Active Shape Model-Based Object Tracking in Panoramic Video

Daehee Kim, Vivek Maik, Dongeun Lee, Jeongho Shin, and Joonki Paik

Image Processing and Intelligent Systems Laboratory, Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia, and Film, Chung-Ang University, 221 Huksuk-Dong, Tongjak-Ku, Seoul 156-756, Korea
http://ipis.cau.ac.kr

Abstract. Active Shape Model (ASM) paradigm is a popular method for image segmentation where a priori information about the shape of the object of interest is available. The effectiveness of the method is contingent upon a correct correspondence between model points and the features extracted from the image. Extensive application of these models soon revealed one of their limitations when, for a given model point, no obvious salient point can be found in the image. The primary cause of such limitation is due to weak edges and presence of abrupt noise which is the case with low light surveillance video images. In this paper we propose a fusion-based panoramic tracking algorithm of in low light images using multiple sensors. The proposed algorithm uses an IR and CCD sensor for image capture. The proposed tracking system consists of three steps: (i) pyramid based fusion algorithm, (ii) reconstruction of panoramic image, and (iii) active shape model (ASM)-based tracking algorithm. The experimental results show that the proposed tracking system can robustly extract and track objects on panoramic images in real-time.

1 Introduction

Video surveillance system is useful for monitoring large and complex environments such as large building, airport, etc. Motion detection and object tracking play an important role in video-based surveillance system, some of which are, i) adaptive background generation and background subtraction [1, 2], ii) selective pixel integration [3], and iii) region based tracking [4]. In this paper we target our algorithm for low light visual images where above mentioned method fails to perform. Using fusion-based tracking we can integrate information from CCD and IR sensors and perform effective target tracking. For integrating and processing multi-sensor image we used panoramic imaging. The main advantage by doing so is they improve field of view to incorporate wide image observation region. Panoramic algorithm is carried out using direct linear transformation (DLT) and the singular value decomposition (SVD) algorithm. Panoramic algorithm is merged with the pyramid bared fusion to over come heavy computational overhead and provide better visual quality.

V.N. Alexandrov et al. (Eds.): ICCS 2006, Part IV, LNCS 3994, pp. 922 – 929, 2006.

2 Pyramid Based Fusion Algorithm

2.1 Pyramid Construction for Image Fusion

The pyramid representation can be used both for assessing the salience of the source image features, and for the reconstruction of the final image result. The following definitions for the pyramid are used. The fusion method described within this paper use a Laplacian pyramid representation. Laplacian pyramids are constructed for each image using the filter subtracts decimates (FSD*)* method. Thus the k'th level of the FSD Laplacian pyramid, k L, is constructed from the corresponding Gaussian pyramid level *k* based on the relationship.

$$L_k = G_k - wG_k = G_k(1-w), \tag{1}$$

Where *w* represents a standard binomial Gaussian filter, usually of 5x 5 spatial pixels extent. When constructing the FSD Laplacian, due to the decimation process and the fact that w is not an ideal filter, a reconstruction of the original image based on the FSD Laplacian pyramid incurs some loss of information.

2.2 Feature Saliency Computation

The feature saliency computation process, labeled sigma, expresses a family of functions that operate on the pyramids of both images yielding saliency pyramids. In practice, these functions can operate on the individual pixels or on a local region of pixels within the given pyramid level. The saliency function captures the importance of what is to be fused. When combining images having different focus, for instance, a desirable saliency measure would provide a quantitative measure that increases when features are in better focus. Various such measures, including image variance, image gradients, have been employed and validated for related applications such as auto focusing. The saliency function only selects the frequencies in the focused image that will be attenuated due to defocusing. Since defocusing is a low pass filtering process, its effects on the image are more pronounced and detectable if the image has strong high frequency content. One way to high pass filter an image is to determine its Laplacian or second derivative in our case.

$$\nabla^2 L_K = \frac{\partial^2 L_k}{\partial x^2} + \frac{\partial^2 L_k}{\partial y^2}, \tag{2}$$

In order to accommodate for possible variations in the size of texture elements, we computer the partial derivative by using a variable spacing between the pixels used to compute the derivatives. Hence a discrete approximation to the modified Laplacian is given by,

$$ML(i,j) = \left|2I(i,j) - I(i-1,j) - I(i+1,j)\right| + \left|2I(i,j) - I(i,j-1) - I(i,j+1)\right|, \tag{3}$$

Finally, the focus measure at a point (i, j) is computed as the sum of modified Laplacian values, in a small window around (i, j), that are greater than a threshold value.

$$F(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} M_k(x,y) \, for M_k(x,y) \geq T_1 \tag{4}$$

The parameter determines the window size used to compute the focus measure. In contrast to auto focusing methods, we typically use a small window of size, i.e. $N = 1$. The above equation can be referred to as sum modified Laplacian (SML).

3 Reconstruction of Panoramic Image Construction Algorithm

Most feature-based correspondence algorithms use DLT and SVD for 3D-transformation and interpolation. We begin with a simple linear algorithm for determining H given a set of four 2D to 2D point correspondences, $x_i \leftrightarrow x_i^{'}$. The transformation is given by the equation $x_i^{'} = Hx_i$. Note that this is an equation involving homogeneous vectors; thus the 2-vectors $x_i^{'}$ and Hx_i are not equal, they have the same direction but may differ in magnitude by a non-zero scale factor. The equation may be expressed in terms of the vector cross product as $x_i^{'} \times Hx_i = 0$. This form will enable a simple linear solution for H to be derived.

The SVD is one of the most useful matrix decompositions, particularly for numerical computations. Given a square matrix A, the SVD is a factorization of A as $A = UDV^{\mathrm{T}}$, where U and V are orthogonal matrices, and D is a diagonal matrix with non-negative entries. Note that it is conventional to write V^{T} instead of V in this decomposition. The decomposition may be carried out in such a way that this is always done. Thus a circumlocutory phrase such as "the column of V corresponding to the smallest singular value" is replaced by "the last column of V ."

4 Active Shape Tracker

The original ASM was first proposed by Cootes [2], [3]. Detection, analysis, and tracking a human body in a video sequence is a major application area for the ASM because the shape of the human body has unique combination of head, torso, and legs, which can be modeled with small number of parameters. In this section we briefly revisit ASM theory including three steps: (a) shape variation modeling, (b) model fitting, and (c) local structure modeling.

4.1 Shape Variation Modeling

Given a frame of input video, initial landmark points should be assigned on the contour of the object either manually or automatically. Good landmark points should be at or close to the desired boundary of each object. A particular shape *X* is represented by a set of *n* landmark points which approximate its outline as

$$X = [x_1, x_2, \cdots, x_n, y_1, \cdots, y_n]^T. \tag{5}$$

Different sets of such landmark points make a training set. A shape in the training set is normalized in scale, and aligned with respect to a common frame, as shown in Fig.1. Although each aligned shape is in the *2n*-dimensional space, we can model the shape with a reduced number of modes using the principal component analysis (PCA) analysis. The main modes of the template model, X , are then described by the eigen-vectors φ of the covariance matrix C, with the largest eigen values [9].

4.2 Model Fitting

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

$$E = (y - Mx)^T W^T (y - Mx), \tag{6}$$

where M represents the geometric transformation of scaling (s), translation (t), and rotation (θ). For instance if we apply the transformation to a single point, denoted by $[p,q]^T$, we have

$$M\begin{bmatrix} p \\ q \end{bmatrix} = s\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} p \\ q \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{7}$$

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

$$x_p = M^{-1} y. \tag{8}$$

Finally, the parameters are updated as

$$b = \phi^T (x_p - \bar{x}). \tag{9}$$

4.3 Local Structure Modeling

In order to interpret a given shape in the input image based on ASM, we must find a set of parameters that best match the model to the input shape. 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_i, *i*=1…, *n,* be the normalized derivative of a local profile of length K across the *i*-th landmark point, $\bar{g}$ and S_g the corresponding mean and covariance, respectively. The nearest profile can be obtained by minimizing the following Mahalanobis distance between the sample and mean of the model as

$$f(g_{i,m}) = (g_{i,m} - \overline{g})^T S_g^T \ (g_{i,m} - \overline{g}), \tag{10}$$

where $g_{i,m}$ represents the shifted version of g_i by m samples along the normal direction of the corresponding boundary. Figures 1 and 2 gives a schematic representation of the ASM. Fig.2. shows the result of PCA and local structure model fitting.

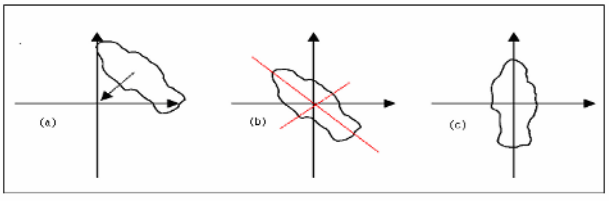


Fig. 1. Shape alignment: (a) Move centroid to origin, (b) find major axes of the shape,(c) rotation for alignment

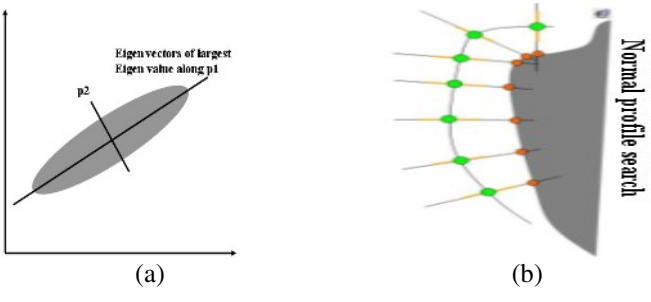


Fig. 2. (a) PCA analysis and (b) Local structure model fitting

5 Experimental Results

In this section we present the results of the proposed algorithm. Fig. 6 gives some examples of pyramid banned fusion algorithm. As can be seen the CCD sensor fails to provide complete object information in both canes due to low illumination environment. But fusion with IR image in Fig 3(f) results in much more distinguishable object for tracking. In Fig 4 we provide results of panorama view. The panoramic reconstruction is carried out on fused image to provide wide view for object tracking. The object tracking results after panorama reconstruction is carried out using ASM.

Experimental results show tracking in the dark no problem. This makes the best use of an advantage of CCD-camera and IR-camera.

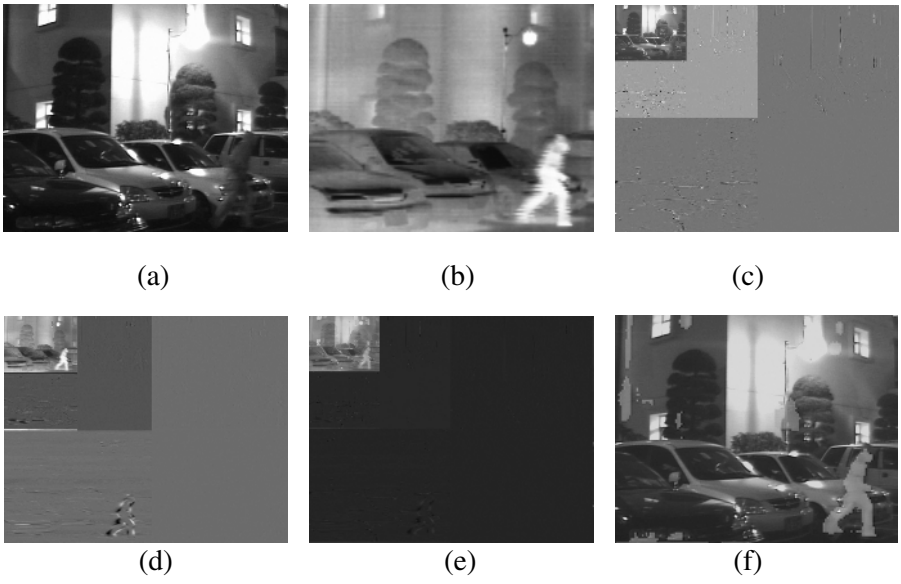


Fig. 3. Results of multi-dimensional fusion of IR and CCD camera sequences. (a) input ccd image, (b) input IR image, (c) daubechies wavelet representation of CCD, (d) daubechies wavelet representation of IR, (e) fusion result in wavelet domain, (f) reconstruction of fused wavelet.

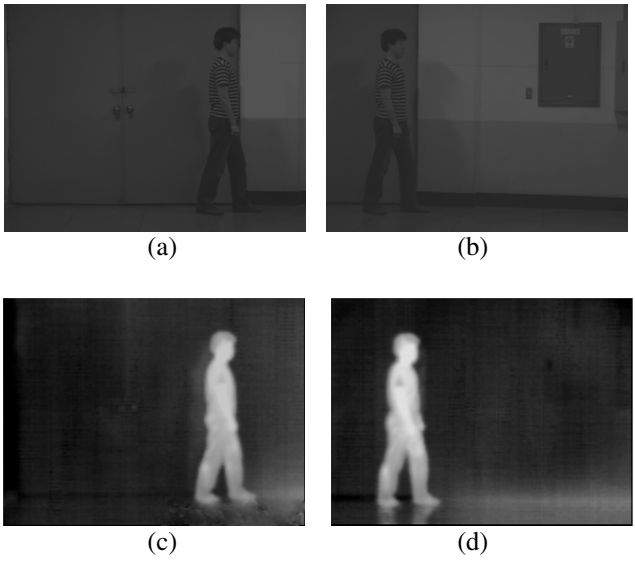
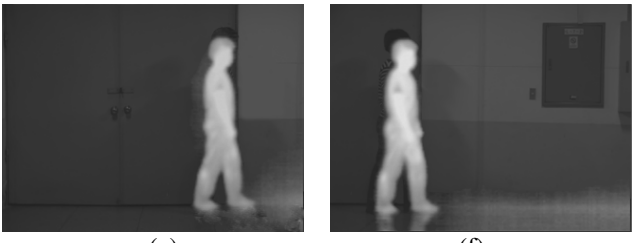


Fig. 4. Results of multi-dimensional fusion of IR and CCD camera sequences. (a), (b) input CCD image, (c),(d) input IR image, (e), (f) reconstruction of fusion



(e) (f)

Fig. 4. (Continued)

Fig. 5. Generated panoramic image

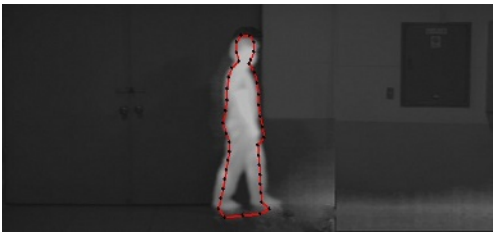


Fig. 6. ASM tracking on panoramic image

6 Conclusion

In this paper we proposed an automatic fusion-based panoramic tracking system. Pyramid fusion was extended to IR and CCD sensors. Active shape tracker was used to carry about object tracking. Experimental results prove the effectiveness of the proposed algorithm in real-time and low light environment.

References

[1] I. Haritaoglu D. Harwood and L.S Davis, "*W4: Real-time surveillance of people and their activities,*" IEEE Trans. Pattern Analysis, Machine Intelligence, vol. 22 no. 8, pp. 809-830, August 2000.

[2] A. Koschan, S. K. Kang, J. K. Paik, B. R. Abidi, and M. A. Abidi, "*Video object tracking based on extended active shape models with color information*," Proc. 1st European Conf. Color in Graphics, Imaging, Vision, pp. 126-131, University of Poitiers, France, April, 2002.

[3] P. Chang and J. Krumm, "*Object recognition with color occurrence histograms*," IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins, CO, June, 1999.

[4] T. Horprasert, D. Harwood, and L.S. Davis, "*A robust background subtraction and shadow detection*," Proc. ACCV'2000, Taipie, Taiwan, January 2000.

[5] C. Chen, W. Hsieh, J. Chen, "*Panoramic appearance-based recognition of video contents using matching graphs*," IEEE Transactions on Systems, Man, and Cybernetics-PART B: Cybernetics, vol.34, no. 1, February 2004.

[6] S. Kim, J. Kang, J. Shin, S. Lee, J. Paik, S. Kang, B. Abidi, and M. Abidi, "*Optical flow-based tracking of deformable object using a non-prior training active feature model*," PCM 2004, LNCS, vol. 3333, pp. 69-78, December 2004.

[7] Z. Zhu, A. Hanson, E. Riseman, "*Generalized parallel-perspective stereo mosaics from airborne video*," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 2, pp. 226-237, February 2004.

[8] R. Patil, P. Rybski, T.Kanade, and M. Veloso, "*People detection and tracking in high resolution panoramic video mosaic*," Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1323-1328, September 2004.

[9] C. R. Wren, A. Azerbayejani, T. Darrell, and A. P. Pentland, "*Pfinder: Real-time tracking of the humand body,*" IEEE Trans. Pattern Anal. Machine Intell., vol. 19, pp. 780-785, July 1997.