trajectories of features. From left to right, the three persons are referred by the first, second and third persons. Feature segmentation results for the three walking persons are shown by Figure 4 (b) (c) and (d), respectively. Several features on the second person's legs were not correctly marked. The reason is that these features were not correctly tracked during feature tracking. It can be seen in Figure 4 (a) that these features drifted to the right foot of the person. Despite this and a small number of errors, the majority of the feature points are successfully clustered on the correct persons. The empirical posterior distributions of motion and validity vector parameters related to these three persons have been estimated by samples and their weights. Due to space limitation, details on estimation of empirical posterior distribution have been omitted from this paper.

4. Conclusions

In this paper, a Bayesian approach is designed to solve the MIMOME problem using SIS. In the proposed method,

a balancing step is added into the SIS procedure to preserve samples of low weights so that all objects have enough samples to propagate empirical motion and validity vector distributions. The proposed approach has been tested on both synthetic and real image sequences. Satisfactory results have been obtained. Moreover, similar SIS algorithms with balancing can be applied to other applications when one needs to simultaneously estimate parameters of multiple sources using mixed data from all these sources.
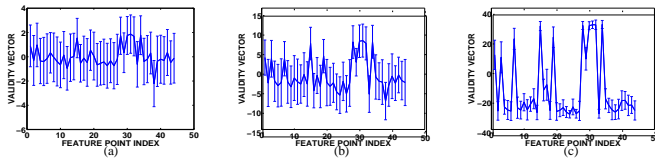

Figure 3. Temporal evolution of validity vector samples. These figures show the empirical means (thick curves) and the standard deviation (thin bars) of the validity vectors using the clustered samples related to the background in the first example. (a), (b) and (c) show these values at the initial, middle and last time instants.
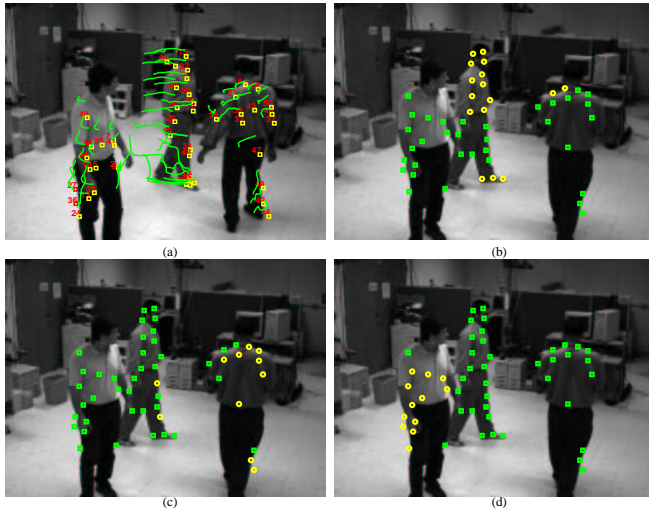

Figure 4. (a) shows feature points (marked by squares) in the last image of the sequence and their trajectories. Feature points belonging to the first, second and third persons are marked by circles in (b), (c) and (d), respectively.

References

[1] G. Adiv, "Determining 3-D motion and structure from optical flow generated by several moving objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **7**, pp. 384–401, July 1985.

[2] J. Costeira and T. Kanade, "A multibody factorization method for independently moving-objects," *International Journal of Computer Vision* **29**, pp. 159–179, September 1998.

[3] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision* **9**, pp. 137–154, November 1992.

[4] T. Darrell, A. Azarbayejani, and A. Pentland, "Robust estimation of multiple models in the structure from motion domain," Tech. Rep. TR-281, MIT, Media Lab, 1994.

[5] A. Azarbayejani and A. Pentland, "Recursive estimation of motoin, structure, and focal length," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17**, pp. 562–575, 1995.

[6] S. Soatto and P. Perona, "3d transparent structure segmentation and multiple motion estimation from monocular perspective image sequences," in *Proc. of the IEEE Workshop on motion of nonrigid and articulated objects, Austin, TX*, pp. 228–235, IEEE Computer Society Press, (Los Alamitos, CA), 1994.

[7] S. Soatto, R. Frezza, and P. Perona, "Motion estimation on the essential manifold," in *European Conference on Computer Vision, Stockholm, Sweden*, pp. B:61–72, 1994.

[8] M. Han and T. Kanade, "Reconstruction of a scene with multiple linearly moving objects," in *IEEE Computer Vision and Pattern Recognition, Hilton Head, SC*, pp. II:542–549, 2000.

[9] A. Shashua and A. Levin, "Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects," in *International Conference on Computer Vision, Vancouver, Canada*, pp. II: 592–599, 2001.

[10] G. Qian and R. Chellappa, "Moving targets detection using sequential importance sampling," in *IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT*, 2001.

[11] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *J. Amer. Statist. Assoc.* **93**, pp. 1032–1044, 1998.
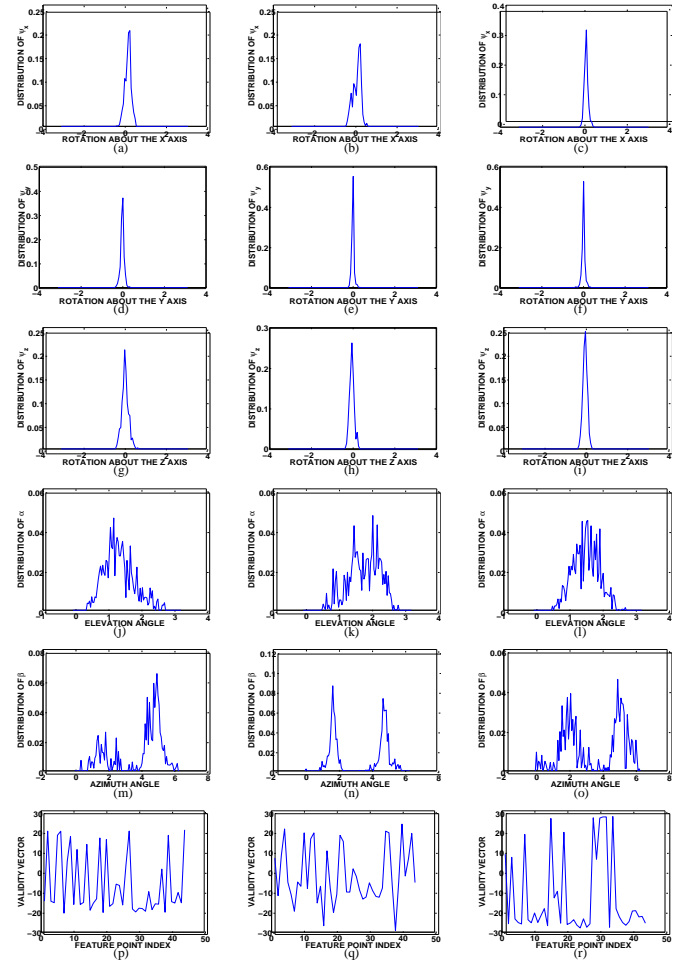
Figure 5. The first five figures at the top of the first column are the marginal posterior distributions of the motion parameters and the last one shows the empirical means of the validity vector associated with the first person (walking to the right) and the results related to the second person and the background are shown in the figures in the second and third columns, respectively.