ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

Face Detection and Tracking in Video Sequence using Fuzzy Geometric Face Model and Mean Shift

P. S. Hiremath, Manjunath Hiremath, Mahesh R.
Department of Computer Science
Gulbarga University, Gulbarga - 585106
Karnataka, India
hiremathps53@yahoo.com, manju.gmtl@gmail.com, maheshswamy99@gmail.com

Abstract : Humans make use of face as an important cue for identifying people. This makes automatic face detection very crucial from the point of view of a wide range of commercial and law enforcement applications. While traditional face detection is typically based on still images, face detection and tracking from video sequences has become prominent research domain. In this paper, we propose a novel algorithm which segments the face region in video images using fuzzy geometric face model in key frame. The mean shift is used to track the face along the video sequence. Contrary to current techniques that are based on huge learning databases and complex algorithms to get generic face models, the proposed method handles simple face detection and tracking approach. The proposed method is implemented and evaluated with numerous experiments on videos containing large variations of head motion, light condition, and expressions. The experimental results show that the proposed method is effective in detecting and tracking faces in videos.

Key words : Face detection, fuzzy geometric face model, mean shift, face tracking.

INTRODUCTION

Face detection is required as the first step of the automatic face image analysis system. Face detection system has been widely investigated in recent years because it overlies many areas of application such as face recognition, man-machine interaction systems, visual communication systems, video-surveillance, etc. However, face detection is a challenging task due to variation in illumination, scale, location, orientation and pose. Many methods for face detection in still images and in video sequences have been reported in the literature, which have achieved some encouraging results.

A comprehensive survey on methods of face detection in images can be found in [1-6]. In general, face detection techniques can be divided into two categories: model-based technique and feature-based technique. The first one assumes that a face can be represented as a whole unit. Several statistical learning mechanisms are explored to characterize face patterns, such as neural network, Bayesian classifier and boosting algorithm. The second category considers a face as a collection of components. Important facial features such as eyes, nose and mouth are extracted from face image, and by using their locations and relationships, the faces in image are detected. Among feature-based face detection methods, using skin color as a detection cue is very popular. Skin color is important and powerful information for human face. Identification of skin region in an image can be used as the first step in face detection process in color images. Many researchers use skin color models to locate potential face regions and then examine the locations of faces by analyzing each face candidate's shape and physical geometric information. In order to represent the human face, finding the efficient invariant features for face detection is still an open problem.

Real-time object tracking is a critical task in computer vision applications. Many tracking algorithms have been proposed to overcome the difficulties arising from noise, occlusion, clutter and changes in the foreground object or in the background environment. Among the various tracking algorithms, mean shift tracking algorithms have recently become popular due to their simplicity and efficiency [10, 13, 16, 17, 18]. In the object tracking system, face tracking is pervasive application. In face tracking system, face tracker estimates the face trajectory by locating its position in every frame of the sequence. While this information may be sufficient for some applications (e.g. detecting the presence of an intruder), other applications require additional data, like knowing the orientation, extension or even the precise contour of the faces at every frame (e.g. facial expression recognition).

The mean shift algorithm was originally proposed by Fukunaga and Hostetler[14] for data clustering. It was later introduced into the image processing community by Cheng[11]. Bradski[15] modified it and developed the Continuously Adaptive Mean Shift (CAMSHIFT) algorithm to track a moving face. Comaniciu and Meer[12] successfully applied mean shift algorithm to image segmentation[12] and Comaniciu et al,[13] have applied it to object tracking[13]. Mean Shift is an iterative kernel-based deterministic procedure which converges to a local maximum of the measurement function with certain assumptions on the kernel behaviors. Furthermore, mean shift is a low complexity algorithm, which provides a general and reliable solution to object tracking and is independent of the target representation.

In the present paper, a novel algorithm is proposed for the face region segmentation in a video sequence of images using fuzzy geometric face model, which employs the mean shift to track the face along the video streaming.

MATERIALS AND METHODS

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

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

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. Each subset contains 20, 42, 13 videos, respectively, from 20 human subjects. The second dataset is recorded by a SONY DFW-V500 camera at Computer Vision Laboratory, University of California, San Diego, in 2004. It includes two subsets, one each for training and testing, of 30 videos from another 15 different human subjects [7].

The proposed algorithm is experimented with the video sequences drawn from the above data sets.

Fig.1: Face detection using fuzzy geometric face model (a) Original input image (b) Skin region extraction (c) Construction of fuzzy face model (d) detected face

PROPOSED METHODOLOGY

The proposed methodology comprises the application of fuzzy geometric face model for face detection and mean shift for face tracking in video sequences, which are described below.

A. Fuzzy Geometric Face Model for Detection

The fuzzy geometric face model for face detection [8] is applied to the input video sesquence. 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 and the face detection algorithm. The fuzzy rules 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. The frames of the input video sequence are expected to contain, not too dark or too bright, images. If the input frame of the video sequence is a color frame, it is converted into gray scale image. The gray scale frame is filtered using the Sobel horizontal edge emphasizing filter and utilizing the smoothing effect by approximating a vertical gradient. In the filtered frame, objects of interest (facial features) are brighter than the background. In order to make the essential facial features clearly visible, filtered image is converted into binary frame by simple global thresholding. Further, the frame is denoised by morphological operations, in which opening operation is performed to remove noise, and then the closing operation is performed to remove holes. Then the active pixels are grouped into maximal connected blocks to get the regions or blocks which are labeled. After the labeling process, for each feature block, its center of mass(x, y), orientation θ, bounding rectangle and the length of semi major axis are computed. The resultant images in different steps of the face detection algorithm using fuzzy geometric face model are shown in the Fig. 1.

B. Mean Shift Algorithm

There are various methods available for tracking an object in a video sequence. They can be categorized as: deterministic and probabilistic. Deterministic method looks for the local maxima of a similarity measure between the object model and the target iteratively, and mean-shift is one such technique for tracking a moving object. Mean shift is a versatile algorithm that has found a lot of practical applications - especially in the computer vision where the dimensions are usually low. Hence, the mean shift is used to perform lots of common tasks in machine vision. The mean shift algorithm is a non-parametric method. It provides accurate localization and efficient matching without expensive exhaustive search. It is an iterative process, which computes the mean shift value for the current point position, then moves the point to its mean shift value as the new position, and compute the mean shift until it fulfills certain condition [9].

Mean shift is a nonparametric density gradient estimator. It is employed to derive the object candidate that is the most similar to a given model while predicting the next object location. In other words, it starts from the position of the model in the current frame and then searches in the model's neighborhood in next frame, followed by finding best candidate by maximizing a similarity function. Finally, the same process is repeated in the next pair of frames. Mean shift considers feature space as an empirical probability density function. If the input is a set of points, then Mean shift considers them as sampled from the underlying probability density function. If dense regions are present in the feature space, then they correspond to the mode (or local maxima) of the probability density function. For each data point, mean shift associates it with the nearby peak of the dataset's probability density function and defines a window

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

around it and computes the mean of the data point. Then it shifts the center of the window to the mean and repeats the algorithm till it converges. After each iteration, hence, the window shifts to a denser region of the dataset.

A target is usually defined by a rectangle or an ellipsoidal region in the image. Most existing target tracking schemes use the color histogram to represent the rectangle or ellipsoidal target. In this paper, a new target representation approach is adopted by using the fuzzy geometric face model and mean shift algorithm. Firstly, the target representation in the mean shift tracking algorithm is described in the following: Denote by $\{x_i^*\} i = 1 \cdots n$ the normalized pixel positions in the target region, which is supposed to be centered at the origin point. The target model $\hat{q}$ corresponding to the target region is computed as

$$\begin{cases} \hat{q} = \{\hat{q}_u\} u = 1 \cdots m, \\ \hat{q}_u = C \sum_{i=1}^{n} k(\| x_i^* \|^2) \delta(b(x_i^*) - u) \end{cases} \quad (1)$$

where $\hat{q}_u$ represent the probabilities of feature u in target model $\hat{q}$, m is the number of feature spaces, δ is the Kronecker delta function, $b(x_i^*)$ associates the pixel x_i^* to the histogram bin, $k(x)$ is an isotropic kernel profile and constant C is a normalization function defined by

$$C = 1 / \sum_{i=1}^{n} k(\left\| x_i^* \right\|^2)$$

(2)

Similarly, the target candidate model $\hat{p}(y)$ corresponding to the candidate region is given by

$$\begin{cases} \hat{p}(y) = \{\hat{p}_u(y)\} u = 1 \cdots m \\ \hat{p}(y) = C_h \sum_{i=1}^{n_h} k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta(b(x_i) - u) \end{cases} \quad (3)$$

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)}$$

(4)

where $\hat{p}_u(y)$ represents the probability of feature u in the candidate model $\hat{p}(y)$, $\{x_i\} i = 1 \cdots n_h$ denote the pixel positions in the target candidate region centered at y, h is the bandwidth and constant C_h is a normalization function.

In order to calculate the likelihood of the target model and the candidate model, a metric based on the Bhattacharyya coefficient is defined between the two normalized histograms $\hat{p}(y)$ and $\hat{q}$ as follows:

$$\rho(\hat{p}(y), \hat{q}) = \sum_{u=1}^{m} \sqrt{\hat{p}_u(y) \hat{q}_u}$$

(5)

The distance between $\hat{p}(y)$ and $\hat{q}$ is then defined as

$$d(\hat{p}(y), \hat{q}) = \sqrt{1 - \rho(\hat{p}(y), \hat{q})} \quad (6)$$

Minimizing the distance given by the Eq.(6) is equivalent to maximizing the Bhattacharyya coefficient given by the Eq.(5). The iterative optimization process is initialized with the target location y_0 in the previous frame. Using Taylor expansion around $\hat{p}_u(y_0)$, the linear approximation of the Bhattacharyya coefficient is obtained as

$$\rho(\hat{p}(y), \hat{q}) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{\hat{p}_u(y_0) \hat{q}_u} + \frac{1}{2} C_h \sum_{i=1}^{n_h} w_i k\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)$$

(7)

where

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y_0)}} \delta(b(x_i) - u) \quad (8)$$

Since the first term in Eq.(7) is independent of y, to minimize the distance in the Eq.(6) is to maximize the second term in the Eq.(7). In the iterative process, the estimated target moves from y to a new position y_1, which is defined as

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n_h} w_i g\left(\left\| \frac{y - x_i}{h} \right\|^2 \right)} \quad (9)$$

When the kernel g with the Epanechnikov profile designed by the Eq.(9) is reduced to

$$y_1 = \frac{\sum_{i=1}^{n_h} x_i w_i}{\sum_{i=1}^{n_h} w_i}$$

(10)

By using Eq.(10), the mean shift tracking algorithm finds, in the new frame, the most similar region to the object.

The mean shift algorithm can be extended to be used with sequence of colour images and to fit with the varying scale of the object throughout the sequence. It works well in the absence of occlusion.

C. Proposed Method

The basis for the proposed algorithm for face detection and tracking in a video sequence is the object-tracking algorithm for a moving target. A rectangular target window is defined in an initial frame of detected face region by applying fuzzy geometric face model, and then the data within that window is processed to track the object in the video sequence using

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

mean shift algorithm. The block diagram of the proposed method for face detection and tracking in a video sequence is shown in the Fig. 2.

The algorithm of the proposed method is given below:
Algorithm: Face detection and tracking

(1) Input video sequence.
(2) Extract all the frames from the input video sequence, and then select first video frame as key frame.
(3) Apply the fuzzy geometric face model[9] for searching the face region in the key frame using prominent face features, namely, eyes and mouth, to detect the face.
(4) Select the next video frame and perform the mean shift computation to determine the object motion from one video frame to the next frame.
(5) Draw the rectangular box for the detected face in the frame.
(6) Repeat the Step 4 and 5 till the end of the input video sequence, which results in tracking the detected face in the video sequence.

geometric face model. Then, the consecutive frames from a video sequence and their corresponding mean shift are estimated and face is tracked. In the proposed method, single frontal face in the video frames with different motions, head tilts, lighting conditions, expressions and backgrounds are considered.

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

Parameters	Jifeng Ning et.al[9]	Shaohua Zhou et al. [19]	Proposed Method
Video	Face	Face	Face
#video frames	740	800	395
Frame rate	120 fps	--	15 fps
Frame Size	352x240	240x360	640x480
Occlusion	No	Yes	Yes

The proposed approach yields better average detection and tracking, which is robust and almost real time. The implementation of the proposed method is evaluated with numerous experiments, which yielded encouraging results that demonstrate the effectiveness in detecting and tracking faces in videos. The proposed method can be extended for multiple faces in a video sequence by considering multiple face detection and tracking algorithms.

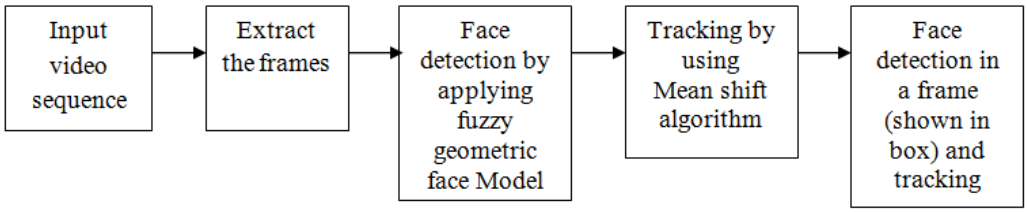


Fig. 2: Block diagram of the proposed approach

EXPERIMENTAL RESULTS AND DISCUSSION

The extensive and representative experiments are performed to illustrate and testify the proposed fuzzy geometric face detection and mean shift tracking based model. The experimentation of the proposed approach is carried out using the Honda/UCSD Video database [7]. 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 [8], and then, the detected face is tracked by using mean shift algorithm [9]. The experimental results of the face detection and tracking in video sequence by the proposed method are shown in the Fig. 3.

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

CONCLUSION

In this paper, a novel method for detection and tracking of a face in a video sequence based on the fuzzy geometrical face model and mean shift tracking is presented. The human face is detected by feature extraction process based on fuzzy

ACKNOWLEDGEMENT

The authors are indebted to the University Grants Commission, New Delhi, for the financial support for this research work under UGC-MRP F.No.39-124/2010 (SR).

REFERENCES

[1] A. Yilmaz, O. Javed and M. Shah, "Object tracking: A survey", ACM Comput. 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, Engineering and Technology, pp.293-301.
[4] Andrea F. Abate, Michele Nappi, Daniel Riccio, Gabriele Sabatino, "2D and 3D face recognition: A survey", Pattern Recognition Letters 28, 2007, pp.1885–1906.
[5] Ing-Sheen Hsieh, Kuo-Chin Fan, and Chiunhsiun Lin, "A Statistic Approach to the Detection of Human Faces in Color Nature Scene", Pattern Recognition, 35, 2002, pp.1583-1596.
[6] J. G. Wang and T. N. Tan, "A New Face Detection Method Based on Shape Information", IEEE Transactions on Pattern Recognition Letters, Vol. 21, 2000, pp.463-471
[7] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang, David Kriegman, "Visual tracking and recognition using probabilistic appearance

manifolds", Computer Vision and Image Understanding, Vol. 99, 2005, pp.303–331.

[8] P. S. Hiremath and Manjunath Hiremath, "Fuzzy face model for face detection using eyes and mouth features", International Journal of Machine Intelligence, Vol. 3, No. 4, 2011, pp.185-190.

[9] 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

[10] G. Bradski, "Computer vision face tracking for use in a perceptual user interface", Intel Technologies. Journal. Q2(2) (1998), pp.12–21.

[11] Yizong Cheng Y. Cheng, "Mean Shift, Mode Seeking, and Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, NO. 8 (1995), pp.790–799.

[12] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", IEEE Transactions Pattern Analysis and Machine Intelligence. Vol.24(5) (2002), pp.603–619.

[13] D. Comaniciu, V. Ramesh and P. Meer, "Kernel-based object tracking", IEEE Transactions Pattern Analysis and Machine Intelligence. Vol. 25(5) (2003), pp.564–575.

[14] K. Fukunaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition", IEEE Trans. Inform. Th.21(1) (1975), pp.32–40.

[15] G. Bradski, "Computer vision face tracking for use in a perceptual user interface", Intel Technol. J.2(2) (1998), pp.12–21.

[16] I. Haritaoglu and M. Flickner, Detection and tracking of shopping groups in stores, Proc. IEEE Conf. Computer Vision and Pattern Recognition, Kauai, Hawaii, 2001, pp. 431–438.

[17] Q. A. Nguyen, A. Robles-Kelly and C. Shen, "Enhanced kernel-based tracking for monochromatic and thermographic video", Proc. IEEE Conf. Video and Signal Based Surveillance(2006), pp.28–33.

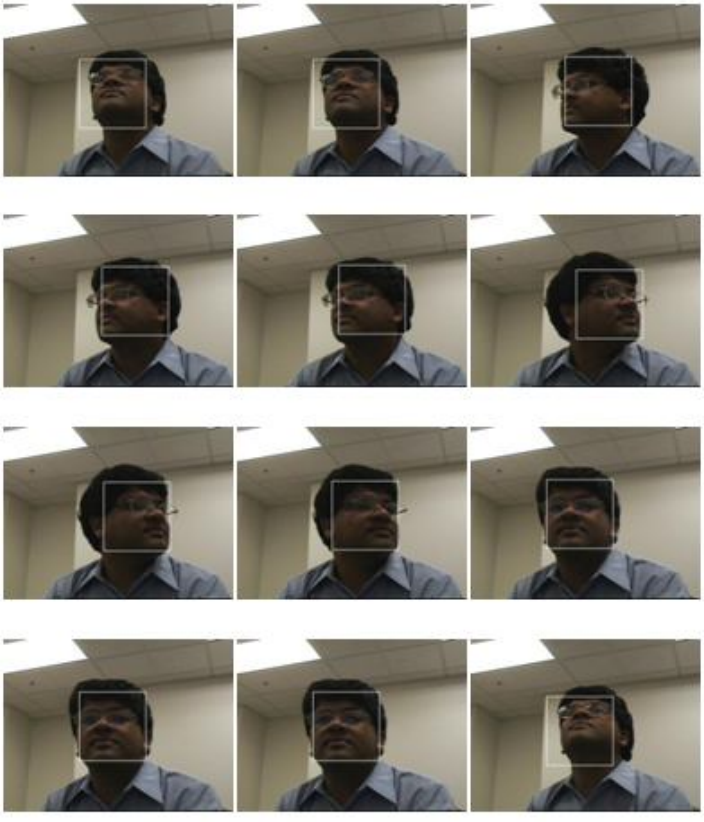
[18] C. Yang, D. Ramani and L. Davis, "Efficient mean-shift tracking via a new similiarity measure", Proc. IEEE Conf. Computer Vision and Pattern RecognitionI(2005), pp.176–183.

[19] 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.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 41-46 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

(a)



(b)

Fig 3: The experimental results obtained by the proposed method of the face detection and tracking in a sample video sequence