Quasi-Feature based Panoramic Video Creation for Multiview Object Tracking System

Changhan Park and Kyung-Hoon Bae*
Advanced Technology Research & Development Center, Samsung Thaels. Co., Ltd.
*Corresponding author E-mail: khbae.bae@samsung.com

Abstract: *In this paper we present an efficient approach to building a panoramic video from mutiview cameras and to tracking objects. The proposed panoramic video creation module consists of two functions: (i) selecting four quasi-feature points in two adjacent frames acquired by corresponding multiple cameras and (ii) mosaicing the two images. Four pairs of selected quasi-feature points play a role of similarity reference in registering two adjacent frames. The mosaicing step uses the direct linear transformation (DLT) algorithm. The proposed tracking algorithm uses the active contour model (ASM), which is robust against partial occlusion. The proposed tracking module consists of four functions: (i) landmark point assignment, (ii) principal component analysis (PCA), (iii) modeling of local structure, and (iv) model fitting. While most conventional panoramic image creation methods are pixel-based, the proposed feature-based method provides more accurate tracking result. In the experiment, the created panoramic images exhibit high quality, which enables robust, real-time video tracking.*
Keywords: *Object tracking, panoramic video, Image mosaicing, multiview camera.*

1. Introduction

Image mosaicing techniques have attracted a growing attention in many application areas, such as: video stabilization and compression, background generation, virtual environment, and panoramic photography[1]. Recently, its application was extended to video tracking in the wide space such as airport and building. The problem of tracking and recognizing non-rigid objects in video sequences becomes crucial in many video surveillance applications. Examples include a motion detector in video recording systems, action analysis for animation, medical imaging, and human computer interaction (HCI).

Image mosaics can be constructed by aligning and blending partially overlapped images. Various methods to register overlap areas of two adjacent images have been proposed using an 8-parameters perspective transformation[2], a polynomial transformation with higher degree of freedom[3], or other geometric corrections[4]. However, the registration is usually implemented on the entire overlapping area, which is usually too large for a single global transformation to get an acceptable result. Thus local corrections have to be introduced to deal with both discontinuities and distortions. The problem of the overlap-based registration still exists for most commercial software to generate panoramic views. The existing research performs projective transformation from only four features in two overlapping images[5].

The proposed panoramic video creation module consists of: two steps (i) selection of four quasi-feature points in two adjacent frames acquired by corresponding 4 cameras and (ii) mosaicing the two images. The proposed quasi-feature extraction algorithm selects four points that are not co-linear in the reference image. In creating a panoramic image, overlapping ratio of two images is 30% to 90%. The reference frame is first divided into three areas and block-based registration is performed in the third block by using mean absolute difference (MAD) values. The block of size can be experimentally selected, and we used 20×10 blocks for the experiment. We additionally search a quasi-feature point from centroid of the selected block. In this process, an internal block consists of a flat and a texture region. The extracted four feature points play a role in evaluating parameters of the projective transform by using the DLT.

Tracking a deformable object in consecutive frames is a fundamental problem in video surveillance systems. There have been various researches for video-based object extraction and tracking. One of the simplest methods is to track regions of difference between a pair of consecutive frames[6], and its performance can be improved by using adaptive background generation and subtraction. Although a simple difference-based tracking method is efficient in tracking an object under noise-free circumstances, it often fails under noisy, complicated background. The tracking performance is further degraded if a camera moves either intentionally or unintentionally. For more robust analysis of an object, shape-based object tracking algorithms have been developed, which utilize a priori shape information of an

International Journal of Advanced Robotic Systems, Vol. 6, No. 2 (2009)
ISSN 1729-8806, pp. 109-114

object-of-interest, and project a trained shape onto the closest shape in a certain frame. This type of methods includes active contour model (ACM), condensation algorithm, and ASM.

This paper presents an ASM-based, real-time tracking algorithm for locating a deformable object in the panoramic images. The proposed tracking algorithm is considered to be a physical model that allows the system to accurately predict the potential change in object's shape over time. The detecting procedure extracts moving objects by motion segmentation between frames, and the tracking procedure detects moving regions using, for example, optical flow-based data association. The detected region is further enhanced using ellipsoid localization prior to update of training sets using the smart snake algorithm (SSA). The local structure modeling is carried out using directional regularization operators. The directional regularization reflects the orientation of the edges and finds dominant directional edge.

The paper is organized as follows. In section 2 we introduce a method to extract quasi-features and to build a panoramic image. The ASM is presented in section 3. Section 4 summarizes experimental results, and section 5 concludes the paper.

2. Panoramic Image Creation Using Quasi-Features

There are limits in tracking objects from as single image. Therefore, we propose a method that can track objects in a wide area of background. The proposed method extracts quasi-feature points, and performs the DLT algorithm to build panoramic image. The panoramic video based multiview object tracking follows shown in Fig. 1.

2.1. Quasi-Feature Points and Their Correspondence

The quasi-feature is not the geometrical feature of the image but the feature based on the intensities of the image[7]. The proposed method extracts four feature points in two overlapping images to compute the projective transformation. Feature extraction and block searching processes are illustrated in Fig. 2.

The extraction algorithm of quasi-feature is summarized as follow.

Step 1: Select a block R_b in the reference frame, and extracts the most similar block T_b in the target frame. This

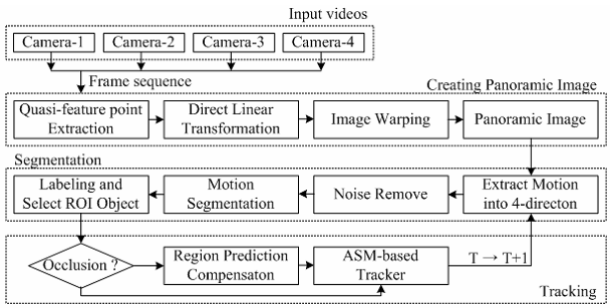


Fig. 1. Block diagram of the proposed ASM-based tracking algorithm in panoramic video

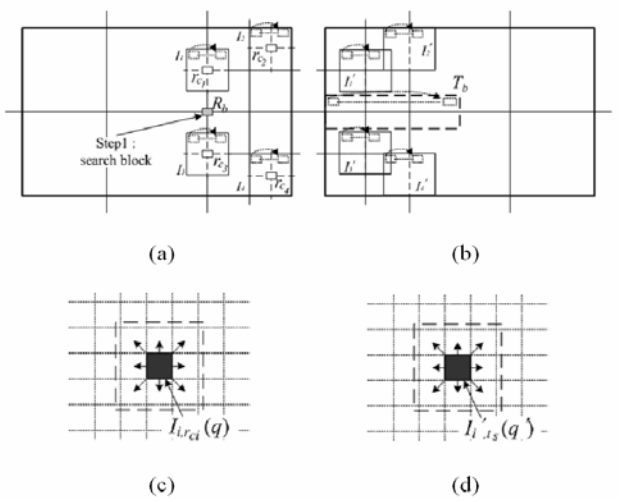


Fig. 2. Proposed block search and quasi-feature extraction: (a) reference frame, (b) target frame, (c) reference quasi-feature point, and (d) target quasi-feature point

process is searching the most similarity block from reference frame to target frame. Repeat the same operations for the limited four sub-blocks, I_1, I_2, I_3, and I_4.

Step 2: Extract strong edge using Canny operator. This process searches many edge blocks that are classified as a flat and textured regions. As a result the optimal quasi-features are extracted in the overlapping region.

Step 2-1: Evaluate r_{c_i} in the landmark region I_i. This detects the optimal region with the minimum MAD, where i represents the index of the block sequence.

$$MAD = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| I_i(m,n) - I_i^{'}(m+x, n+y) \right|, 1 \le i \le 4, \qquad (1)$$

where M and N respectively represent the width and height of r_{c_i}. $I_i(m,n)$ and $I_i(m,n)$ represent intensities of the reference and target frames, respectively.

Step 2-2: Select t_s by comparing r_{c_i} and I'_i, where t_s represents a block with the optimum quasi-feature point.

Step 2-3: Select the centroid of the optimum quasi-feature q in t_s.

Step 3: Compute quasi-feature error of 8-neighborhood pixels between r_{c_i} and t_s as shown in Fig. 2(c) and 2(d), defined as

$$\varepsilon = \sum_i \{ I_{i,r_{c_i}}(q) - I_{i,t_s}^{'}(q^{'}) \}^2 \qquad (2)$$

where $I_{i,r_{c_i}}(q)$ represents the mosaic intensity level at q, and $I'_{i,r_{c_i}}(q')$ the intensity of a pixel from the target frame projected to the same mosaic location q.

2.2. Panoramic Image Using DLT

The proposed panoramic image creation performs the projective transform by using the DLT algorithm from the extracted four quasi-features. The homography H matches in two images as[5]

$$x' = Hx, \tag{3}$$

where $x' = (x', y', z')^T$, and $Hx = (H^{1T}x, H^{2T}x, H^{3T}x)$. Eq. (4) form enables a simple homegeneous linear solution for H as

$$\begin{bmatrix} 0^T & -w'x^T & y'x^T \\ w'x^T & 0^T & -x'x^T \\ -y'x^T & x'x^T & 0^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0 \tag{4}$$

Although there are three equations in Eq. (4), only two of them are linearly independent. Thus each point correspondence gives two equations in the entries of H. It is usual to omit the third equation in solving for H. Consequently, the set of equations forms

$$\begin{bmatrix} 0^T & -w'x^T & y'x^T \\ w'x^T & 0^T & -x'x^T \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0 \tag{5}$$

This can be rewritten as

$$Ah = 0, \tag{6}$$

where H is a 2x9 matrix, and $h = (h^1, h^2, h^3)^T$. We then obtain singular value decomposition (SVD) of matrix A.

$$A = UDV^T. \tag{7}$$

The unit singular vector corresponding to the smallest singular value is the solution h. If H is diagonal with positive entries, arranged in descending order, h is the last column of V. The matrix H is determined from h as in Eq. (6).

3. Multiview Object Tracking System

Localization and shape of an object are the fundamental factors for video tracking. Within the class of deformable models, the ASM obtains boundary shape by prior information. Therefore the ASM is one of the best-suited approaches in the sense of both accuracy and efficiency. The basic concepts of ASM consist of modeling the contour of the silhouette of an object in the image by parameters in order to align the changing contours in the image frames to each other. The proposed tracking module consists of four steps: (i) landmark point assignment, (ii) PCA, (iii) modeling of local structure, and (iv) model fitting. In this section we present about fundamental ASM.

3.1. Landmark Point Assignment

Given a frame of input video, suitable landmark points should be assigned on the contour of the object. The feature point of object boundary called landmark points. The role of landmark points is controlling the shape of model contours. Good landmark points should be consistently located from one image to another. In a two-dimensional image, we represent n landmark points by a $2n$-dimensional vector as

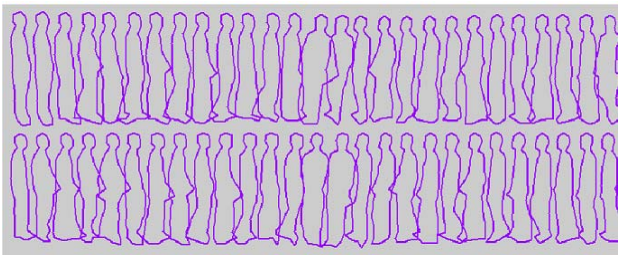


Fig. 3. Training set of 56 shapes (m=56)

$$x = [x_1, \cdots, x_n, y_1, \cdots, y_n]^T, \tag{8}$$

where x and y represent a coordinate of pixels, respectively.

A typical setup in our system consists of 42 manually assigned landmark points of 42 (n=42). The role of landmark points is controlling the shape of model contours. More specifically, the initially assigned landmark points are updated by minimizing the deviation from the original profile, which is normal to the boundary at each landmark point.

3.2. Training Set

A set of n landmark points represents the shape of the object. Fig. 3 shows a set of 56 different shapes, called a training set. The shape of object made with landmark points. We must aligning shapes into a common coordinate frame.

3.3. Principal Component Analysis (PCA)

Although each shape in the training set is in the $2n$-dimensional space, we can model the shape with a reduced number of parameters using the PCA technique. Suppose we have m shapes in the training set as shown in Fig. 3, presented by x_i, for i=1,…,m. The nonlinear version of this constraint is discussed in.

3.4. Model Fitting

We can find the best pose and shape parameters to match a shape in the model coordinate frame, x, to a new shape in the image coordinate frame, y by minimizing the error function.

$$E = (y - Mx)^T W (y - MX), \tag{9}$$

where W is a diagonal matrix whose elements are weighting factors for each landmark point and M represents the geometric transformation of rotation θ, translation t, and scaling s. The weighting factors are set in relation to the displacement between the computed positions of the old and the new landmark points along the profile. If the displacement is large, then the corresponding weighting factor in the matrix is set low; if the displacement is small, then the weighting is set high.

Given a single point, denoted by $[x_0, y_0]^T$, the geometric transformation is defined as

$$M\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = s\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (10)$$

After the set of pose parameters, $\{\theta, t, s\}$, is obtained, the projection of *y* into the model coordinate frame is given as

$$x_p = M^{-1}y. \quad (11)$$

Finally, the model parameters are updated as

$$b = \Phi^T(x_p - \bar{x}) \quad (12)$$

As the result of the searching procedure along profiles, the optimal displacement of a landmark point is obtained. The combination of optimally updated landmark points generates a new shape in the image coordinate frame, *y*. This new shape is now used to find the nearest shape using Eq. (9). After computing the best pose, denoted by *M*, this new shape is projected into Φ, which contains principal components of the given training set. This process updates the model parameter *b*. As a result, only similar variation corresponding to the principal components can affect the model parameters.

3.5. Modeling a Local Structure

A statistical, deformable shape model can be built by assignment of landmark points, PCA, and model fitting steps. In order to interpret a given shape in the input image based on the shape model, we must find the set of parameters that best match the model to the image. Show Fig. 4, if we assume that the shape model represents boundaries and strong edges of the object, a profile across each landmark point has an edge-like local structure. Let g_j, *j=1, … , n*, be the normalized derivative of a local profile of length *K* across the j^{th} landmark point, $\overline{g_j}$ and *S* the corresponding mean and covariance, respectively. The nearest profile can be obtained by minimizing the following Mahalanobis distance between the sample and the mean of the model as

$$f(g_{j,m}) = (g_{j,m} - \overline{g_j})S_j^{-1}(g_{j,m} - \overline{g_j})^T \quad (13)$$

where $g_{j,m}$ represents g_j shifted by m samples along the normal direction of the corresponding boundary. In practice, we use a hierarchical ASM technique because it provides a wider range for the nearest profile search.

4. Experimental Results

We used 4 CCD cameras 320×240, indoor video sequences to test tracking performance for deformable object. Illumination changes in outdoor video sequences are not considered. Fig. 5(a) shows a panoramic image created by the proposed method. Fig. 5(b) shows the result of blank removal. Fig. 5(c) shows the final result of brightness compensation. Fig. 6 shows aligned centroids of the corresponding shapes in 40 frames.

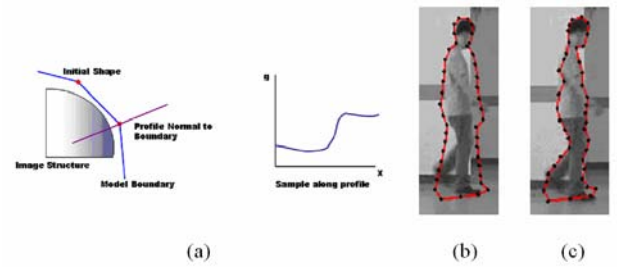


Fig. 4. Local structure: (a) search profile (b) initial shape (c) after local structure

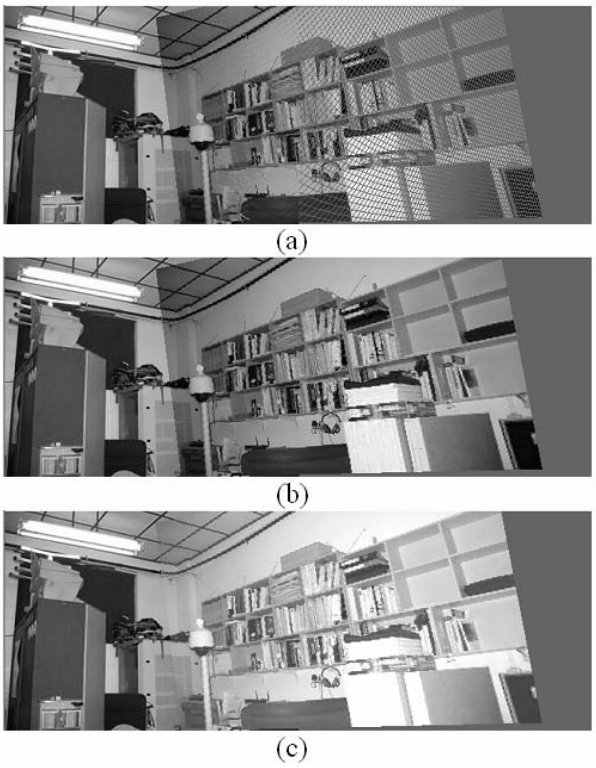


Fig. 5. A sequence of panoramic image creation: (a) the initially created panoramic image, (b) the result of blank removal, and (c) the result of brightness compensation

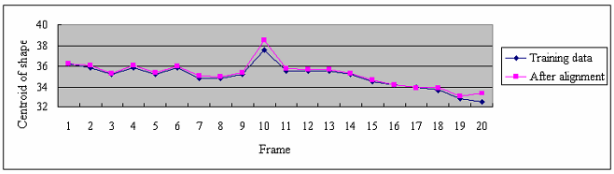
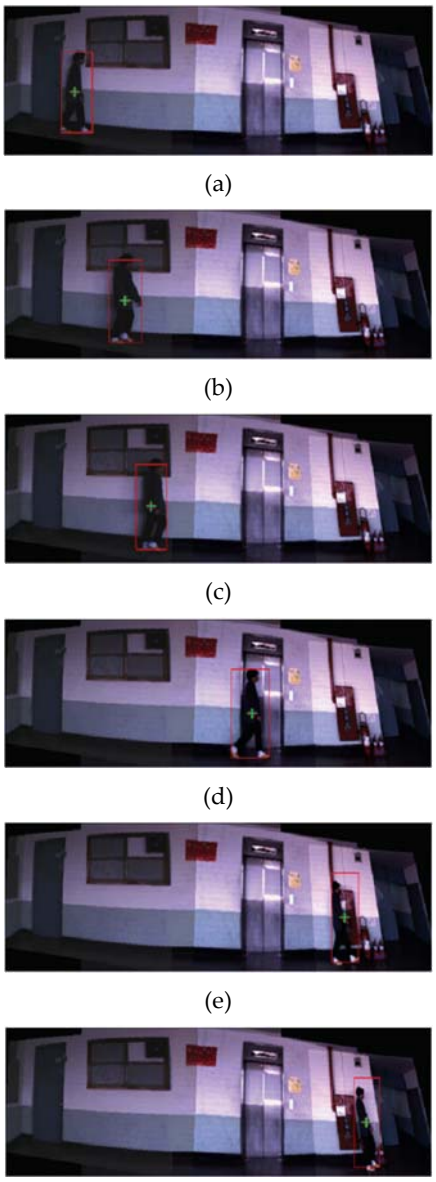


Fig. 6. Result alignment for training data

Fig. 7 shows the tracking result in the created panoramic video. In Fig. 7 the model fitting by using directional regularization before and after occlusion enables acceptable degree of occlusion handling. In the presence of occlusion the ASM-based tracker utilizes feature vectors of the region tracker. The region prediction and compensation algorithm is used to update the motion blob and region sizes using information from the previous frames. As a result the proposed ASM-based tracker can effectively deal with the occlusion problem. Table 1 shows the tracking speed can be obtained 20 f/s using ASM which can be considered as real-time. We also get tracking speed that can be obtained 24 f/s using down sampled image. Wavelet method has much complexity to make wavelet transformation.



(a) (b) (c) (d) (e) (f)

Fig. 7. Tracking results in a panoramic video with occlusion handling: (a) the 15th frame, (b) the 35th frame, (c) the 52nd frame, (d) the 77th frame, (e) the 94th frame, and (f) the 115th frame.

Method	Images size	Landmark point	Number of iterative	Frame second
Not processing	680×240	42	17	20.32
Down sampling	340×120	21	14	24.12
Wavelet	340×120	21	14	10.53

Table 1. Performance of each method

5. Conclusion

This paper presents an efficient approach to building a panoramic video from 4 cameras and object tracking. The proposed panoramic video creation module consists of two steps: (i) selecting four quasi-feature points in two adjacent frames acquired by corresponding multiple cameras and (ii) mosaicing the two images. Four pairs of selected quasi-feature points play a role of similarity reference in registering two adjacent frames. The mosaicing step uses the DLT algorithm. The centroids of four pairs of selected quasi-feature points define the proper landmark. We also presented the method that classifies flat and texture regions to extract the optimum quasi-feature point. In the experiment, the proposed quasi-feature point extraction provides improved results. The tracking speed was 24 f/s, which can be considered as real-time. Finally the proposed ASM-based tracker can effectively solve the occlusion problem. While most conventional panoramic image creation methods are pixel-based, the proposed feature-based method provides more accurate tracking result. Based on the experimental results, the created panoramic images exhibit high quality, which enables robust, real-time tracking.

6. References

[1] R. Marzotto, A. Fusiello, & V. Murino (2004). High resolution video mosaicing with global alignment, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, 692-698, 0-7695-2158-4, June 2004, IEEE Computer Society 2004, Washington, DC, USA

[2] H. Shum & R. Szeliski (2000). Construction of panoramic image mosaics with global and local alignment. *Int. Journal Computer Vision*, 36, 2, 101-130, 0920-5691

[3] M. Irani, P. Anandan, J. Bergen, Kumar, R., Hsu, S (1996). Efficient representations of video sequencese and their applications. *Signal process. Image Commun.*, 8, 4, 327-351, 0923-5965

[4] Z. Zhu, G. Xu, E. M. Riseman & A. R. Hanson (2006) Fast construction of dynamic and multi-resolution 360-degree panorama from video sequence. *Image Vision Comput.*, 24, 1, 13-26

[5] R. Hartley, A. Zisserman (2004) *Multiple view geometry in computer vision 2nd edition*, Cambridge University press, 0521540518

[6] I. Haritaoglu, D. Harwood & L. Davis (2000) W-4: Real-time surveillance of people and their activities, *IEEE Trans. Pattern Analysis, Machine Intelligence*, 22, 8, 809-830, 0162-8828

[7] D. Kim, Y. Yoon & J. Choi (2001) A quasi-feature based image mosaic algorithm using modified block matching criteria, *Trans. IEE Japan*, 121-C, 5, 892-898 0385-4221