MOVING TARGETS DETECTION USING SEQUENTIAL IMPORTANCE SAMPLING

Gang Qian and Rama Chellappa

Center for Automation Research
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742-3275

ABSTRACT

In this paper, we present a new technique for detecting moving targets from image sequences captured by moving sensors. Feature points are detected and tracked through the image sequences. A validity vector is used to describe the consistency of feature trajectories with sensor motion. By using the sequential importance sampling method, an approximation to the posterior distribution of the sensor motion and the validity vector is derived and the feature points belonging to the moving target are then segmented out. Real image examples are included.

1. INTRODUCTION

Detection of moving targets from a moving platform is a very important task in video surveillance applications. Mainly three types of moving targets detection algorithms have been developed. The first type of algorithms are *2D* algorithms [1]. They are applicable when the observed scene can be well modeled by a flat surface or the camera only rotates and zooms. The *2D* algorithms have difficulties in processing sequences containing rich variations in the 3D scene structure. The second type of algorithms are called *3D* algorithms [2], where the structure from motion (SfM) problem is solved using feature correspondences from the sequences. Multiple solutions to camera motion relative to the background and moving targets are typically found. Although the *3D* algorithms are supposed to handle sequences with general camera motion and scene structure, due to the inherent ambiguity in SfM [3], the *3D* algorithms can work well only in some specific cases such as when the camera can be modeled by orthographic projection, etc. The third approach is called *plane+parallax* algorithm [4]. In [4], a plane registration process using the dominant 2D parametric transformation is applied to remove the effects of camera rotation, zoom and calibration. Since the residual motion field is due to the presence of moving objects and camera translation, the moving targets are identified by looking at image regions violating the epiploar constraints if the global epipole can be extracted correctly. However, it is a challenge to extract the global epipole (or focus of expansion) from a noisy translational motion field corrupted by moving objects. Various robust regression techniques such as M-estimators and random sampling consensus paradigm (RANSAC) have also been used for computing the fundamental matrix [5] and then the moving targets are segmented out. A common drawback of the above methods is that since only two, three or four frames from the sequences are used to detect moving targets, the close temporal relationship of the motion pattern of the moving targets in a video sequence is not explored.

Given a series of image frames, the moving target detection result (noisy or might be partially wrong) obtained from the previous frames should be able to help refine moving target detection using current and incoming frames and make the detection algorithm more robust to observation noise in a recursive fashion. Although the importance of temporal correlation in the motion field has received some attention in detecting independent motion [6] from 2D scenes (aerial sequences), a sound theoretical computational framework is still needed to utilize the temporal correlation in detecting moving targets in both 2D and 3D scenes.

In this paper, we develop a recursive algorithm for moving targets detection. A validity vector is used to describe the segmentation of the feature points. During recursion, the validity vector evolves such that the entries corresponding to the background have large positive values while other entries corresponding to the feature points on the moving objects have negative values. By using a sampling method called sequential importance sampling (SIS) [7], an approximation to the posterior distribution of the camera motion and the validity vector is obtained. Our approach can deal with both 2D and 3D scenes and since the temporal relationship of the moving targets in the entire video sequence is taken into account during the sequential sampling procedure, our algorithm is robust to observation noise and gives better targets detection results.

2. THEORY OF SEQUENTIAL IMPORTANCE SAMPLING

The SIS technique is a recently proposed method for approximating the posterior distribution of the state vector for a pos-

Prepared through collaborative participation in the Advanced Sensors Consortium (ASC) sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0001.

sibly nonlinear dynamic system [7]. It is a very practical tool for prediction, filtering and smoothing of nonlinear and/or non-Gaussian state space models and has been used in object tracking [8] as well as verification [9] where the algorithm is usually called the *Condensation* algorithm. Usually, the state space model of a dynamic system is described by observation and state equations. If the measurement is denoted by $\mathbf{y}_t$ and the state parameter by $\mathbf{x}_t$, essentially, the observation equation provides the conditional distribution of the observation given the state, $f_t(\mathbf{y}_t|\mathbf{x}_t)$. Similarly, the state equation gives the Markov transition distribution from time t to the next time, $q_t(\mathbf{x}_{t+1}|\mathbf{x}_t)$. The goal is to find the posterior distribution of the states $\mathcal{X}_t = (\mathbf{x}_1, \cdots, \mathbf{x}_t)$ given all the available observations up to t, $\pi_t(\mathcal{X}_t) = P(\mathcal{X}_t|\mathcal{Y}_t)$ where $\mathcal{Y}_t = \{\mathbf{y}_i\}_{i=1}^t$. One way to represent an approximation to the posterior distribution is by a set of samples and their corresponding weights.

Definition [7] *A random variable X drawn from a distribution g is said to be* **properly weighted** *by a weighting function w(X) with respect to the distribution π if for any integrable function h,*

$$E_g h(X) w(X) = E_\pi h(X).$$

A set of random draws and weights $\{x^{(j)}, w^{(j)}\}_{j=1}^m$, is said to be properly weighted with respect to π if

$$\lim_{m\to\infty} \frac{\sum_{j=1}^m h(x^{(j)}) w^{(j)}}{\sum_{j=1}^m w^{(j)}} = E_\pi h(X)$$

for any integrable function h.

Suppose $\{\mathcal{X}_t^{(j)}\}_{j=1}^m$ is a set of random samples properly weighted by the set of weights $\{w_t^{(j)}\}_{j=1}^m$ with respect to π_t and let g_{t+1} be a trial distribution. Then the recursive SIS procedure that obtains the random samples and weights properly weighting π_{t+1} is as follows.

SIS steps: for $j = 1, \cdots, m$,
(A) Draw $X_{t+1} = \mathbf{x}_{t+1}^{(j)}$ from $g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})$. Attach $\mathbf{x}_{t+1}^{(j)}$ to form $\mathcal{X}_{t+1}^{(j)} = (\mathcal{X}_t^{(j)}, \mathbf{x}_{t+1}^{(j)})$.
(B) Compute the "incremental weight" u_{t+1} by

$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\mathcal{X}_{t+1}^{(j)})}{\pi_t(\mathcal{X}_t^{(j)}) g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})}$$

and let $w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$.
It can be shown [7] that $\{\mathcal{X}_{t+1}^{(j)}, w_{t+1}^{(j)}\}_{j=1}^m$ is properly weighted with respect to π_{t+1}.

Hence, the above SIS steps can be applied recursively to get the properly weighted set for future time instants when corresponding observations are available. The choice of the trial distribution g_{t+1} is very crucial in SIS since it directly affects the efficiency of the proposed SIS method. In our approach, g_{t+1} is chosen as

$$g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t) = \pi_t(\mathbf{x}_{t+1}|\mathbf{x}_t)$$

due to the convenience it provides during the computation and satisfactory performance from the associated SIS method. It can be shown that in this case $u_{t+1} \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}, \mathcal{Y}_t)$.

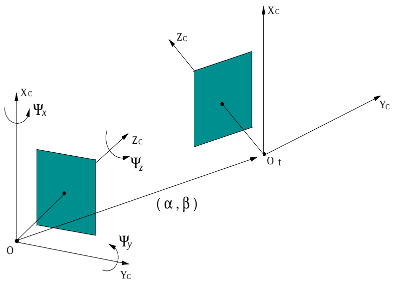


Fig. 1. Camera coordinate system and motion model

3. BAYESIAN MOVING TARGETS DETECTION

In our approach to moving targets detection, we first detect a set of feature points in the first frame. Typically, due to the intensity differences between moving targets and the background, some of the feature points will be located on the moving targets. The detected feature points then are tracked through the video sequence. By using the SIS technique, the posterior distribution of the camera motion can be approximated by a set of properly weighted motion samples and the feature points on the moving targets can be segmented out simultaneously.

3.1. A State Space Model

Parameterization Two 3D Euclidean coordinate systems used in this paper are shown in Fig. 1. System O and O_t, respectively denote the camera-centered coordinate frames before camera motion and at time t when the camera moves. Assuming that the camera internal calibration parameters are known, five parameters are enough to describe the camera motion.

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

$\psi = (\psi_x, \psi_y, \psi_z)$ are the rotation angles of the camera about the coordinate axes of the initial frame O. (α, β) are the elevation and azimuth angles of the camera translation direction, also measured in O. The *validity vector* ν_t describes the segmentation of the feature points. If M feature points are tracked through the sequence, ν_t is an M-dimensional vector. Each entry of ν_t is associated with a feature point. It indicates the possibility that this point belongs to the background.

State space model Given the above parameterization, the segmentation of the moving objects and the motion of the camera can be well described using a state space model with motion parameters $\mathbf{m}_t$ and the validity vector ν_t as its state vector, i.e.

$$\mathbf{x}_t = (\mathbf{m}_t, \nu_t)$$

and the perspective projection of the feature points on the image plane as the observations. The corresponding state space model can be expressed as follows.

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{x}_t + n_x && (1) \\ \mathbf{y}_t &= Proj(\mathbf{x}_t, \mathcal{S}_t) + n_y && (2) \end{aligned}$$

where n_x is the dynamic disturbance of the system, describing the time varying property of the state vector. $\mathbf{y}_t$ are the image positions of the features at time t $Proj(\cdot)$ denotes the perspective projection. It is a function of camera motion $\mathbf{m}_t$, validity vector ν_t and scene structure $\mathcal{S}_t$. In the following section, it can be shown that the likelihood function $f(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})$ can be obtained without knowing $\mathcal{S}_t$. Hence in the moving target detection procedure, $\mathcal{S}_t$ is not required to be explicitly computed.

3.2. SIS Formulation

Trial functions Based on the above state space model, we would like to design an SIS method for finding an approximation to the posterior distribution of the state parameters, $\pi_t(\mathbf{x}_t) = P(\mathbf{x}_t|\mathbf{y}_t)$. As we mentioned above, the trial distribution in the SIS procedure used in our approach is chosen as

$$g_{t+1}(\mathbf{x}_{t+1}|\mathbf{x}_t) = \pi_t(\mathbf{x}_{t+1}|\mathbf{x}_t)$$

If no prior knowledge about motion is available, a random walk will be a suitable alternative for modeling the dynamic motion of the camera. Hence the trial distribution for the motion parameters $\mathbf{m}_t$ is simply the one step Markovian state transition distribution $q_{t+1}(\mathbf{m}_{t+1}|\mathbf{m}_t)$. Therefore, during the SIS step (A), we draw samples from the distribution of $\mathbf{m}_t + n_m$. For the validity vector, the samples at $t+1$ are drawn via

$$\nu_{t+1} = \gamma\nu_t + \xi(\mathbf{m}_t, \mathbf{y}_t) + n_\nu \quad (3)$$

where n_ν is the dynamic noise in the validity vector and γ is an exponentially forgetting factor. Both of them represent the possible time-varying property of the validity vector. n_ν can be approximated using a Gaussian random vector. $\xi(\cdot)$ is a function used to update the current validity vector.

$$\xi(\mathbf{m}_t, \mathbf{y}_t) = (\frac{e_{th}}{e+1})^2 - sign(\lfloor e/e_{th} \rfloor)\frac{e+1}{e_{th}} \quad (4)$$

where $e = e(\mathbf{m}_t, \mathbf{y}_t)$ is the distance between the observed feature points and their associated epipolar lines given the motion parameters $\mathbf{m}_t$. e_{th} is a pre-chosen threshold for the above distance. This sampling step essentially plays the role of temporal integration of validity vector. The "incremental weight" u_{t+1} in this case is proportional to the likelihood function of the observation given the motion parameters, i.e.

$$u_{t+1} \propto f(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \quad (5)$$

Likelihood function In this case, the likelihood function of the observation given the state parameter is obtained via

$$f(\mathbf{y}_t|\mathbf{x}_t) \propto I_{\{\sum sign(s(i))>7\}}(\nu_t^+)\sum_{i=1}^{M}\nu_t^+(i)\exp\{-\frac{\epsilon}{\sigma_u^2+\sigma_v^2}\} \quad (6)$$

where

$$\nu_t^+(i) = \begin{cases} \nu_t(i), & \nu_t(i) > 0 \\ 0, & otherwise \end{cases} \quad (7)$$

and ϵ is given by

$$\epsilon = \frac{\sum_{i=1}^{M} e_t(i)^2 \nu_t^+(i)}{\sum_{i=1}^{M} \nu_t^+(i)} \quad (8)$$

<u>SIS Procedure</u>

1. Initialization. Draw samples of the 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. Note the fact that camera does not move does not imply $\mathbf{m}_0 = 0$. 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 distribution 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 here as the image sources instead of sets of image frames in arbitrary orders, 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 σ_ι,δ_α and δ_β can be chosen as some positive numbers. Draw samples of the validity vector $\{\nu_t^{(j)}\}_{j=1}^m$ via (3).
3. Weight computation and re-sampling. Compute the weights of the samples,$\{w_t^{(j)}\}$, using the observed feature correspondence according to the likelihood equation (6). The resulting samples and their corresponding weights $(\mathbf{x}_t^{(j)}, w_t^{(j)})$ are properly weighted with respect to $\pi_t(\mathbf{x}_t)$. Re-sample the above samples.

Since the sample sequences are properly weighted by their weights with respect to the posterior distribution of the state parameters $\mathbf{x}_t$, either the minimum mean square estimate (M MSE) or the maximum posterior (MAP) estimate of $\mathbf{x}_t$ can be obtained by finding the sample mean or locating the modes of π_t, respectively. Once an estimate of the validity vector for the feature points is obtained, classification methods such as the K-means algorithm can be applied to split the features into a group of features on the background and a group of features belonging to the moving objects.

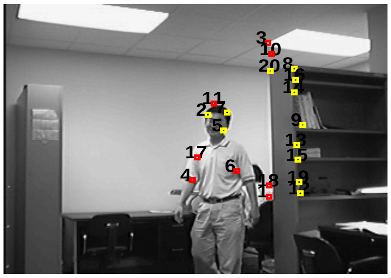
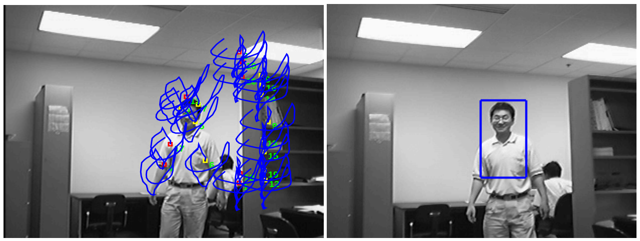


Fig. 2. Feature points locations in the first frame

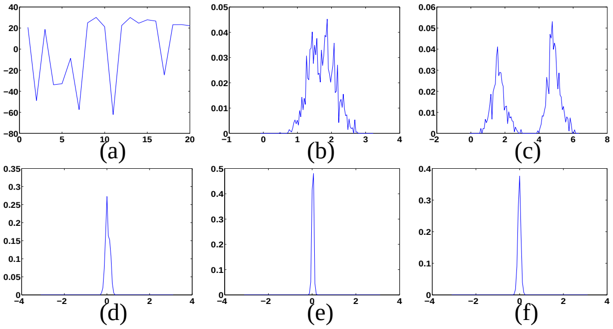


(a) (b)

Fig. 3. (a) is the feature trajectories and (b) shows the detected walking person in the testing sequence.

4. EXPERIMENTAL RESULTS

The proposed moving object detection algorithm has been tested using several video sequences and satisfactory results have been obtained. One of the experimental result is shown here. Fig. 2 shows the features detected in the first frame of a sequence containing a walking person. This sequence was captured by a moving camera. Fig. 3 (a) shows the feature trajectories in the sequence. It can be seen that some of the feature points are on the walking person and the rest of the features are on the background. By using the algorithm described here, the points on the walking person can be segmented out from the feature set and the camera motion respect to the background is estimated simultaneously. Fig. 4 (a) shows the sample mean of the validity vector at the last frame of the sequence and Fig. 4 (b)-(f) show the posterior distribution of the motion parameters. It can be seen that the feature points on the walking person have negative entries in the validity vector. The detection result of the walking person is shown in Fig. 3 (b).



(a) (b) (c)

(d) (e) (f)

Fig. 4. (a) is the mean of the validity vector and (b)-(f) are the a posterior distributions of the motion parameters at the last frame in the sequence. (b) and (c) are the distribution of the translation angles α and β, respectively. (d),(e) and (f) are the distribution of the rotational angles about X, Y and Z axes.

5. CONCLUSIONS

A recursive algorithm for moving targets detection from a moving platform using the SIS technique is presented in this paper. Both 2D and 3D scenes can be handled by the proposed method. Since the temporal relationship of the moving target segmentation between adjacent frames in the entire video sequence has been taken into account by the SIS procedure, our approach is very robust to observation noise. Although in this paper, we assume that the internal calibration parameters are known, a much weaker calibration will also give very similar results. When only the principal point (the intersection point of the optical axis of the camera with the image plane) is given and the field of view (FOV) of the camera is unknown, a similar SIS procedure can be developed to simultaneously recover the FOV of the camera, segment moving targets, and estimate camera motion. Future research will focus on the extraction of the boundary of the moving targets.

6. REFERENCES

[1] M. Irani, B. Rousso, and S. Peleg, “Computing occluding and transparent motions,” *International Journal of Computer Vision* **12**, pp. 5–16, February 1994.

[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] G. Young and R. Chellappa, “Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14**, pp. 995–1013, October 1992.

[4] M. Irani and P. Anandan, “A unified approach to moving object detection in 2d and 3d scenes,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**, pp. 577–589, June 1998.

[5] P. H. S. Torr, “Geometric motion segmentation and model selection,” in *Philosophical Transactions of the Royal Society A*, J. Lasenby, A. Zisserman, R. Cipolla, and H. Longuet-Higgins, eds., pp. 1321–1340, Roy Soc, 1998.

[6] R. Pless, T. Brodsky, and Y. Aloimonos, “Independent motion: The importance of history,” in *IEEE Computer Vision and Pattern Recognition, Fort Collins, CO,*, pp. II:92–97, 1999.

[7] J. S. Liu and R. Chen, “Sequential monte carlo methods for dynamic systems,” *J. Amer. Statist. Assoc.* **93**, pp. 1032–1044, 1998.

[8] J. MacCormick and A. Blake, “A probabilistic contour discriminant for object localisation,” in *International Conference on Computer Vision, Mumbai, India*, pp. 390–395, 1998.

[9] B. Li and R. Chellappa, “Simultaneous tracking and verification via sequential posterior estimation,” in *IEEE Computer Vision and Pattern Recognition, Hilton Head, SC*, pp. II:110–117, 2000.