Enhanced Face Detection and Tracking In Video Sequence Using Fuzzy Face Model and Sparse Representation Technique

Manjunath Hiremath[1], P. S. Hiremath[2]

[1]Department of Computer Science,
Central University of Karnataka, Aland Road, Kalaburgi – 585 367

[2] Department of Computer Science (MCA),
KLE Technological University, BVBCET, Hubli -580 031, Karnataka

Abstract

The human face detection from video sequences is an important biometric component in the design of intelligent human computer interaction systems for video surveillance, face recognition, emotion recognition and face database management. In this paper, an automatic and robust method to detect human faces from video sequences and track the same is proposed. A novel algorithm for segmentation of face regions in video images based on fuzzy geometric face model is developed. The sparse representation algorithm is used to track the face along the video sequence. The proposed method, which is developed as a simple face detection and tracking approach, is implemented and evaluated with numerous experiments on videos containing large variations of head motion, light condition, and facial expressions. The experimental results show that the proposed method is effective in detecting and tracking human faces in videos.

Keywords: Face detection, Segmentation, Fuzzy geometric face model.

1. INTRODUCTION

The identification of objects and tracking their motion in a video is a significant research area, which has many important applications. The human face detection in a video frame and its feature extraction are considered to be essential requirements for intelligent vision-based security surveillance systems. Further, it can be used as a preprocessing mechanism for high-level applications, such as face recognition, facial gesture recognition, face liveness recognition, and so on. Since the human faces are non-rigid and have a high degree of variability in color, liveness and pose, several features of human face are not common to other pattern detection issues. Hence, human face detection is a more complex task in the computer vision. In addition, face variability occurs in inter-spatial relationship among facial features, namely, eyes, nose and mouth, and also in facial expressions. Occlusion and illumination conditions can also change the overall appearance of a face. Such face variabilities often lead to the large intra-class variations of the face distribution being highly nonlinear and complex in any space, which is linear to the original image space.

An effective face detection system shall detect the presence of all human faces in the image and give rough estimation of the positions of all detected faces in real time. The representation of face patterns, which strongly affect the performance of an automatic face detection system, is important for feature extraction and classification. A good representation has essentially small within-class variations, large between-class variations, and is robust to transformations without changing the class labels. Furthermore, its extraction shall be devoid of manual operation.

The literature on methods of face detection in digital images is abundant and can be found in [1-4,17-26]. The feature-based approaches form one face tracking category. Maurer and Malsburg [5] and McKenna et al. [6] used the features extracted by Gabor filters as the cues for face tracking. This approach is usually insensitive to global intensity changes and illumination changes. However, it is computationally expensive, and hence unsuitable for a real-time application.

The second face tracking category consists of active appearance model (AAM)-based approaches. AAMs are the generative, parametric models of certain visual objects that show both shape and appearance variations [7]. These variations are represented by linear models such as principal component analysis (PCA), which finds a subspace preserving the maximum variance of a given data. A face model can be constructed from training data using the AAM, and the face tracking is achieved by fitting the learned model to an input sequence. Recently, the view-based AAM [8] and the background-robust AAM [9,10] were proposed to handle the problems of face rotation and cluttered background, respectively. However, the fitting process of AAM fails occasionally and is not apt for a real-time application.

The third face tracking category comprises skin-color-based approaches. The property, that human facial colors of different people are clustered in a certain transformed 2-D color space, is used by Yang and Waibel [11], Raja et al. [12], Qian et al. [13], Jang and Kweon [14], Schwerdt and Crowley [15], and Stern and Efros [16] for face tracking. It is easy to implement, fast, and has a low computational cost. However, when the illumination condition changes drastically and the background image has a skin-like color, it certainly fails to track the face. Color adaptation methods have been proposed to overcome this problem. Raja et al. [12] used a Gaussian mixture model to represent the skin-color model and updated the parameters over time as the illumination condition changed. Qian et al. [13] collected pixel samples whose filtering scores, based on the general color model, exceed a prescribed threshold and updated the color model. Jang and Kweon [14] proposed an adaptive color model which used the basic structure of condensation. Stern and Efros [16] used a number of different color models and switched the models during the tracking sequence. However, these approaches, which used only skin color as tracking cue, invariably misidentify skin-colored objects in the background as faces.

In the present paper, a novel algorithm is proposed for the face detection in a video sequence of images based on fuzzy geometric face model, and for face tracking using the sparse representation based classification. The rest of this paper is organized as follows. The Section 2 describes the proposed method. The experimental results and discussion are given in the Section 3. The Section 4 concludes the paper.

2. PROPOSED METHOD

The proposed method for face detection and tracking in a video streaming comprises the following procedures: (i) Face detection using fuzzy geometric face model, (ii) Feature extraction using Gabor filter and HOG filter, (iii) Feature optimization using PCA, and (iv) Face tracking using sparse representation based classification (SRC) algorithm, which are described below.

(i)Fuzzy geometric face model for face detection

Firstly, different frames of the input video sequence are extracted. Then, select the first video frame as the key frame. The fuzzy geometric face model for face detection is applied[27,28] to the key frame as explained in the following: The extracted frame image is preprocessed and then the eyes are searched on the basis of geometrical knowledge of the symmetrical relations between eyes. The other prominent feature, namely, mouth, is searched with respect to the detected eyes using fuzzy rules, which are derived from the knowledge of the relative positions of the facial features in the human faces and the trapezoidal fuzzy membership functions to represent the uncertainty of the locations of the facial features due to variations in poses and facial expressions. A sample output of the face detection algorithm using fuzzy geometric face model is shown in the Figure 2.

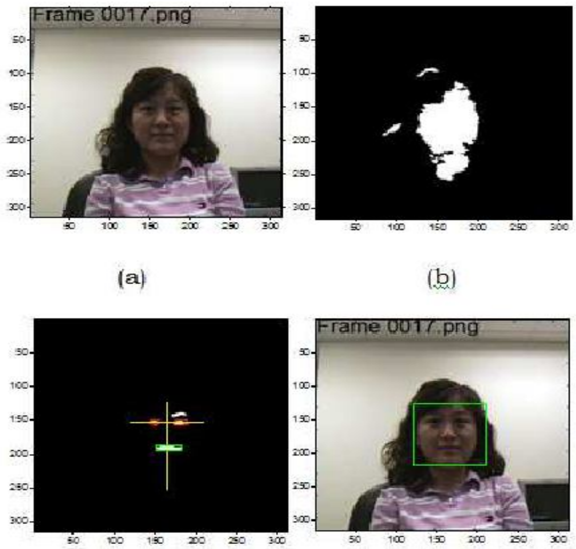


Figure 2: Face detection using fuzzy geometric face model: (a) Original input image, (b) Skin region extraction, (c) Construction of fuzzy face model, (d) Detected face.

(ii) Feature extraction using Gabor filter and HOG filter

In Gabor filter, the size of the input image is chosen in such a way that the filter results have always the same size after subsampling. Initially, all the training images are loaded and converted to grayscale to obtain a single sample in each pixel. Grayscale images have only black and white color channels where black has high intensity information and white color has the least. All the images are resized to 160 $*$ 160 dimensions to obtain same number of feature vectors for all the images. The variables are initialized to calculate edge detection and rotation values and then the feature vector is constructed. In the best resolution level, Gabor filters extract fine image structures of a small image region. In the lower resolution levels, coarse image structures can be extracted over large regions.

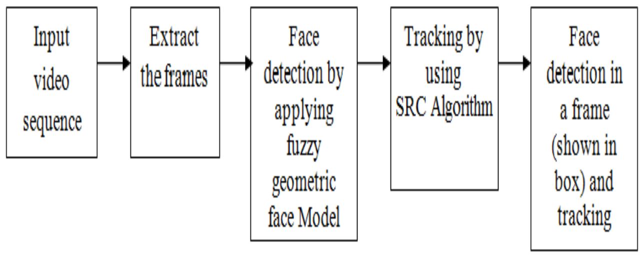


Figure 1. Block diagram of the proposed method

The HOG filters are used as a feature descriptor to extract features from image. The descriptor counts the occurrences of gradient orientation and magnitude in localized portions of an image. The orientation binning is performed to obtain rotation values. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The descriptor blocks are identified to perform normalization and to compute histogram, and then histograms of all cells are concatenated.

(iii) Feature optimization using PCA

The features obtained from Gabor and HOG filters are present in large dimensions, which often contain insignificant features and reduce the prediction accuracy. The dimensionality reduction of feature space is achieved by employing the principal component analysis (PCA). The feature vectors obtained by using PCA convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables.

(iv) Sparse representation based classification (SRC) algorithm

In SRC algorithm, the training gallery is given from a set of dictionary images and test face track is obtained from video data[29]. The sparsity weight parameter λ is initialized as 0.001. The description of the algorithm is as follows. Initially the training sample vector and test face track are given as an input. The feature vectors are normalized to unit length to solve the convex optimization equation. After normalization, the minimum coefficient vector for each track in a frame is obtained. Once the vectors are generated, the residuals are calculated for each class and the residual distribution is evaluated using sparsity index to predict the classes.

The main drawback of using convex optimization is that it uses minimum coefficient vector for each frame in test face track. Minimum coefficient vector may sometime lead to improper residuals during distribution which may lead to misclassification.

The algorithm of the proposed method is given below:

Algorithm:

Step 1. Input video sequence.

Step 2. Extract all the frames from the input video sequence, and then select the first video frame as the key frame.

Step 3. Apply the fuzzy geometric face model for searching the face region in the key frame using prominent face features, namely, eyes and mouth, to detect the face.

Step 4. Perform feature extraction using Gabor filter and HoG filter. Apply PCA for feature optimization.

Step 5. Select the next video frame and perform the sparse representation [29] algorithm to determine the object motion from one video frame to the next frame.

Step 6. Draw the rectangular box for the detected face in the next frame.

Step 7. Repeat the Step 5 and 6 till the end of the input video sequence, which results in tracking the detected face in the video sequence.

3. EXPERIMENTAL RESULTS AND DISCUSSION

To illustrate and test the proposed method, the extensive and representative experiments are performed using the Honda/UCSD Video database. The Honda/UCSD Video Database provides a standard video database for evaluating face detection, tracking and recognition algorithms. Each video sequence is recorded in an indoor environment at 15 frames per second, and each sequence lasted for at least 15 seconds. The resolution of each video sequence is 640x480. Every individual is recorded in at least two video sequences. In each video, the person rotates and turns his/her head in his/her own preferred order and speed, and typically in about 15 seconds, the individual is able to provide a wide range of different poses.

The Honda/UCSD Video Database contains two datasets. The first dataset is recorded by a SONY EVI-D30 camera at Honda Research Institute in 2002. It includes three different subsets, one each for training, testing, and occlusion testing. For proposed experiment only first data set is used. The implementation is done on Intel Core2Quad PC @ 2.60 GHz machine using MATLAB 7.0. The different frames of the input video sequence are extracted. The extracted frame image is preprocessed and then the facial features, namely, eyes and mouth, are searched by using fuzzy geometric face model, and then, the detected face is tracked by using SRC algorithm.

The experimental results are analyzed in terms of the two performance metrics, namely, precision and recall, which are defined as:

$$\textit{Precision}: \; P = \frac{A}{A+C} \times 100\%$$

$$\text{Recall}: \; R = \frac{A}{A+B} \times 100\%$$

where A is number of faces detected, B is number faces not detected, and C is the number of non-faces detected. The experimental results of the face detection using fuzzy geometric face model are given in Figure 3. The experimental results of tracking of face in video sequence are given in Figure 4.

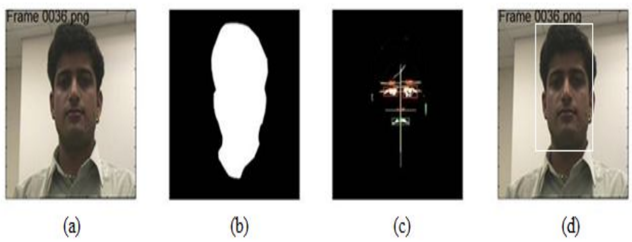
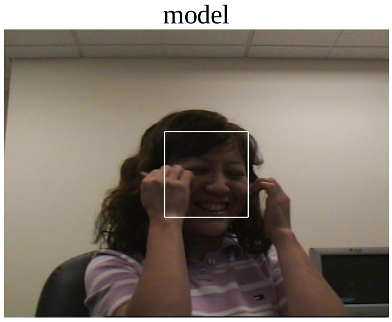


Figure 3. Face detection using Fuzzy geometric face model



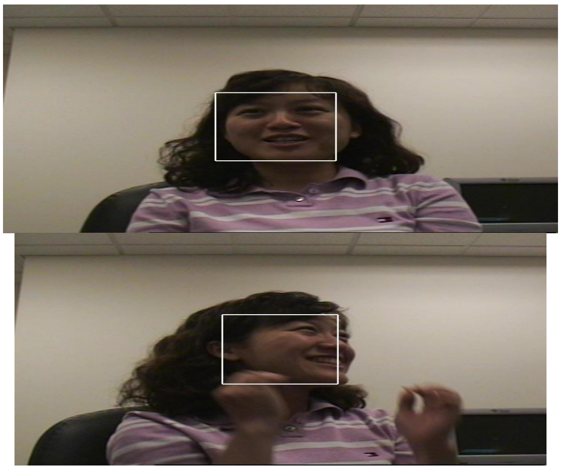


Figure 4. Face tracking in video sequence using Sparse Representation based classification algorithm

The comparison of the tracking results obtained by the proposed method and other methods in the literature are shown in the Table 1.

4.CONCLUSION

In this paper, a novel method of the detection and tracking of a face in a video is proposed. The detection of human face is achieved by using fuzzy geometric face model and face tracking is done by using sparse representation based classification algorithm. The single frontal face in the video frames with different motions, head tilts, lighting conditions, expressions and backgrounds are considered. The proposed approach yields better average detection and tracking, which is robust and runs in almost real time. It can be extended for multiple faces in a video sequence.

REFERENCES

[1] A. Yilmaz, O. Javed and M. Shah, “Object tracking: A survey”, ACMComput. Surv. 38(4) (2006).

[2] Kinjal A Joshi, Darshak G, Thakore, “A survey on moving object detection and tracking in video surveillance system”, International Journal of Soft Computing and Engineering Vol. 2, No. 3, 2012,pp.44-48.

[3] Huafeng Wang, Yunhong Wang andYuanCao, “Video based face recognition: A Survey”, World Academy of science, Engineeringand Technology, pp.293-301.

[4] Andrea F. Abate, Michele Nappi, Daniel Riccio, Gabriele Sabatino, “2D and 3D face recognition: A survey”, Pattern RecognitionLetters 28, 2007, pp.1885–1906.

[5] T. Maurer, C. Malsburg, Tracking and learning graphs and pose on image sequences of faces, in: Proceedings of the International Conference onAutomatic Face and Gesture Recognition, 1996, pp. 176–181.

[6] S. McKenna, S. Gong, R. Würtz, J. Tanner, D. Banin, Tracking facial feature points with Gabor wavelets and shape models, in: Proceedings of the International Conference on Audio and Video-based BiometricPerson Authentication, 1997, pp. 35–42.

[7] T. Cootes, G. Edwards, C. Taylor, Active appearance models, in: Proceedings of the European Conference on Computer Vision, 1998, pp.484–498.
[8] T. Cootes, G. Wheeler, K. Walker, C. Taylor, View-based activeappearance models, Image Vision Comput. 20 (9) (2002) 657–664.
[9] J.-W. Sung, D. Kim, A background robust face tracking using active contour technique combined active appearance model, in: Proceedings of the International Conference on Biometrics, 2006, pp. 159–165.
[10] J.-W. Sung, D. Kim, Large motion object tracking using active contour combined active appearance model, in: Proceedings of the International Conference on Computer Vision Systems, 2006, p. 31.
[11] J. Yang, A. Waibel, A real-time face tracker, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, 1996, pp. 142–147.
[12] Y. Raja, S. McKenna, S. Gong, Colour model selection and adaptation in dynamic scenes, in: Proceedings of the European Conference onComputer Vision, 1998, pp. 460–474.
[13] R. Qian, M. Sezan, K. Matthews, A robust real-time face tracking algorithm, in: Proceedings of the International Conference on Image Processing, 1998, pp. 131–135.
[14] G. Jang, I. Kweon, Robust real-time face tracking using adaptive color model, in: Proceedings of the International Symposium on Mechatronics and Intelligent Mechanical System for 21 Century, 2000.
[15] K. Schwerdt, J. Crowley, Robust face tracking using color, in: Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2000, pp. 90–95.
[16] H. Stern, B. Efros, Adaptive color space switching for tracking undervarying illumination, Image Vision Comput. 23 (3) (2005) 353–364.
[17] S. Birchfield, An elliptical head tracker, in: Proceedings of the 31stAsilomar Conference, 1997, pp. 1710–1714.
[18] M. Pardàs, E. Sayrol, A new approach to tracking with active contours, in: Proceedings of the International Conference on Image Processing,2000, pp. 259–262.
[19] X. Bing, Y. Wei, C. Charoensak, Face contour tracking in video using active contour model, in: Proceedings of the European Conference onComputer Vision, 2004, pp. 1021–1024.
[20] S. Birchfield, Elliptical head tracking using intensity gradients and color histograms, in: Proceedings of the Computer Vision and Pattern Recognition, 1998, pp. 232–237.
[21] H.-S. Lee, D. Kim, S. Lee, Robust face tracking using skin color and facial shape, in: J. Kittler, M.S. Nixon (Eds.), Lecture Notes in Computer Science, vol. 2688, AVBPA, 2003, pp. 302–309.
[22] M. Isard, A. Blake, Condensation-conditional density propagation for visual tracking, Int. J. Comput. Vision 29 (1) (1998) 5–28.
[23] D. MacKay, Learning in graphical models, Introduction to Monte CarloMethods, MIT Press, 1998, pp. 175–204 (Chapter 7).
[24] M. Isard, A. Blake, Icondensation: unifying low-level and high-level tracking in a stochastic framework, in: Proceedings of the EuropeanConference on Computer Vision, 1998, pp. 893–908.
[25] A. Blake, M. Isard, Active Contours, Springer, Berlin, 1998.
[26] A. Blake, M. Isard, D. Reynard, Learning to track the visual motion ofcontours, Int. J. Artif. Intell. 78 (1995) 101–134.
[27] P. S. Hiremath, Manjunath Hiremath, " Fuzzy geometric face model for face detection based on skin color fusion model", Proceedings of International Conference on Advances in Computing Advances in Intelligent Systems and Computing, M. S. Ramaiah Institute of Technology, Bangalore, 4-6 July 2012, Vol.174, pp 977-982, DOI : 10.1007/978-81-322-0740-5_118, Print ISBN-978-81-322-0739-9. Springer India (2012)
[28] P. S. Hiremath, Manjunath Hiremath, "Fuzzy face model for face detection using eyes and mouth features", Proceedings of the 3rd International conference on Cognition and Recognition, 9-10 December 2011, Department of Studies in Computer Science, University of Mysore, Mysore, ICCR 2011.
[29] Shruthi Nagendraa,∗, R. Baskarana and S. Abirami, "Video-Based Face Recognition and Face-Tracking using SparseRepresentation Based Categorization", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015).
[30] Jifeng Ning, Lei Zhang, David Zhang and Chengke Wu, "Robust Object Tracking using Joint Color-texture Histogram", International Journal of Pattern Recognition and Artificial Intelligence Vol. 23, No. 7 (2009), pp.1245–1263
[31] Shaohua Zhou, Rama Chellappa, Baback Moghaddam, "Visual Tracking and Recognition Using Appearance-Adaptive Models in Particle Filters", IEEE Transactions on Image Processing, Vol. 13, No. 11, 2004, pp. 1491-1506.
[32] P. S. Hiremath, Manjunath Hiremath, Mahesh, "Face Detection and Tracking in Video Sequence using Fuzzy Geometric Face Model and Motion Estimation", International Journal of Computer Applications (0975-8887), Volume 58, No.15, November2012
[33] P. S. Hiremath, Manjunath Hiremath, Mahesh, Face Detection and Tracking in Video Sequence using Fuzzy Geometric Face Model and Mean Shift, International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013), ISSN 2278-3091

Table 1. Comparison of the tracking results obtained by the proposed method and other methods

Parameters	Jifeng Ning et.al[30]	Shaohua Zhou et al. [31]	Hiremath et. al. [32,33]	Proposed Method
Video	Face	Face	Face	Face
#video frames	740	800	395	395
Frame rate	120 fps	--	15 fps	15 fps
Frame Size	352x240	240x360	640x480	640x480
Occlusion	No	Yes	Yes	Yes