International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

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 anface 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.

Manuscript published on 30 September 2019
* Correspondence Author
Ramakrishna B B, Department of CSE, CIT, Ponnampet, Karnataka, India, Email: ramakrishu@gmail.com
M. SharmilaKumari,Department of CSE, PACE, Mangalore, Karnataka, India, Email: sharmilabp@gmail.com

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 facesshould 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 conceivableforward-tracking technique is Mean-shift method. In Mean shift method kernel function is defined tofind the distance between sample points and its mean shift, also the weightcoefficient is reverse with the distance. As the closer the distance is, the bigger theweight coefficient is.
Defined Mean shift algorithm is a non-parametric process [9]. Mean shiftofferstruelocalization and efficient matching without costlycomprehensive search. It is aniterative procedure, in other words, first calculate the mean shift value for the new pointposition, after that move the point to its mean shift value which is the new position, thencalculate the mean shift until it satisfy certain condition. The standard of Mean shiftprocedure 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 Epanechnikov$M_E(x)$kernel, and its kernel profile is

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

Where B_z is the volume of the unit z-dimensional sphere, $\{x_i\}$, (i=1,…,n) be anrandom 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 $\{\check{T}_u\}$ where u =1 2…m, m bins of histograms is derived from anelliptic region centered at y_0 , to remove the influence of the target scale it isnormalized to a unit circle, its pixel coordinates$\{z_i^*\}$.

$$\check{T}_u = \frac{\sum_{i-1}^{n}(1-|z_i^*|^2)\delta(c(z_i^*)-u)}{\sum_{i-1}^{n}(1-|z_i^*|^2)} \qquad (2)$$

where n is the number of pixels, δ is the Kronecker delta function:

$$\delta(z) = \begin{cases} 1, & z = 0 \\ 0, & otherwise \end{cases} \qquad (3)$$

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,…,n according. Then determine the new location of the target regionaccording to calculate the new probability value and determine that value. Onlyif the correspondence between the new targets region and thetarget mode is very less than that between the old target region and the model.

Carryout the remaining processes of this step toshift the target region half waybetween the new and old locationsand 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 object model in the current frame and then examines in the model's neighbourhood in next new frame, followed by calculatingfinest 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 invariantto 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 toorientation and scale.

(2) Keypoint localization: In a frame at each candidate location, a complete model is fit toidentify scale and location. Keypoints are chosen on basis of measures of theirstability.

(3) Orientation assignment: Multiple orientations are allocated to each key point location on basis of local image gradient directions. All imminentprocesses areperformed on image data that has been converted relative to the allocatedorientation, scaleand location for each feature, thus providing invariance to all thesetransformations.

(4) Generation of keypoint descriptors: At the selected scale local image gradients are measuredin the region nearby each keypoint. These selected gradients areconverted into a representation which acceptsmajor levels of local change inshape distortion and illumination.

III. COMBINED METHOD: SHIFT AND MEAN SHIFT

Proposed tracking algorithm is an effective integration using of Mean shift and SIFTfeature tracking. The proposed approach will apply asimilarity measurement between two neighbouring frames in terms of color and SIFTcorrespondence. Theoretically, a face track will be made if SIFT and Mean shift featuretracking lead to calculate probability distributions likecolor and intensity withinthe resultantarea in the next frame of the image (normally the two probabilitydistributions should be indistinguishable if the scenario does not change to considerably). Whenexpectation–maximization algorithm is applied in order to track a fulllikelihood estimate using the results from Mean shift and SIFTcorrespondence.

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 $m_q$in a rectangular (or ellipse) region centred at $y_0$and window size W.

$$\sum_{q-1}^{p} m_q = 1 \qquad (4)$$

In which $m_q$is the probability distribution of color q.

In the next step SIFT method is applied on the next frame. If there are manycorresponding points between the reference object region and a candidate region, anaffine matrix is estimated. Picking up the four corners ofthe rectangle model in the first frame, the affine matrix is used to transform thelocation of the old four corners to new ones, averaging its new positions. Whicheffectsin a new centre point *y*1_ *SIFT*. If there are no more sufficient pairs of matched pointsoriginate in the second frame, just average these positions and then obtain its *y*1_ *SIFT*. Thecandidate region cantered at *y*1_ *SIFT* is described by the color histogramh_u(*y*1_ *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 centre that is associated with a large Bhattacharyya coefficient is selected as the new box centre *y*1 for the current frame

(5)

$$y1 = \begin{cases} y1_{\text{SIFT}}, & if\ \rho_{SIFT} \geq \rho\_Meanshift \\ y1_{\text{Meanshift}}, & otherwise \end{cases}$$

If the coefficient is larger than a threshold specified previous by application requires (that is ρ > 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 fame. TheMean 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.

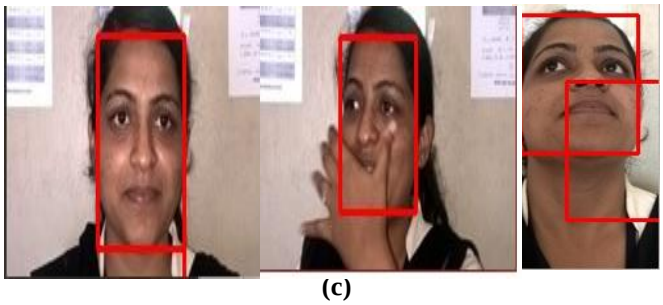
IV. RESULTS:

A number of experiments conducted to evaluate the proposed face tracking algorithm.Various experiments were performed to verify the validity of the combinedmethod. The new method has been applied to track person face in the room and outdoor onthe street who were occluded by different objects. The camera being fixed, furthergeometric constraints and also background subtraction can be exploited to improvethe tracking process. In the examples shown in figure1.2 the combined method successfully coped withpartial 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.

(a)

(b)



(c)

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.

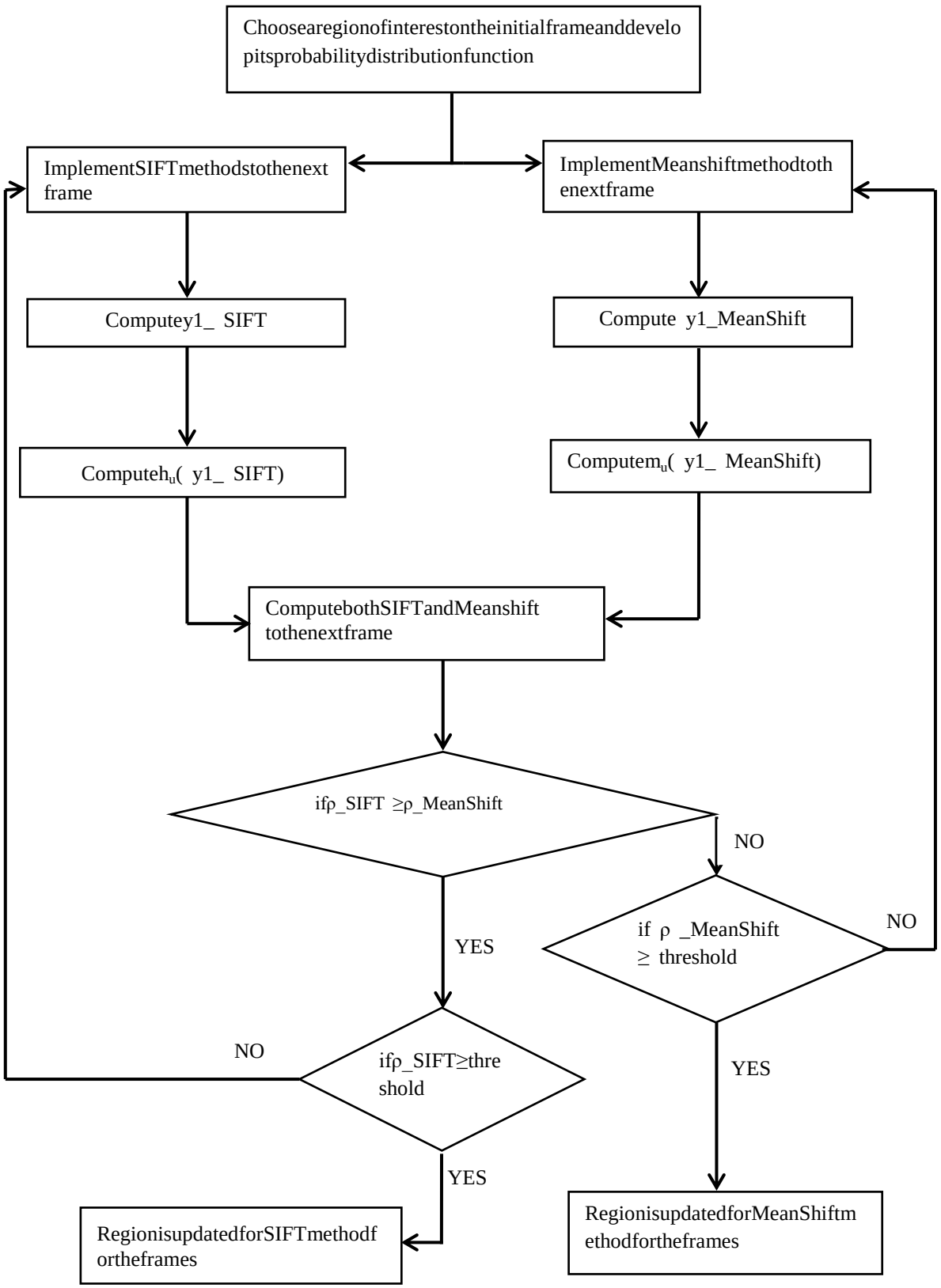


Figure 1.1: Process of combined method for face tracking

Table 1: The Performance of Proposed Face tracking System on

Algorithm	Frame Size	# of frames	FP rate (%)	FN rate (%)
Mean Shift	1280*720	150	62	57
SIFT			43	41
Combined			36	9

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

1. P.J. Butt and et al. Object tracking with a moving camera. In Proc. 2nd Int.Conf. on Computer Vision, pp. 2-12, 1989
2. A. Gyaourova, C. Kamath, S. and C. Cheung. Block matching object tracking. LLNLTechnical report, October, pp. 22 -26, 2003.
3. Y. Rosenberg, M. Werman. Real-Time Object Tracking from a Moving Video Camera:A software approach on PC. Applications of Computer Vision, WACV '98.Proceedings, pp.784-793,1998.
4. A. Turolla, L. Marchesotti and C.S. Regazzoni. Multiple camera object tracking invideo surveillance applications. Applications of Computer Vision. pp. 132–135,2001.
5. Y. Wang, J. Doherty and R. Van Dyck. Moving object tracking in video - Proc.Conference on Information Sciences and Systems, Princeton, NJ, pp.546-549,March 2000.
6. Çiˇ gdemEroˇ gluErdem and Bülent San - Video Object Tracking With Feedback ofPerformance Measures - IEEE Transactions on circuits and systems for videotechnology, vol. 13, no. 4, April 2003.
7. Probabilistic Object Tracking Using Multiple Features,DavidSerby, Esther-Koller-Meier, Luc Van Gool, Computer Vision Laboratory (BIWI), ETH Z¨urich,Switzerland,1998.
8. D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects usingmean shift, CVPR, pp. 142–149, 2000.
9. P. Perez, C. Hue, J. Vermaak, and M. Gangnet, Color-based probabilistic tracking,ECCV, pp. 661–675, 2002.
10. M. Isard, and A. Blake. Condensation – conditional density propagation for visualtracking, IJCV, 29(1):5–28, 1998.
11. M. Isard, and A. Blake, ICondensation: Unifying low-level and high-level tracking ina stochastic framework, ECCV, pp. 893–908, 2010.
12. G. David Lowe. Object recognition from local scale-invariant features. Proceedings ofthe International Conference on Computer Vision. 2. pp. 1150–1157,1997.
13. G. David Lowe. Distinctive image features from scale-invariant keypoints.International Journal of Computer Vision, 60(2), pp, 91-110, 2004.
14. Collins RT. Mean-Shift blob tracking through scale space. In: Danielle M, ed. IEEE Int'l Conf. on Computer Vision and Pattern Recognition, Vol 2. Baltimore: Victor Graphics, pp.234-240, 2003
15. P. Torr, Motion Segmentation and Outlier Detection, PhD thesis, Department ofEngineering Science, University of Oxford, 1995.
16. Y. Cui, S. Samarasekera, Q. Huang. Indoor Monitoring Via the CollaborationBetweena Peripheral Senson and a Foveal Sensor, IEEE Work-shop on VisualSurveillance, Bomba y, India, 2-9, 1998.
17. G. R. Bradski, Computer Vision Face T racking as a Component of a Perceptual UserInterface, IEEE Work. onApplic. Comp. Vis., Princeton, 214-219, 2013.
18. S.S. Intille, J.W. Davis, A.F. Bobick, Real-Time Closed-World T racking. IEEEConf. on Comp. Vis. and Pat. Rec., Puerto Rico, 697-703, 1997.
19. C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, Pfinder: Real-Time Tracking ofthe Human Body, IEEE Trans. Pattern Analysis Machine Intell, 19:780-785, 1997.
20. A. Eleftheriadis, A. Jacquin. Automatic Face Location Detection and Tracking forModel-Assisted Coding of Video Teleconference Sequences at Low Bit Rates,Signal Processing- Image Communication, 7(3): 231-248, 1995.

AUTHOR PROFILE

Ramakrishna B B received the BE and M.Tech degree from Visvesharaya Technological 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 Engineeringat 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.