Automatic Video Object Tracking Using a Mosaic-Based Background

Young-Kee Jung[1], Kyu-Won Lee[2], Dong-Min Woo[3], and Yo-Sung Ho[4]

[1] Honam University, Gwangju, 506-090, Korea
ykjung@honam.ac.kr
[2] Daejeon University, Daejeon, 300-716, Korea
kwlee@dju.ac.kr
[3] Myongji University, Yongin, 449-728, Korea
dmwoo@mju.ac.kr
[4] Gwangju Institute of Science and Technology, Gwangju, 500-712, Korea
hoyo@gist.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 information. 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 estimated 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.

Keywords: Object Tracking, Object Detection, Image Mosaic.

1 Introduction

In recent years, there have been various research works on image segmentation for object-based coding using the MPEG-4 standard [1,2,3,4,5,6]. MPEG-4 supports a content-based representation of visual objects in the compressed bitstream domain. Extraction of moving objects plays a key role in such kind of applications. For video conferencing and surveillance, we use a camera system 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 of moving objects play quite important roles.

However, it is difficult to extract video objects and to design general and robust solutions to problems involved. Conventional object extraction and tracking schemes could not be applied to the video sequence taken from an active camera because the moving camera creates image changes due to its own motion. Although a few references [5,6] have addressed the problem of object segmentation

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 866–873, 2004.

and tracking using an active camera, they cannot segment the moving object in an arbitrary pan and tilt angle.

In this paper, we attempt to resolve these problems. First, we trace a moving object based on the image mosaic background with an active camera. Second, we have focused on how to match images for the general transformations, i.e. for the cases when the camera pans, rotates, and tilts in any directions. An affine model is utilized to generate the image mosaic background. An image mosaic is a panoramic image reconstructed from multiple frames in the video sequence [7,8]. Affine models provide greater flexibility in modelling a 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 motions are not representatives 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 and tracking, and control of the pan-tilt camera. After background and foreground regions are identified based on dominant motion estimates, camera motion is then estimated on the background by applying parametric affine motion estimation. The image mosaic background is integrated in a single image. Finally, after we detect and trace the moving object using background, we command the pan-tilt controller to position the moving object at the center of the camera.

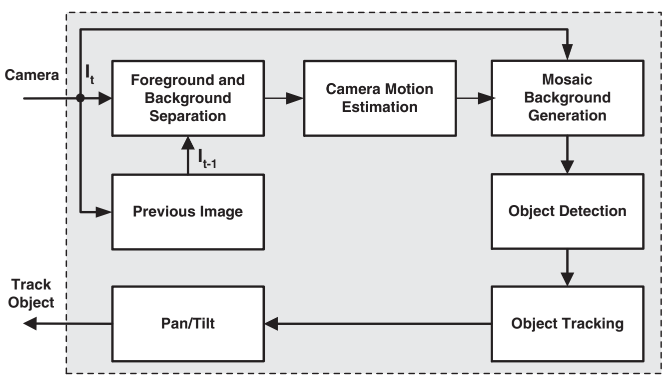


Fig. 1. Proposed Object Tracking Algorithm

2.1 Background and Foreground Separation

The discrimination between background and foreground is based on block-based motion estimation. After the dominant motion is extracted by clustering the block vectors, regions of 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.*

Block-based Motion Estimation. In this paper, the modified block-based estimator is used to track changes of the individual block while the global motion estimation step is introduced for deriving a single representative affine motion. Each frame of the 320x240 pixel resolution is divided into non-overlapping of 32x24 pixels. For the block motion estimation, a 9x9 window region with the 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 the lower standard deviation than a certain threshold value. The extracted 9x9 template 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.

Background Region Extraction. In order to extract the background motion, we compute a dominant motion by the following procedure:

(1) For all block motion vectors, count the number of times that a motion vector is used.

(2) Obtain the most and second-most popularly used motion vectors.
(3) Average the two motion vector candidates.

If the motion of the block is similar to the dominant motion, we regard this block as the background block. Finally, foreground or noise blocks are removed.

2.2 Camera Motion Estimation

After the background motion is discriminated from the other motions, the camera motion is estimated from the background. In this way, the camera motion estimate is not disturbed by the presence of foreground objects.

The camera motion is modelled by a parametric affine motion model of six parameters. Once we estimate the six parameters using the least square method from the background motion vectors, we compensate the camera motion through the inverse affine motion transformation.

Let (x, y) be a block 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

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

Since we use the affine motion model of six parameters, the motion vector can be expressed as

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

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

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

where N is the number of motion vectors in the same frame.

By substituting Eq. 2 into Eq. 3, we have

$$E(a) = \sum_{i=1}^{N} \{[v_X(x_i, y_i) - (a_1 x + a_2 y + a_3)]^2 + [v_Y(x_i, y_i) - (a_1 x + a_2 y + a_3)]^2\} \tag{4}$$

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

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

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. In order 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 chosen 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

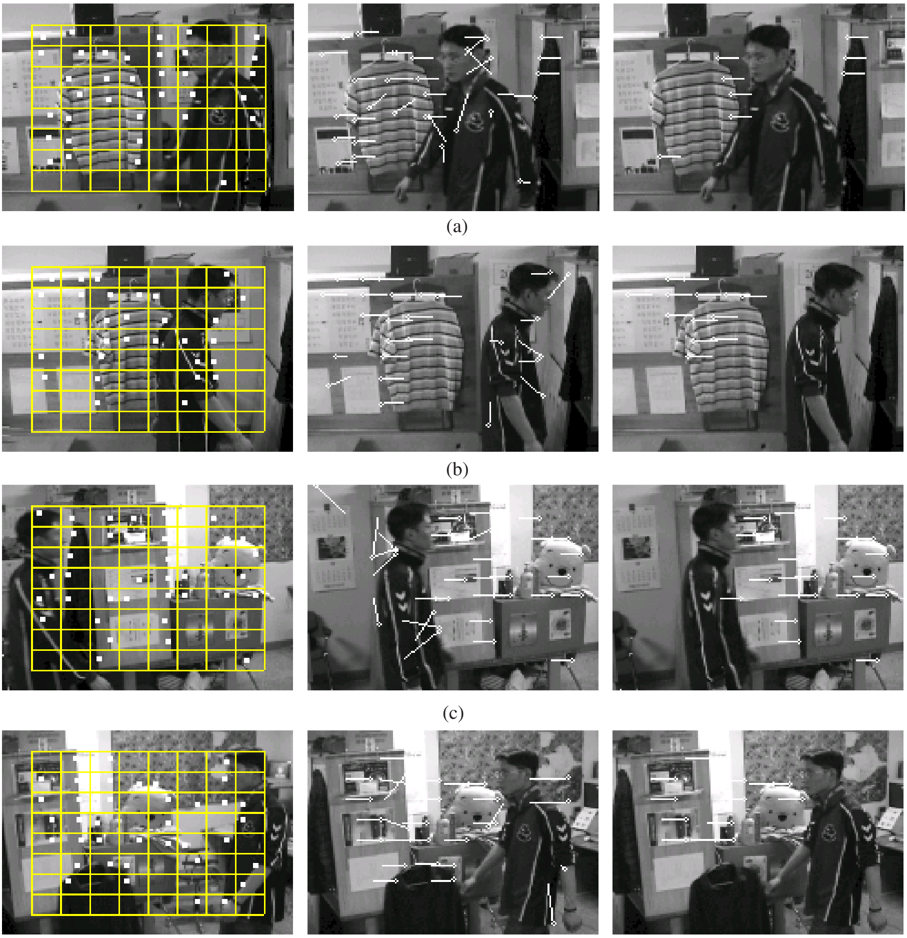
During the camera motion estimation process, the affine motion parameters for each frame are estimated sequentially with respect to its previous frame. From the camera motion information, we can extract the corresponding region of the background mosaic. The moving objects are then segmented by subtracting 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 2D token-based tracking scheme using Kalman filtering [9]. 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 to obtain linear minimum variance (LMV) estimates of motion parameters.

Fig. 3. Initial Image Mosaic Background

3 Experimental Results and Analysis

The proposed tracking system has been tested on several video sequences in indoor environments. Fig. 3 shows the initial mosaic background from 16 images. Four types of sequences are captured, as shown in Fig. 4; right-panning and left-moving person, right-panning and right-moving, left-panning and right-moving person, left-panning and left-moving person.



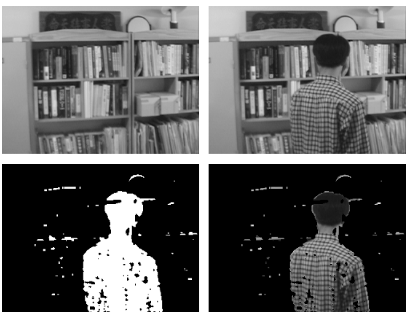
(a) (b) (c) (d)

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

In 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 result of block motion vector estimation. The result of background motion separation is represented in the most right image of Fig. 4(a).

Fig. 5(a) shows the mosaic of a room from a sequence of 30 images. Fig. 5(b) shows a current image with one person and the corresponding background from the mosaic. Fig. 5(b) also shows the subtraction image between the current image and the corresponding background image. This subtraction result has some noise blobs due to small errors of camera motion estimation. We utilize a morphological opening operation to remove the noise blobs and the largest blob is chosen to the moving person in Fig. 5(b). The moving person is traced and marked with a white rectangle in Fig. 5(c).

(a)



(b)

(c)

Fig. 5. Tracking Results: (a) mosaic background from a room sequence, (b) extracted object, (c) tracking object

4 Conclusions

In this paper, we propose a new algorithm for moving object tracking using the image mosaic background. We also propose an efficient camera motion estimation algorithm based on background motion to obtain image mosaic integration. In order to build the mosaic, the frames are aligned with respect to a coordinate system and updated. By subtracting the current frame from the corresponding background region, we segment the moving objects. Simulation results demonstrate that the proposed algorithm successfully builds the background mosaic and segments foreground objects.

Acknowledgement. This work was supported in part by grant No. R01-2002-000-00336-0 from the Basic Research Program of the Korea Science & Engineering Foundation, and in part by the Post-doctoral Fellowship Program of Korea Science & Engineering Foundation. This work was also supported in part by GIST, in part by the Ministry of Information and Communication (MIC) through the Realistic Broadcasting Research Center at GIST, and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

References

1. Gould, K., Shah, M.: The Trajectory Primal Sketch: A Multi-Scale Scheme for Representing Motion Characteristics. IEEE Conf. of CVPR, (1989) 79-85
2. Rouke, O., Badler: Model-based Image Analysis of Human Motion using Constraint Propagation. IEEE Trans. on PAMI, Vol.3, No.4 (1980) 522-537
3. Lee, K.W., Kim,Y.H., Jeon, J., and Park, K.T.: An Algorithm of Moving Object Extraction Under Visual Tracking without Camera Calibration. Proceedings of ICEIC, (1995) 151-154
4. Lipton, A.J., Fujiyoshi, H., and Patil, R.S.: Moving target classification and tracking from real time video. IEEE Workshop on Applications of Computer Vision, (1998) 8-14
5. Ye, Y., Tsotsos, J. K., Bennet, K., and Harley, E.: Tracking a person with prerecorded image database and a pan, tilt, and zoom camera. Proc. IEEE Workshop on Visual Surveillance, (1998) 10-17
6. Hat, S., Saptharishi, M., and Khosla, P.K.: Motion detection and segmentation using image mosaics. Proc. IEEE Int. Conf. Multimedia and Expo., (2000) 1577-1580
7. Szeliski, R., and Shum, H.: Creating full view panoramic image mosaics and environment maps. Computer Graphics Proceedings, (1997) 251-258
8. Irani, M., Anandan, P., Bergen, J., Kumar, R., and Hsu, S.: Mosaic Representation of video sequences and their applications. Signal Processing: Image Commnunication, special issue on Image and Video Semantics: Processing, Analysis, and Application, Vol. 8, No. 4, (1996) 673-676
9. Jung, Y.K., Ho, Y.S.: Robust Vehicle Detection and Tracking for Traffic Surveillance. Picture Coding Symposium, (1999) 227-230