Active Object Tracking using Image Mosaic Background

Dong Kyu Kim[1], Young Kee Jung[1] and Yo-Sung Ho[2]

[1]Department of Computer Engineering, Honam University
59-1 Seobong-dong Kwangsan-gu, Kwangju, 560-090, Korea
Tel. +82-62-940-5404, Fax. +82-62-940-5667
E-mail : kitoq@hotmail.com, ykjung@honam.ac.kr

[2]Kwangju Institute of Science and Technology (K-JIST)
1 Oryong-dong Puk-gu, Kwangju, 500-712, Korea
Tel. +82-62-970-2211, Fax. +82-62-970-2204
E-mail : hoyo@kjist.ac.kr

Abstract: In this paper, we propose a panorama-based object tracking scheme for wide-view surveillance systems that can detect and track moving objects with a pan-tilt camera. A dynamic mosaic of the background is progressively integrated in a single image using the camera motion informatin. For the camera motion estimation, we calculate affine motion parameters for each frame sequentially with respect to its previous frame. The camera motion is robustly estimatd on the background by discriminating between background and foreground regions. The modified block-based motion estimation is used to separate the background region. Each moving object is segmented by image subtraction from the mosaic background. The proposed tracking system has demonstrated good performance for several test video sequences.

1. Introduction

For the automated surveillance system, we use a camera to watch moving objects in the restricted area. If objects move outside the field of view, the camera should pan or tilt such that they always stay within its filed of view. In those applications, motion detection and tracking for moving objects play quite important roles.

Although various research works have addressed these application areas [1-5], it is difficult to design general and robust solutions to the problems involved. This difficulty mainly stems from the complicated relationship between the motion of objects in the 3-D scene and the apparent motion of brightness patterns in the sequence of 2-D projections of the scene. Information about the relative depth of objects is lost in the projection, and the observed motion in the image plane can result from other phenomena than the object motion in the scene, such as changes in the lighting conditions. Moreover, the presence of observation in the 2-D image sequence is in itself a non-trivial task because of the presence of observation noise, occlusions and temporal aliasing. Especially, for the active camera, because the moving camera creates image changes due to its own motion, object tracking with the mobile camera is a very challenging task.

In this paper, we attempt to address these problems. First, we trace a moving object based on the image mosaic background for wide surveillance. Second, we utilze the affine model to generate the image mosaic background. The image mosaic is a panoramic image that is constructed from multiple frames in the video sequence [6,7]. The affine model provides greater flexibility in modeling the global motion, being able to represent rotation, dilation and shear as well as translation. Third, the camera motion is robustly estimated on the background by discriminating between background and foreground regions. Therefore, the camera motion estimate is not spoilt by the presence of outliers due to foreground objects whose motion is not representative of the camera motion.

2. Proposed Tracking Algorithm

As shown in Fig. 1, the proposed tracking system consists of five functional parts: foreground and background region separation, camera motion estimation, mosaic background, object detection & tracking, and control of the pan-tilt camera. The system first identifies background and foreground regions based on dominant motion estimates. Camera motion is then estimated on the background by applying the parametric affine motion estimation algorithm. The image mosaic background content is integrated in a single image. Finally, after we detect the moving object, we trace it at the center of the camera.

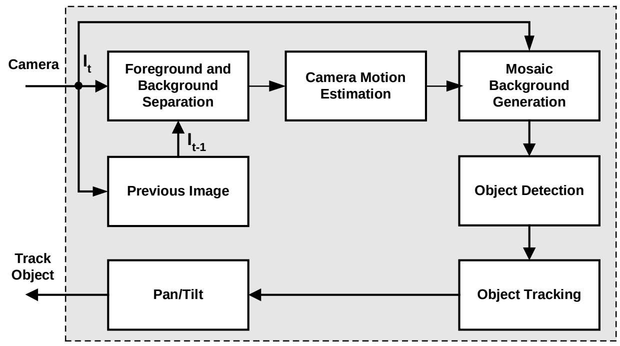


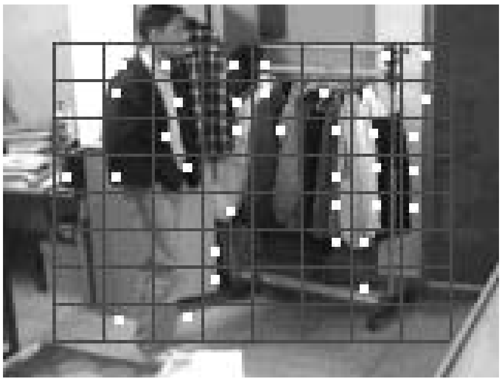
Fig. 1. Proposed Object Tracking Algorithm

2.1 Background and Foreground Separation

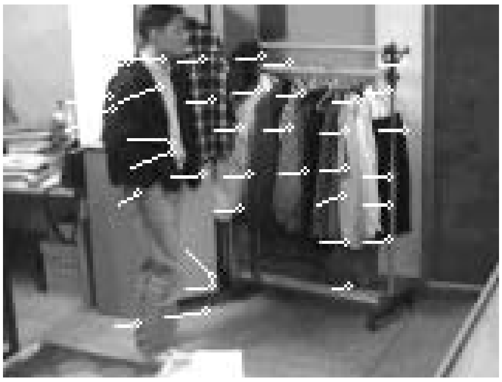
Discrimination between background and foreground is based on block-based motion estimation. A dominant motion is extracted by a clustering of the block vectors. Then, regions moving according to the dominant motion are identified as background, and otherwise as foreground. This separation has the following two steps: block-based motion estimation and background region extraction.

2.1.1 Block-based Motion Estimation

In this paper, the modified block-based estimator is used to track changes of individual blocks while the global motion estimation step is introduced for deriving a single representative affine motion. Each frame of 320x240 resolution is divided into multiple 32x24 blocks, and for block motion estimation, a 9x9 window region with maximum standard deviation is extracted within each block, as shown in Fig. 2.



(a)



(b)

Fig. 2. Block-based Motion Estimation (a) the selected 9x9 region for block motion estimation (b) the extracted block vectors

However, in low contrast areas, resulting motion vectors are unreliable. In order to overcome this problem, we apply the activity criterion to filter out unreliable blocks with lower standard deviation than a certain threshold value. The 9x9 template that was extracted is correlated in the search region. After we locate the correlation peak, a motion vector is associated with each block. The block motion vector holds the displacement of the block between the current and the previous frames.

2.1.2 Background Region Extraction

In order to extract the background motion, we compute a dominant by the following steps:

(a) For all block motion vectors, count the number of times that a motion vector is used.
(b) Obtain the most and second-most used motion vectors.
(c) Average the two candidates motion vectors.

Finally, if the motion of the block is similar to the dominant motion, we will consider this block as the background block; otherwise, blocks of the foreground block or noise block are removed.

2.2 Camera Motion Estimation

After discriminating the background motion from other motions, we estimate the camera motion in the background. In this way, the camera motion estimate cannot be spoilt by the presence of outliers due to foreground objects, whose motion is not representative of the camera motion.

The camera motion can be modeled by a parametric affine motion model with six parameters. We first estimate the six parameters using the least square method from the background motion vectors. Once motion parameters are obtained, we compensate the camera motion through the inverse affine motion transformation.

Let (x, y) be a block vector position in the previous frame, and (x', y') be the position in the current frame. Then, we can represent the motion vector (v_X, v_Y) by

$$\begin{pmatrix} v_X(x,y) \\ v_Y(x,y) \end{pmatrix} = \begin{pmatrix} x'-x \\ y'-y \end{pmatrix} \tag{1}$$

The affine motion model can be represented with the six parameters as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_4 & a_5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_3 \\ a_6 \end{pmatrix} \tag{2}$$

In order to estimate the six affine motion parameters, we define an error function to be minimized by

$$E(a) = \sum_{i=1}^{N} \{ [v_X(x_i, y_i) - \hat{v}_X(x_i, y_i)]^2 + [v_Y(x_i, y_i) - \hat{v}_Y(x_i, y_i)]^2 \} \tag{3}$$

where N is the number of motion vectors in the same frame. By substituting Eq. (2) into Eq. (3), we have

$$E(a) = \sum_{i=1}^{N} \left\{ [V_X(x_i, y_i) - (a_1 x + a_2 y + a_3)]^2 + [V_Y(x_i, y_i) - (a1x + a2y + a3)]^2 \right\} \tag{4}$$

The optimal values of the six parameters can be estimated by the least square method. The resulting equation is represented by

$$\sum_{i=1}^{N} \begin{bmatrix} x_i^2 & x_i y_i & x_i & 0 & 0 & 0 \\ x_i y_i & y_i^2 & y_i & 0 & 0 & 0 \\ x_i & y_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_i^2 & x_i y_i & x_i \\ 0 & 0 & 0 & x_i y_i & y_i^2 & y_i \\ 0 & 0 & 0 & x_i & y_i & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} v_X(x_i, y_i) \cdot x \\ v_X(x_i, y_i) \cdot y \\ v_X(x_i, y_i) \\ v_Y(x_i, y_i) \cdot x \\ v_Y(x_i, y_i) \cdot y \\ v_Y(x_i, y_i) \end{bmatrix} \tag{5}$$

2.3 Mosaic Background Generation

Once the affine parameters have been calculated, we can warp all the images with respect to the common coordinate system. To create the final mosaic image, we map the transformation parameters for each frame into the reference coordinate system by concatenating the transformation matrices. In this paper, we have arbitrarily selected the first image as the reference, and warped all other images into the first image's coordinate system. Using the camera motion information, a dynamic mosaic of the background is progressively integrated and stored in a single image.

2.4 Object Detection and Tracking

During the camera motion estimation process, the affine motion parameters for each frame are estimated sequentially with respect to its previous frame. From the camera motion information, we can extract the corresponding region of the background mosaic. The moving objects are then segmented by subtracting between the current frame and the corresponding background region.

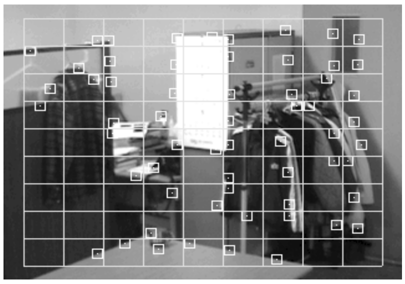
Once the object region is detected, we can track the object efficiently by predicting the next coordinate from the observed coordinate of the object centroid. We design a 2-D token-based tracking scheme using Kalman filtering [8]. The center position and the size of the object are used as the system states to be estimated. After we define the system model and the measurement model, we apply the recursive Kalman filtering algorithm in order to obtain linear minimum variance (LVM) estimates of motion parameters.

3. Experimental Result and Analysis

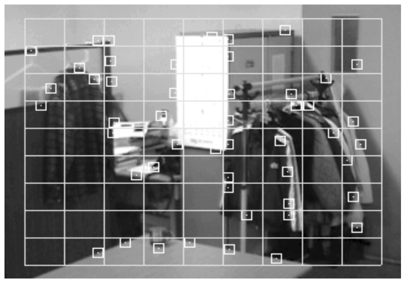
In this paper, we have tested the proposed tracking system on several video sequences of indoor environments. Fig. 3 shows the block feature selection results for three activity thresholds. A high activity threshold diminishes the number of the block features. In our experiment, we use 35.0 as the threshold value for the tracking system.

As shown in Fig. 4, four types of sequences are captured; right-panning and left-moving person, right-panning and right-moving person, left-panning and right-moving person, left-panning and left-moving person. In the case of Fig. 4(a), the right panning of camera causes one motion. A moving person occurs the other motion. The background motion is separated by dominant motion vector extraction. The center image of Fig. 4(a) displays the results of the block motion vector estimation algorithm. The result of the background motion separation is represented in the right image of Fig. 4(a).

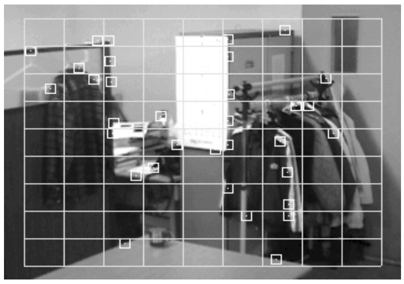
Fig. 5(a) shows the mosaic of a room from a sequence of 30 images. Fig. 5(b) and Fig. 5(c) show the current image with one person and the corresponding background from the mosaic, respectively. Fig. 5(e) is the subtraction image between the current image and the corresponding background image. This subtracted result has some noise blobs due to small motion errors. In order to remove the noise blobs, we utilize a morphological opening operation annd obtain the result in Fig. 5(f). The largest blob is chosen to the moving person in Fig. 5(g) and Fig. 5(d).



(a)



(b)



(c)

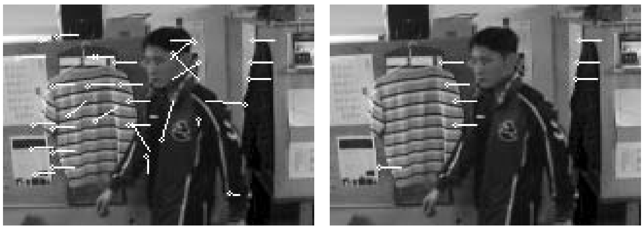
Fig. 3. Block feature Selection for 3 Activity Thresholds: (a) TH(25.0) (b) TH(35.0) (c) TH(45.0)

4. Conclusions

In this paper, we have proposed a new algorithm for moving object detection and tracking using the image mosaic technique. We have also presented an efficient algorithm for camera motion estimation based on the background motion to obtain image mosiac integration. In order to build the mosaic, we have aligned the frames with respect to the coordinate system and have updated. Subtracting between the current frame and the corresponding background region then segments the moving objects. As we have seen from the experimental results, the proposed algorithm builds the background mosaic and segments the foreground objects successfully.

Acknowledgement

This work was supported in part by grant NO. R05-2002-000-00868-0 from the Basic Research Program of the Korea Science & Engineering Foundation. This work was also supported in part by K-JIST, in part by KOSEF through UFON, and in part by MOE through BK21.



(a)

(b)

(c)

(d)

Fig. 4. Background Motion Separation: (a)right-panning and left-moving person, (b)right-panning and right-moving, (c)left-panning and right-moving person, (d)left-panning and left-moving person

(a)

(b) (c)

(d) (e)

(f) (g)

Fig. 5. Tracking Result: (a) mosaic of a room from a sequence (b) current image (c) corresponding background image (d) extracted object (e) subtract result (f) morphological opening (g) moving person extraction

References

[1] G. Johansson, "Visual Perception of Biological Motion and a Model for Its Analysis," *Perception and Psychophysics*, Vol. 14, pp. 201-211, 1973.

[2] K. Gould and M. Shah, "The Trajectory Primal Sketch", "A Multi-Scale Scheme for Representing Motion Character-istics," *IEEE Conf. of CVPR*, pp. 79-85, 1989.

[3] O. Rouke and Badler, "Model-based Image Analysis of Human Motion using Constraint Propagation," *IEEE Trans. on PAMI*, Vol. 3, No. 4, pp. 522-537, 1980.

[4] K.W. Lee, Y.H. Kim, J. Jeon and K.T. Park "An Algorithm of Moving Object Extraction Under Visual Tracking without Camera Calibration," *Proceedings of ICEIC*, pp. 151-154, 1995.

[5] A.J. Lipton, H. Fujiyoshi and R.S. Patil, "Moving target classification and tracking from real time video," *IEEE Workshop on Applications of Computer Vision*, pp. 8-14, 1998.

[6] R. Szeliski and H. Shum, "Creating full view panoramic image mosaics and environment maps," *Computer Graphics Proceedings*, pp. 251-258, 1997.

[7] M. Irani, P. Anandan, J. Bergen, R. Kumar and S. Hsu, "Mosaic Representation of video sequences and their applications Signal Processing," *Image Commnu., special issue on Image and Video Semantics: Processing, Analysis, and Application*, Vol. 8, no. 4, pp. 673-676, 1996.

[8] Y.K. Jung and Y.S. Ho, "Robust Vehicle Detection and Tracking for Traffic Surveillance," *Picture Coding Symposium*, pp. 227-230, 1999.