Unconstrained Human face Tracking in live Video

Ramakrishna B. B, M. SharmilaKumari

Abstract: In surveillance applications visual face detection and tracking becomes an essential task. Many algorithms and technologies have been developed to automatically monitor pedestrians or other moving objects and to track the detected face. One main difficulty in face tracking, among many others, is to choose suitable features and models for detecting and tracking the target. For tracking of faces there are some common features are considered like color, intensity, shape and feature points. In this paper we discuss about mean shift based face tracking based on the color, optical flow tracking based on the intensity and motion, SIFT face tracking based on scale invariant local feature points. Mean shift is then combined with local feature points. Initial results from tries have shown that the implemented method is able to track target face with different pose variation, rotation, partial occlusion and deformation.

Keywords: Face detection, Face tracking, Mean shift, SIFT, NPD

I. INTRODUCTION

Face detection and tracking in live videos is an essential task within the field of computer vision. It aims at locating a moving face or several ones in time using a digital camera. An algorithm analyses the video frames and outputs the location of moving faces within the video frame. So it can be defined as the process of segmenting a face object of interest from a video scene and keeping track of its motion, orientation, occlusion etc. so as to extricate useful information by means of some algorithms. Its primary task is to find and follow a moving face object or several faces in image sequences. The rise of powerful computers and the growing need for mechanized video exploration have created a lot of interest in face tracking algorithms. The use of visual face tracking is appropriate in the tasks of computerized surveillance, vehicle routing, traffic monitoring, human-computer interaction and so forth. Computerized video surveillance compacts with real time monitoring of people or vehicles in full or restricted environments leading to tracking and analysis of activity of the subjects in the field of view. Mainly there are three key stages in video surveillance: detection of faces from exciting moving objects, tracking faces from such objects, and analysis of such object tracks to identify their behaviour.

II. RELATED WORK

a. Mean Shift: Accurate face tracking under the limitation of low computational complexity makes a challenge. In real-time applications like surveillance and monitoring [3], actual user interfaces [4], smart rooms [6], and video compression [7, 8] all require the ability to track face in moving objects. In general, tracking of visual faces should be possible either by forward-tracking or by back-tracking. The forward-tracking method estimates the points of the regions in the current frame with the segmentation result obtained for the frame in the preceding image. The back-tracking based methodfragments foreground region in the present image and then creates the correspondence of regions between the frames from the previous image. For building up correspondence, a few several face templates are used. A conceivable forward-tracking technique is Mean-shift method. In Mean shift method kernel function is defined to find the distance between sample points and its mean shift, also the weight coefficient is reverse with the distance. As the closer the distance is, the bigger the weight coefficient is. Defined Mean shift algorithm is a non-parametric process [9]. Mean shifftereestruelocalization and efficient matching without costly comprehensive search. It is aniterative procedure, in other words, first calculate the mean shift value for the new point position, after that move the point to its mean shift value which is the new position, then calculate the mean shift until it satisfy certain condition. The standard of Mean shift procedure can be gained by analysing entire process.

Mean shift is applied in real-time face tracking is published in [10] called as kernel based tracking or Mean shift tracking. The shape and size of the interest face area is normally described by two kinds of kernel function. The first one is Epameknikov $M_E(x)$ kernel, and its kernel profile is

$$M_E(x) = \begin{cases} \frac{1}{2}B_z^{-1}(z+2)(1-V^TV) & \text{if } V^TV < 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where $B_z$ is the volume of the unit $z$-dimensional sphere, \{x\}_i, (i=1, ..., n) be an random set of n points in the $z$-dimensional Euclidean space $R_z$. With the contour of this kernel, the pixels near the centre could weight higher value than the pixels in the edges which helps in tracking the centrepoint of the interested object [11, 12]. The kernel-based face tracking algorithm (Mean shift algorithm) is as follows:
The target model \( \{ \hat{T}_u \} \) where \( u = 1, 2, \ldots, m \) bins of histograms is derived from an elliptic region centered at \( y_0 \), to remove the influence of the target scale it is normalized to a unit circle, its pixel coordinates \( \{ z'_i \} \).

\[
\hat{T}_u = \frac{\sum_{i=1}^{n} (1 - |z'_i|^2) \delta(c(z'_i) - u)}{\sum_{i=1}^{n} (1 - |z'_i|^2)}
\]

where \( n \) is the number of pixels, \( \delta \) is the Kronecker delta function:

\[
\delta(z) = \begin{cases} 
1, & z = 0 \\
0, & \text{otherwise}
\end{cases}
\]

Use \( y_0 \) from earlier frame location which as an initial position to estimate pixel in the current frame. Develop weights for \( i=1,2,\ldots,n \) according. Then determine the new location of the target region according to calculate the new probability value and determine that value. Only if the correspondence between the new targets region and the target mode is very less than that between the old target region and the model.

Carry out the remaining processes of this step to shift the target region half way between the new and old locations and then evaluate the similarity function in this identified new location [16].

The Mean shift approach is a non-parametric density gradient estimator. It is utilized to develop the object candidate which is the most similar to a given model while calculating the next new objects face location. That is it starts from the position of the model object in the current frame and then examines in the model’s neighbourhood in the next new frame, followed by calculating finest candidate by maximizing a similarity function [17, 18]. At last, repeats the same process in the next pair of new frames.

b. Scale Invariant Feature Transform (SIFT): Scale Invariant Feature Transform algorithm is an approach for detecting and extracting local feature descriptors which are sensibly invariant to changes in the scaling, rotation, illumination, image noise and small changes in viewpoint. SIFT algorithm is first proposed by David Lowe in 1999, and further modified and developed. The SIFT features have many gains such as SIFT features are all natural features of images. They are constructively invariant to image scaling, rotation, illumination, viewpoint, translation, noise etc. SIFT algorithm is suitable for fast and accurate matching in amass of feature databases [20].

SIFT algorithm contains following stages for extracting and detecting features

1. Scale-space extrema detection: The initial stage of calculation searches overall measures and image localities. This is implemented affectively by means of a difference of Gaussian function to recognize potential interest points in the frame that are invariant to rotation and scale.
2. Keypoint localization: In a frame at each candidate location, a complete model is fit to identify scale and location. Keypoints are chosen on basis of measures of their stability.
3. Orientation assignment: Multiple orientations are allocated to each key point location on basis of local image gradient directions. All imminent processes are performed on image data that has been converted relative to the allocated orientation, scale and location for each feature, thus providing invariance to all these transformations.
4. Generation of keypoint descriptors: At the selected scale keypoint gradients are measured in the region nearby each keypoint. These selected gradients are converted into a representation which accepts major levels of local change in shape distortion and illumination.

III. COMBINED METHOD: SHIFT AND MEAN SHIFT

Proposed tracking algorithm is an effective integration using of Mean shift and SIFT feature tracking. The proposed approach will apply as similarity measurement between two neighboring frames in terms of color and SIFT correspondence. Theoretically, a face track will be made if SIFT and Mean shift feature tracking lead to calculate probability distributions like color and intensity within the resultant area in the next frame of the image (normally the two probability distributions should be indistinguishable if the scenario does not change considerably). Whenscenarios maximization algorithm is applied in order to track a full likelihood estimate using the results from Mean shift and SIFT correspondence. In this combined method first, select a region of interest on the first frame as a reference object model. This reference model is described by its probability distributions function (PDF) estimation which is a m-bin color histogram \( m \) in a rectangular (or ellipse) region centred at \( y_0 \) and window size \( W \).

\[
\sum_{q=1}^{p} m_q = 1
\]

In which \( m_q \) is the probability distribution of color \( q \).

In the next stage SIFT method is applied on the next frame. If there are many corresponding points between the reference object region and a candidate region, an affine matrix is estimated. Picking up the four corners of the rectangle model in the first frame, the affine matrix is used to transform the location of the old four corners to new ones, averaging its new positions. Which effect in the second frame, just average these positions and then obtain its y1_ SIFT. The candidate region is centered at y1_ SIFT is described by the color histogram \( h(y1_\text{ SIFT}) \).

After this apply mean shift to the new frame in parallel. Calculate a new centre position of Mean shift, y1_MeanShift, and the corresponding color histogram \( m_q \) (y1_MeanShift).

The bounding box that is associated with a large Bhattacharyya coefficient is selected as the new box centre y1 for the current frame.
If the coefficient is larger than a threshold specified previous by application requires (that is $\rho > T$), it is assumed that the target in the current frame is not occluded. The region of interest is updated for SIFT matching process for the following frames.

In general there are two parallel models used in this tracking method, one is Mean shift and the other is SIFT. Figure 1.1 describes entire process of combined method to track the face from the frame. The Mean shift model is based on color information from the reference frame (currently, from the first frame), and remains unchanged during the whole tracking process. For SIFT model, the reference model will be replaced if the Bhattacharyya coefficient surrounded the SIFT matched points is larger than a predefined threshold.

IV. RESULTS:

A number of experiments conducted to evaluate the proposed face tracking algorithm. Various experiments were performed to verify the validity of the combined method. The new method has been applied to track person face in the room and outdoor on the street who were occluded by different objects. The camera being fixed, further geometric constraints and also background subtraction can be exploited to improve the tracking process. In the examples shown in figure 1.2 the combined method successfully coped with partial occlusions of different colors, target scale variations and rotation.

A lot of parameters are utilized to quantify the execution of the proposed constant face detection and tracking framework. These parameters are Frame Per Second (FPS), which denotes the real-time framework speed or time proficiency, False Negatives (FN), which is the quantity of faces that have not been tracked, and False Positives (FP), which is the quantity of non-face objects that have been tracked.

![Results](image)

**Figure 1.2:** (a) Results after tracking faces from the frames 11, 74, 122 using Mean Shift; (b) Results after tracking faces from the frames 11, 74, 122 using SIFT; (c) Results after tracking faces from the frames 11, 74, 122 using proposed method.

The proposed face detection and tracking system is tested on real time captured videos. In order to lower the demand of computational power of the proposed system, the Normalized Pixel Difference (NPD) face detection algorithm is executed every 15th frame for face detection. The combined tracking is executed using the two feature extraction methods, SIFT and Mean Shift. Table 1 shows the system performance in terms of FN rate and FP rate when tested on real time videos. For experiment 7 FPS are considered for all three algorithms.

From Table 1, it is shown that using Mean Shift, the FP rate is 62%, and the FN rate is 57%. Using SIFT, the FP rate is 43%, and the FN rate is 41%. The system performance using combined Mean shift and SIFT, the FP rate is 36%, and the FN rate is 9%. When we compare all the three algorithms, combined algorithm is better than when using SIFT and Mean shift for face tracking.
Unconstrained Human face Tracking in live Video

Choose a region of interest on the initial frame and develop a probability distribution function.

Implement SIFT method to the next frame

Compute $y_1_{SIFT}$

Compute $h_u(y_1_{SIFT})$

Implement MeanShift method to the next frame

Compute $y_1_{MeanShift}$

Compute $m_u(y_1_{MeanShift})$

Compute both SIFT and MeanShift to the next frame

Region is updated for SIFT method for the frames

Region is updated for MeanShift method for the frames

If $\rho_{SIFT} \geq \rho_{MeanShift}$,

If $\rho_{MeanShift} \geq$ threshold,

NO

YES

Region is updated for SIFT method for the frames

Region is updated for MeanShift method for the frames

Figure 1.1: Process of combined method for face tracking
Table 1: The Performance of Proposed Face tracking System on

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Frame Size</th>
<th># of frames</th>
<th>FP rate (%)</th>
<th>FN rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Shift</td>
<td>1280*720</td>
<td>150</td>
<td>62</td>
<td>57</td>
</tr>
<tr>
<td>SIFT</td>
<td></td>
<td></td>
<td>43</td>
<td>41</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td>36</td>
<td>9</td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper we have explained the basics of visual object tracking and introduced some of the research work done in the field and five aspects of its important applications as well as the difficulties in visual object tracking and the dealing algorithms. A solution to enhance the performance of classical SIFT and mean shift object tracking has been presented in this paper. This work integrated the outcomes of SIFT feature correspondence and mean shift tracking. The approach applied a similarity measurement between two neighbouring frames in terms of color and SIFT correspondence. Finally, some experimental results of the integration of mean shift and SIFT feature tracking were presented. After the implementation, received results shows that the proposed method could produce better solutions in face tracking in different scenarios. Therefore this algorithm is considered as an effective visual face tracking algorithm.

REFERENCES


AUTHOR PROFILE

Ramkrishna B B received the BE and M.Tech degree from Visvesharaay University, Karnataka, India. He is currently working as professor in the department of computer science and engineering, CIT, Ponnampet, Karnataka. His Research interest are face detection and tracking, image and vision processing, computer networks. He published more than 10 research papers in reputed journals and conferences.

Dr.Sharmilakumari distinguished professor and Head in the Department of Computer science and Engineering at PA College of Engineering, Mangalore, Karnataka. She received BE from Mangalore University, M.Tech from NITK Surathkal and Ph.D from Magalore university, Karnataka. Her research interest includes image and vision processing, biometrics, text processing, face recognition. She published and presented many research papers in national and international conferences.