2010 Fourth Pacific-Rim Symposium on Image and Video Technology

Motion Detection for Rapidly Moving Cameras in Fully 3D Scenes

Dong Zhang
Department of Control Science and Engineering
Zhejiang University
Hangzhou, China
zhangs@zju.edu.cn

Ping Li
Department of Control Science and Engineering
Zhejiang University
Hangzhou, China
pli@iipc.zju.edu.cn

***Abstract*—This paper presents a novel method for detecting motion regions in image sequences obtained by rapidly moving cameras in fully 3-D scenes. The proposed method mainly focuses on the situations that the backgrounds of the image sequences change rapidly. It has three innovations over existing methods: First, it presents a new initialization method to fast and sparsely gather information of the background model, while traditional methods utilize a complicated training step. Second, a novel model updating scheme is proposed for establishing the on-line sparse background model iteratively. This is the main contribution of the proposed method and this enables the method to work in a 3-D scene which totally changes through the image sequence, while most other methods can only work with a pre-modeled scene or with a camera that moving in limited scope. Third, a novel two-stage model of the background and foreground motion regions is proposed and the foreground motion regions are detected using maximum a posterior estimation. The method is tested on various challenging image sequences captured by freely moving cameras and results show that it is very effective and robust.**

***Keywords*-Motion detection; 3-D motion segmentation; Dynamical vision**

I. Introduction

This paper investigates a fundamental issue in computer vision: detecting motion regions in image sequences captured by moving cameras. It is the foundation of tracking, object recognition, pose reconstruction and action recognition. Traditional methods can work well with stationary cameras, and if there is a moving camera, the usual way is to use ego-motion compensation techniques. However, the scope of these methods is restricted to scenes that the backgrounds can be well approximated by planes or in which the cameras move with static camera centers. This severely limits its applications in natural scenes.

We propose a novel method for motion region detection with rapidly moving cameras, which can work well in rapidly changing 3-D scenes without the pre-construction of the full background model. It first utilizes some frames (less than three) to sparsely initialize the background model; then it uses a novel method to update the background model with the newly developed 3-D segmentation techniques; finally, it presents a new two-stage spatial-color model for detecting the foreground motion regions in the image sequences.

Due to the large volume of works of motion region detection, we need to draw the distinction between our work and the previous ones: Most of the previous methods focus on the stationary cameras, while the proposed method deals with freely moving cameras. The earliest method of stationary cameras utilizes the frame-differencing techniques [7], which is originated in the late 1970s. Methods emerged later mainly focused on the appearance model of the background, like per-pixel Gaussian model [1], Kalman Filters for updating pixel colors in [9], Gaussian mixture model in [15], non-parametric kernel density estimation in [2] and the joint spatial-color model in [14]. Research into relaxing the assumption of a stationary camera has mainly relied on the ego-motion compensation, such as ([6], [11]). However, these methods have severe limitations in that they can work only if the background can be well approximated by a plane or the camera moves while the camera center stays stationary. In the plane + parallax framework ([4], [21]), a homography is estimated between frames and a registration process follows to remove the effects of camera rotation, zoom and calibration. Such methods assume the presence of a dominant plane in the scene, making it possible to estimate the homography. Layer-based methods ([8], [16], [19]) model the scene as piece-wise planar scenes and cluster the pieces by some measure of motion coherency. However, all the methods above do not work well in a 3-D scene with a freely moving camera, which is the focus of the proposed method.

Several popular methods ([10], [22]) have a training step to learn the background model parameters. This training step must be repeated for each scene; however, training information may not always be available, especially when faced with a moving camera. As a counterpart, the proposed method employs a initialization stage to acquire the information of the background. It does not need many frames for the initialization, because in our framework, there is no need to establish the full background model. Moreover, the initialization process has no constraint on the camera motion.

Only a small amount of work could be found concerning the motion region detection problem in fully 3-D scenes with freely moving cameras. For instance, Sheikh's work [13] uses the assumption that the background is the spatially

978-0-7695-4285-0/10 $26.00 © 2010 IEEE
DOI 10.1109/PSIVT.2010.81

dominant "rigid" entity in the image with an orthographic camera model to segment the feature point trajectories and utilizes RANSAC [3] to extract the background trajectories. In summary, the proposed method has many innovations over the existing ones: First, it presents a new initialization method to fast and sparsely gather information of the background model, while most of the methods utilize a complicated training step. Second, a novel model updating scheme is proposed for establishing the on-line background model iteratively. This is the main contribution of the proposed method and this enables the method to work in a fully 3-D scene which totally changes through the image sequence, while most of the fore mentioned methods can only work with a pre-modeled scene or with a camera that moving in limited scope. Third, a novel two-stage model of the background and foreground motion regions is proposed and the foreground motion regions are detected using maximum a posterior estimation. This makes it possible to utilize sparse trajectories to build the background/foreground motion region models. This modeling method is necessary, because, in order to get an accurate 3-D motion segmentation result, it is impossible to track enough points with existing modeling methods (e.g. joint spatial-color model). This is a dilemma that more points tracked means much more erroneous points tracked and worse segmentation result.

The following sections are organized as followed: Section II, detailed discussion of the 3-D motion segmentation problem; Section III, the on-line model of the sparse background; Section IV, the two-stage model of the background/foreground motion regions; Section V, experimental results; Section VI, conclusions and future work.

II. 3-D Motion Segmentation

Over the past few years, several methods for 3-D motion segmentation have been proposed, which can be roughly separated into two categories [17]: affine methods ([18], [20]) assume an affine projection model and perspective methods ([5], [12]) assume a perspective projection model. The proposed motion detection method does not rely on the camera model. However, since the perspective methods are not mature enough, only the affine methods are considered [17] in this paper.

Let $\{x_{fp} \in \mathbb{R}^2\}_{p=1\ldots P}^{f=1\ldots F}$ be the projections of P 3-D points $\{X_p \in \mathbb{P}^3\}_{p=1}^{P}$ lying on a rigidly moving object[17] onto F frames of a moving camera. Under the affine projection model, the images satisfy the equation

$$x_{fp} = A_f X_p \tag{1}$$

,where $A_f = K_f \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_f & t_f \\ 0^T & 1 \end{bmatrix}$ is the affine camera matrix at frame f, $K_f \in \mathbb{R}^{2\times 4}$ is the camera calibration parameter and $(R_f, t_f) \in SE(3)$is the object pose parameters relative to the camera.

Let $W_1 \in \mathbb{R}^{2F\times P}$ be the matrix whose P columns are the image point trajectories $\{x_{fp}\}_{p=1}^{P}$. From 1 , W_1 can be decomposed into the motion matrix $M_1 \in \mathbb{R}^{2F\times 4}$ and the structure matrix $S_1 \in \mathbb{R}^{P\times 4}$:

$$W_1 = M_1 S_1^T \tag{2}$$

$$\begin{bmatrix} x_{11} & \cdots & x_{1P} \\ \vdots & \ddots & \vdots \\ x_{F1} & \cdots & x_{FP} \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_F \end{bmatrix}_{2F\times 4} \begin{bmatrix} X_1 & \cdots & X_P \end{bmatrix}_{4\times P} \tag{3}$$

.

Follow the introduction in [17], the 2D trajectories of a set of 3-D points seen by a rigidly moving camera (the columns of W_1) live in a subspace of $\mathbb{R}^{2F}$ with the dimension $d_1 = rank(W_1) =$ 2,3 or 4.

Consider about n rigid-body motions relative to a moving camera:

Let $W = \begin{bmatrix} W_1 & W_2 & \ldots & W_n \end{bmatrix} \Gamma \in \mathbb{R}^{2F\times P}$ be the data matrix where the columns of $W_i \in \mathbb{R}^{2F\times P_i}$ are the P_i trajectories associated with the ith moving object. $P = \sum_{i=1}^{n} P_i$ and $\Gamma \in \mathbb{R}^{P\times P}$ is an unknown permutation matrix. It is clear that W can be factorized like this:

$$\begin{aligned} W &= \begin{bmatrix} \hat{M}_1 & \hat{M}_2 & \ldots & \hat{M}_n \end{bmatrix} \begin{bmatrix} \hat{S}_1 & & & \\ & \hat{S}_2 & & \\ & & \ddots & \\ & & & \hat{S}_n \end{bmatrix} \Gamma \\ &= M S^T \Gamma \end{aligned} \tag{4}$$

It follows that one possible way of solving the motion segmentation problem is to find a permutation matrix Γ and use decomposition method to factorize W like above. This is the basis of most existing 3-D motion segmentation methods. In this paper, the proposed method utilizes GPCA[18] to obtain the segmentation result W_1 to W_n.

III. On-Line Sparse Background Model Construction

The proposed method includes two steps to construct the sparse background model as following.

A. Model Initialization

The initialization, as a counterpart of the training step of other methods, is performed at the beginning of the image sequence. However, it is different from the latter in that the purpose of it is to extract some background feature points rather than to establish the full background model.

It is difficult, also somewhat unreasonable, to establish the full background model in practice. The fore mentioned methods can work well by pre-modeling the background, only because the scenes are limited to certain form. However, if there is a camera moving freely in a 3-D scene, it is very time-consuming or even impossible to construct the full background model. For example, if the camera moves along a straight line without a termination, the background model could not be trained because the scope of the background is infinite.

The proposed method needs some frames at the beginning of the image sequence to initialize the sparse background model. The number of frames it needed depends on the motion pattern of the camera and typically it is 3. It is better to initialize by frames that contain no other motion regions except the background. However, if the frames for initialization include some foreground motion regions, the method just uses the "dominant rigid object" assumption [13] to select the background points.

If there are F^0 frames provided for initialization, the 3-D motion segmentation techniques could be utilized to segment the feature point trajectories $W^0 \in \mathbb{R}^{2F^0 \times P^0}$ into n^0 different motions(n^0 is usually set empirically in that most 3-D motion segmentation methods could not determine n^0 automatically), where P^0 is the number of trajectories extracted. That is $W^0 = \left[\begin{array}{cccc} W_1^0 & W_2^0 & \cdots & W_{n^0}^0 \end{array} \right] \Gamma$. In this,

$$\begin{aligned} W_i^0 &= \left[\begin{array}{ccc} x_{i,(11)}^0 & \cdots & x_{i,(1P_i^0)}^0 \\ \vdots & \ddots & \vdots \\ x_{i,(F_i^0 1)}^0 & \cdots & x_{i,(F_i^0 P_i^0)}^0 \end{array} \right] \\ &= \left[\begin{array}{ccc} x_{i,(*1)}^0 & \cdots & x_{i,(*P_i^0)}^0 \end{array} \right] \\ &= \left[\begin{array}{ccc} x_{i,(1*)}^0 & \cdots & x_{i,(F_i^0 *)}^0 \end{array} \right]^T \end{aligned} \tag{5}$$

,where P_i^0 is the number of the trajectories of the ith motion, $x_{i,(*j)}^0$ is the jth trajectory and $x_{i,k*}^0$ is the feature points of the kth frame of the ith motion. If $n^0 > 1$, the "dominant trajectories" are considered to be trajectories from background; or else, all are selected as the background trajectories. It is recommended to initialize with frames without motion regions other than the background.

By the end of this step, the background trajectories W_{back}^0 are selected to be the sparse background model.

B. On-Line Model Update

For model updating, F^1 frames are selected to get a new model. There are $F_{overlap}^1$ overlapping frames between these F^1 frames and the F^0 initializing frames. Also, there are F_{fresh}^1 "fresh" frames added and $F^1 = F_{overlap}^1 + F_{fresh}^1$. The goal of the model updating step is to select the background trajectories from these F^1 frames.

The feature point trajectories of these F^1 frames can be segmented using the 3-D motion segmentation techniques into n^1 different motions. So $W^1 = \left[\begin{array}{cccc} W_1^1 & W_2^1 & \cdots & W_{n^1}^1 \end{array} \right] \Gamma$.

The background model of these F^1 frames can be updated with the sparse model of the overlapping frames. That is, for kth of the overlapping frames in these $F_{overlap}^1$ frames, the feature point set $X_{back,(k*)}^1 = \left\{ x_{back,(kj)}^1 \right\} \Big|_{j=1}^{F_{back}^1}$ must has a nonempty intersection with $X_{back,((F^0 - F_{overlap}^1 + k)*)}^0$ of the corresponding frame in previous N^0 frames, because all background points do have the same motion. In short it is $X_{back,(k*)}^1 \bigcap X_{back,((F^0 - F_{overlap}^1 + k)*)}^0 \neq \emptyset$.

The proposed method examines from W_1^1 to $W_{n^1}^1$ to see which trajectory set fits the requirement above and obtain the updated model W_{back}^1.

With W_{back}^1, we can update the background model W_{back}^2 for another F^2 frames in the same way. This process can be performed iteratively for an image sequence of any length and obtain the sparse background model for any successive frames and other point trajectories are considered to be the foreground trajectories. Figure.1 shows the updating process of the sparse background models of an image sequence.

IV. Two-Stage Dense Model for Motion Detection

It is the joint spatial-color model that used in [13], and it works well in many applications, but it is not suitable for this application. The method in [13] is like this: the joint color-location vector is used to build the background/foreground appearance model and then kernel density estimation is employed to label the pixels as background or foreground. Because the color-location vector used is a 5-dimensional vector, there must be enough feature points tracked to build an effective model or else vectors located very sparsely in the 5-dimensional space could only induce a bad model of the background/foreground. In practice, it is difficult to track enough points effectively with the 3-D motion segmentation methods. Since the number of point trajectories extracted highly depends on the feature point matching mechanism, if the threshold of matching algorithm is increased, it will cause serious error in the 3-D motion segmentation results. By analyzing the process of human perception, we got the intuition that human usually first segment the motions with limited spatial information, then use color information to discern the boundaries between motions. Therefore, a new two-stage model is proposed for the background/foreground motion regions.

The objective of the method is to produce a binary labeling $L = \left[\begin{array}{cccc} l_1 & l_2 & \ldots & l_{N_{size}} \end{array} \right]$ for an image with N_{size} pixels[13]. We wish to estimate

$$L^* = \underset{L}{argmax}\, p(L|q) \tag{6}$$

, where q is the point set of the frames.

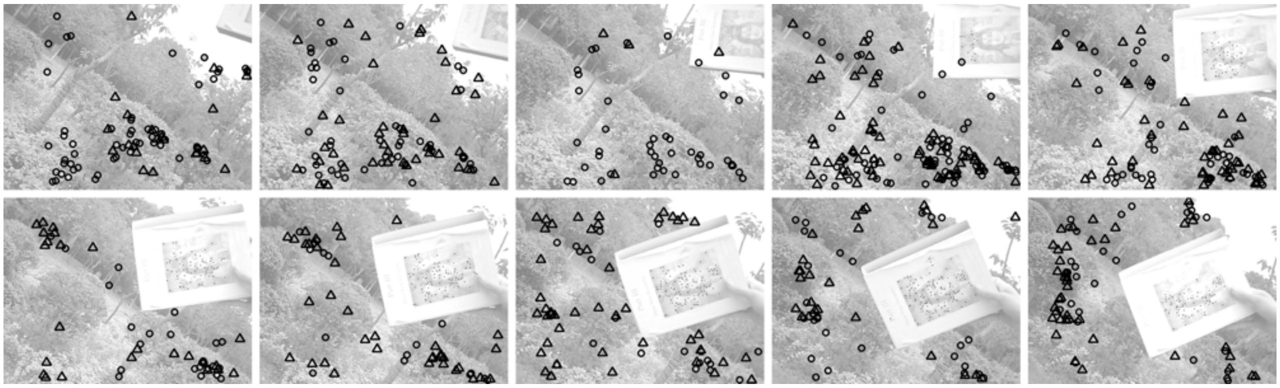

Figure 1. Sparse Background Model Updating Process. The sparse background model of an image sequence (Frame 4 to frame 13, with frame 1 to 3 for model initialization). Markers of triangles are "fresh" background feature points and markers of circles are overlapping background points. For each updating step, 1 fresh frame is added and 3 frames for a group to extract the point trajectories for 3-D motion segmentation.

Applying Bayes Theorem and assuming conditional independence, we can get

$$p(L|q) \propto p(L) \prod_{i=1}^{N_{size}} p(q_i|l_i) \tag{7}$$

.

The likelihood $p(q|L)$ is estimated as follow

$$p(q_i|l_i) = p(q_i|\theta_b)^{l_i - 1} p(q_i|\theta_f)^{l_i} \tag{8}$$

, where $p(q_i|\theta_b)$ is the probability of the pixel belonging to the background and $p(q_i|\theta_f)$ is the probability of the pixel belonging to the foreground.

A. Spatial Models of the Background/Foreground Motion Regions

Background and foreground trajectories are used to create spatial models. For a frame, the background trajectory points at the frame are used to construct a background model $\varphi_b = \begin{bmatrix} y_1^S & y_2^S & \dots & y_{Y^S}^S \end{bmatrix}$, where y_i^S is a location vector(i.e. $y_i = \begin{bmatrix} y_i^S(x) & y_i^S(y) \end{bmatrix}$, it is the location of the pixel in the image) and Y^S is the size of the model.

Follow the introductions in [13] and with the background model above, the log-posterior is

$$\log p(L|q) = (\sum_{i=1}^{N_{size}} (\sum_{j=1}^{N_{size}} (\lambda(l_i l_j + (1 - l_i)(1 - l_j))))) + \sum_{i=1}^{N_{size}} (\log(\frac{p(q_i|\varphi_f)}{p(q_i|\varphi_b)})) l_i \tag{9}$$

, where the first term is a pairwise Markov Random Field as the prior, λ is the smoothness parameter of the prior, and $p(q_i|\varphi_f)$ and $p(q_i|\varphi_b)$ are the kernel density estimations of the probability of a pixel belonging to the foreground and background. In practice, we do not use Markov Random Field for efficiency; however, results could be improved if it is utilized.

B. Color Models of the Background/Foreground Motion Regions

Since the background/foreground feature points extracted are sparsely located in the image, the motion region detection result L^* is not quite good. Therefore, an iterative process is presented to get the color models to refine the results.

Since the coarse motion region detection results are obtained, this can be used to establish the color model. Let $Q^{back} = \left\{ q_1^{back} \quad q_2^{back} \quad \dots \quad q_{N_{back}}^{back} \right\}$ be the pixels that labeled as background and $Q^{fore} = \left\{ q_1^{fore} \quad q_2^{fore} \quad \dots \quad q_{N_{fore}}^{fore} \right\}$ be the pixels that labeled as foreground by L^*, where N_{back} and N_{fore} are the background and foreground point number. We create the background color model as $\omega_b = \begin{bmatrix} y_1^C & y_2^C & \dots & y_{Y^C}^C \end{bmatrix}$, where each y_i^C is a color vector(i.e. $y_i^C = \begin{bmatrix} r_i & g_i & b_i \end{bmatrix}$, $\begin{bmatrix} r & g & b \end{bmatrix}$ defines a color in RGB space) and Y^C is the size of the model.

Finally, we maximize the log-posterior with the prior that only the pixels near the boundaries between background and foreground could be modified. That's because L^* is considered as an initial estimate of color model estimation, so most of the labels in L^* are right. Only the pixels that lie near the boundary of the background and foreground tend to be mislabeled. Thus, the log-posterior is

$$\log p(L|q) = \sum_{i \notin I_{bd}} \frac{\beta}{N_{size}} \cdot (1 - |l_i - l_i^*|) + \sum_{i=1}^{N_{size}} (\log(\frac{p(q_i|\omega_f)}{p(q_i|\omega_b)})) l_i \tag{10}$$

	Planar Object	Bus	3-D Object
Precision	0.905	0.768	0.825
Recall	0.823	0.959	0.857

Table I
THE PRECISION AND RECALL OF SOME TESTING IMAGE SEQUENCES

, where β is the weight coefficient, I_{bd} is the index set of the near boundary points in L^* and $p(x_i|\omega_f)$ and $p(q_i|\omega_b)$ are the kernel density estimations of the probability of a pixel belonging to the foreground and background. In 10, $1-|l_i-l_i^*|$ indicates the difference between l_i and l_i^*. That is, if $l_i=l_i^*$, $1-|l_i-l_i^*|$ is 1, else it is 0.

By maximizing this log-posterior, a new labeling result $L^{*(1)}$ could be obtained. $L^{*(1)}$ is used to establish a new color model and estimate $L^{*(2)}$, and so on. By using some stop criteria, the final result $L^{*(\infty)}$ could be obtained in finite loops (e.g. in the experiment, if there are no more than 30 pixels are modified within one step, the iterative process stops. We could always get $L^{*(\infty)}$ within 5 loops).

V. EXPERIMENTAL RESULTS

The proposed method was tested on a variety of image sequences with freely moving cameras. Results show that the method is effective and very robust. Table V shows the precisions and recalls of three of the tested image sequences and Figure 2,3 and Figure 4 show two experimental results of the tested sequences.

In the first experiment (in Figure 2), sample frames of an image sequence captured by a freely moving camera was tested. There is a big planar object (the box holding by hand) moving in the scene and the 3-D background changes rapidly (it can be seen from the first row of the figure, that the background almost totally changed from the first to the last frame). The first row of the figure is the raw images with tracked feature points superimposed (triangle/circle are the fresh/overlapping background points and '*' are the foreground points). The second row is the ground truth and the third row is the motion detection result of the proposed method. Precision and recall are in Table V.

Figure 3 shows the experimental results of an image sequence with a bus moving in it and Figure 4 shows an image sequence with a 3-D object moving in it. The definitions and algorithms are all the same with Figure 2.

VI. CONCLUSION AND FUTURE WORK

The proposed method is very effective for motion region detection in fully 3-D scenes with rapidly moving cameras. Both the initialization method and updating scheme of the sparse background model perform well in the experiments and the final results are accurate by using the two-stage dense model.

New dense modeling method of the background/foreground (e.g. effective color models other than RGB model) and temporal analysis techniques can be utilized to improve the results. Also, the proposed method can be integrated into the frameworks of robot navigation and object recognition to extend the applications of the latter.

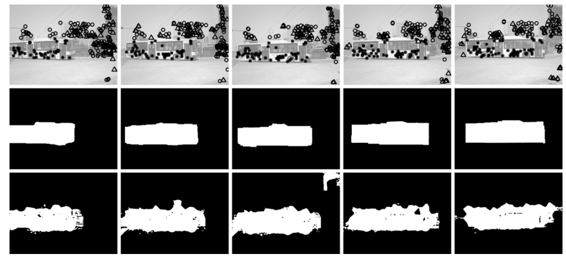

Figure 3. Motion Region Detection for an Image Sequence with a Bus Moving in 3-D a Scene. The sample frame indexes are 4, 5, 6, 7 and 8, while frame 1-3 are initialization frames. The definitions and algorithms are all the same with Figure 2.

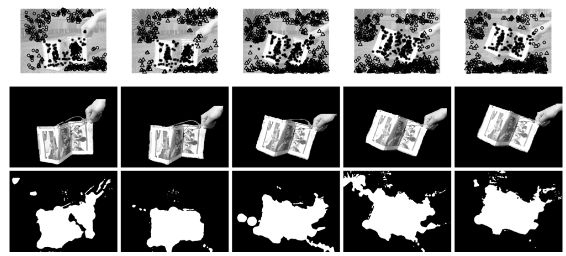

Figure 4. Motion Region Detection for an Image Sequence with a 3-D Object Moving in a 3-D Scene. The frame indexes are 4, 5, 6, 7 and 8, while frame 1-3 are initialization frames. The definitions and algorithms are all the same with Figure 2.

REFERENCES

[1] A. Azarbayejani, C. Wren, and A. Pentland. Real-time 3-d tracking of the human body. *Proceedings of IMAGE'COM*, 96, 1996.

[2] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. volume 90, pages 1151–1163, 2002.

[3] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] H. Guo, Y. Guo, and R. Kumar. Independent motion detection in 3d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 2000. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on.

[5] R. Hartley and R. Vidal. The multibody trifocal tensor: Motion segmentation from 3 perspective views. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 769–775, 2004. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

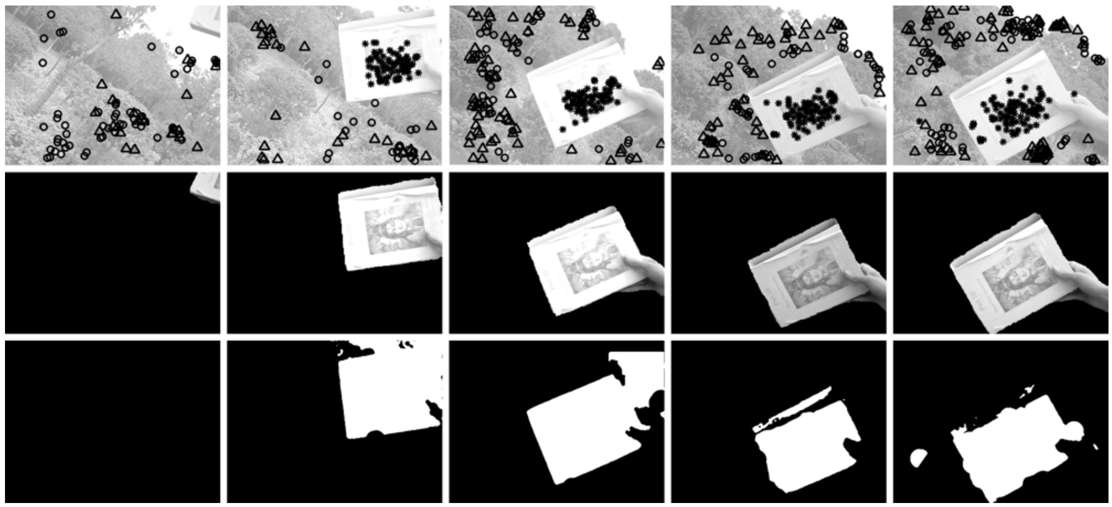

Figure 2. Motion Region Detection for a Moving Planar Object with Fast Changing 3-D Background. The sample frame indexes are 4, 9, 14, 19 and 24, while frame 1-3 are the initialization frames which contain only the background. There are 3 images in a frame group and in each updating step, 1 fresh frame is added and 2 overlapping frames are utilized. The first row of the figure is the raw images with tracked feature points superimposed (triangle/circle are the fresh/overlapping background points and '*' are the foreground points). The second row is the ground truth and the third row is the final result of the proposed method.

[6] E. Hayman and J. Eklundh. Statistical background subtraction for a mobile observer. In *IEEE International Conference on Computer Vision*, volume 1, pages 67–74, 2003. Proceedings of the Ninth IEEE International Conference on Computer Vision.

[7] R. Jain and H. Nagel. On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):206–213, 1979.

[8] Q. Ke and T. Kanade. A subspace approach to layer extraction. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2001. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[9] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell. *Towards Robust Automatic Traffic Scene Analysis in Real-Time*. Jerusalem, Israel, 1994.

[10] A. Monnet, A. Mittal, N. Paragios, V. Ramesh, R. Vision, S. Modeling, and N. Princeton. Background modeling and subtraction of dynamic scenes. pages 1305–1312, 2003. Ninth IEEE International Conference on Computer Vision, 2003. Proceedings.

[11] Y. Ren, C. Chua, and Y. Ho. Statistical background modeling for non-stationary camera. *Pattern Recognition Letters*, 24(1-3):183–196, 2003.

[12] K. Schindler, U. James, and H. Wang. Perspective n-view multibody structure-and-motion through model selection. *Lecture Notes in Computer Science*, 3951:606–619, 2006.

[13] Y. Sheikh, O. Javed, and T. Kanade. Background subtraction for freely moving cameras. 2009.

[14] Y. Sheikh and M. Shah. Bayesian modeling of dynamic scenes for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11):1778–1792, 2005.

[15] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[16] H. Tao, H. Sawhney, and R. Kumar. Object tracking with bayesian estimation of dynamic layer representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 75–89, 2002.

[17] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 4, pages 1–8, 2007. IEEE conference on computer vision and pattern recognition.

[18] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.

[19] J. Xiao and M. Shah. Motion layer extraction in the presence of occlusion using graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1644–1659, 2005.

[20] J. Yan and M. Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. *LECTURE NOTES IN COMPUTER SCIENCE*, 3954:94–106, 2006.

[21] C. Yuan, G. Medioni, J. Kang, and I. Cohen. Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1627–1641, 2007.

[22] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. *Pattern Recognition*, 2:28–31, 2004.