Pose Estimation in Real Time Video using Tri-Line Method

A. Vanitha
Department of Electronics and Engineering
MIT Anna University
Chennai, India
Email: anbu.vanitha17@gmail.com

V. Vaidehi
School of Computing Science and Engineering
VIT
Chennai, India
Email: vaidehi.vijayakumar@vit.ac.in

***Abstract*—Human head pose estimation has become an important issue in the field of computer vision. Over the last decade, many approaches have been introduced to estimate head pose. However, estimating head pose in real-time has proven to be a difficult task. To overcome the drawbacks in existing system, this paper proposes Tri-Line based face pose estimation method for real time videos. The proposed method is faster and accurate compared to other existing head pose estimation methods and also updates the face poses dynamically. In Tri-Line method, pose estimation is calculated based on the distance between the locations of facial landmarks. The proposed Tri-Line method is tested in different databases like YouTube Celebrity database, NRC-IIT Face Video database and manually created Video database for both controlled and uncontrolled environment. Experimental results show that the accuracy of the proposed method is 90 (%) in yaw and 85(%) in pitch on Pointing'04 database.**

Keywords - Pose estimation; Detection; Facial Landmarks; Tri-Line.

I. Introduction

Nowadays many authorization systems fail due to password verification and personal identification number. To overcome this drawback face recognition system is used in many places for user authorization. Face recognition is a process of identifying a person by a machine using features. However to detect and recognize face in video surveillance system is very difficult under controlled and uncontrolled environment in the real time due to pose variations, expression variations, illumination variation and occlusion etc. Detection and recognition of face is an important issue in computer vision. Head pose estimation deals with inferring the orientation of human head in terms of X, Y, and Z Coordinates. Head pose can be classified into three Degree of Freedom namely, Yaw, Pitch and Roll as shown in Fig. 1 [14]. Pose estimation plays a vital role in face recognition in video as the face in the video is likely to move continuously. It is very critical to get the frontal pose of the face in a video in order to extract the feature for recognizing the face correctly.

Existing pose estimation methods need initialization and more training dataset for accurate classification. Hence, not suitable for real time pose estimation. To overcome these drawbacks this paper proposes a Tri-Line method for pose estimation which gives better accuracy compared to other state-of-art methods. This paper proposes Tri-Line method for pose estimation which calculates the face pose i.e yaw and pitch using distance measures. In this method, the distance between eyes, nose and mouth feature point locations gives the orientation of the face pose. As this method is simple and accurate it's found to be suitable for real time face recognition in video.

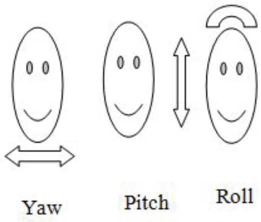

Fig. 1: Degree of Freedom for Pose Estimation

This paper is organized as follows: Section II presents the literature survey, Section III explains the proposed Tri-Line method, Section IV presents the implementation details and Section V discusses the results and Section VI presents the Conclusion.

II. Literature Review

This section gives the review of the state of art algorithms for pose estimation.

Xiaozheng Mou et al. proposed head pose estimation using depth data with Discriminative Random Regression Forest (DRRF). DRRF extracts patches from the query image and passed to the tree for head detection and orientation estimation. DRRF is very sensitive in live mode and also it requires more processing time [1].

De Marsico et al proposed Face Analysis for Commercial Entities (FACE). FACE discards the poor quality samples, uses a manual classification [2]. Correlation index measures the similarity of template image with query image, and classifier gives final decision about the query image.

Athi Narayanan et al proposed Cylindrical and Ellipsoidal Face Models for accurate yaw estimation [3]. Failure occurs due to re-initialization of candidate region during tracking.

978-1-5090-2541-1/16/$31.00 ©2016 IEEE

Tao Xu et al. proposed Saliency Model Based Head Pose Estimation by Sparse Optical Flow [4]. This method extracts feature point from face, also updates the feature points in successive frames for those locations where feature points are missing using sparse optical flow tracking method. The drawback of this method includes error due to initialization.

Gelareh Meydanipour et al. proposed Head Pose Estimation using ContourletSD Transform and GLCM (Gray-Level Co-occurrence Matrix) [5]. ContourletSD transform is applied on images and feature vector is created by GLCM and feature vector is classified using classification techniques.

Min Jiang et al. proposed pose estimation using Active Shape Model. This model is a template matching method, which is suitable for object localization and point based feature extraction [6]. But it fails to locate feature points in the face when the background is complex.

Fanelli et al. proposed DRRF collects depth images of head of different persons and different orientation for training and testing of head pose estimation. Moreover, the classification results fail for small number of images and also it recognizes non-head area as a head in complex background for some cases [7].

Wang et al. proposed Kalman filter with Active Appearance Model (AAM) for head pose estimation, which worked well on limited range of head poses [8]. However the failure can occur for increasing testing images and pose estimation limit.

Jiwen Lu, et al. proposed ordinary preserving manifold analysis approach for pose estimation. The similar samples are considered as single class and dissimilar samples are considered as different class [11]. Main disadvantage of this method is that, it requires more training data and also misclassification occurs due to less samples.

Ananth et al. proposed online sparse matrix Gaussian processes (OSMGP) based on observation of kernels for head orientation. OSMGP Cholesky matrix maintains and updates the training data [12]. It takes more time to maintain and update samples.

Though, they are several methods in pose estimation, there is a need for a simple and accurate pose estimation method which is suitable for video in order to increase face recognition accuracy. Hence, this paper proposes a simple and novel pose estimation using Tri-Line method in real time videos for both controlled and uncontrolled environment. The proposed method gives better accuracy in pose estimation in video under dynamic environment.

III. Proposed Tri-Line Method

The overall architecture of a pose estimation system in real time is shown in Fig. 2.

The proposed Tri-Line method is designed to estimate the pose of the face using seven landmarks and its x and y coordinates. The landmarks are left corner of right eye, right corner of left eye, left corner of mouth, right corner of mouth, right corner of right eye, left corner of left eye and nose tip. Yaw variation can be estimated by calculated the distance between left eye and nose (LC), right eye and nose (RC). LC (2) and RC (3) are calculated using right angle triangle formula. Using min-max function (1) & (4) yaw and pitch can be normalized in the range from 0 to 1. Fig. 3 shows the pose estimation from input video stream. Face is detected using Viola-Jones algorithm; facial landmarks are located using structured output SVM [9], [10]. Tri-Line method and face landmarks are used to estimate the face deviation from frontal to non-frontal.

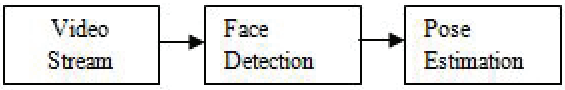

Fig. 2: Pose estimation overall Architecture

The yaw and pitch component using Tri-Line method for pose estimation is given below:

The yaw component for face pose estimation in real time video for both controlled and uncontrolled environment is defined as

$$yaw = (max(LC, RC) - min(LC, RC))/max(LC, RC) \tag{1}$$

Where,

$$LC = \sqrt{a_L^2 + b_L^2} \tag{2}$$

$$RC = \sqrt{a_R^2 + b_R^2} \tag{3}$$

a_L- Distance between left eye and nose with respect to y-coordinate
b_L- Distance between left eye and nose with respect to x-coordinate
a_R- Distance between right eye and nose with respect to y-coordinate
b_R- Distance between right eye and nose with respect to x-coordinate

Normalization of (1) produces result in the range of 0 to 1. Equation 2 and 3 show the right angle triangle formula. Equation 2 is used to estimate the accurate distance between left eye corner and nose tip, (3) estimates the distance between right eye corner and nose tip with respect to x and y coordinates. The values of the (2) and (3) are compared and yaw angle is classified as frontal or non-frontal. If LC is greater than RC, then the pose is oriented towards yaw right and vice versa.

Pitch is estimated using the difference between the y-coordinate of nose tip and left corner of right eye (referred as P_X) as shown in (5) and difference between the y-coordinate of nose and left corner of mouth (referred as P_Y) as shown in (6). If P_X is greater than P_Y, then the pose is oriented towards pitch down and vice versa.

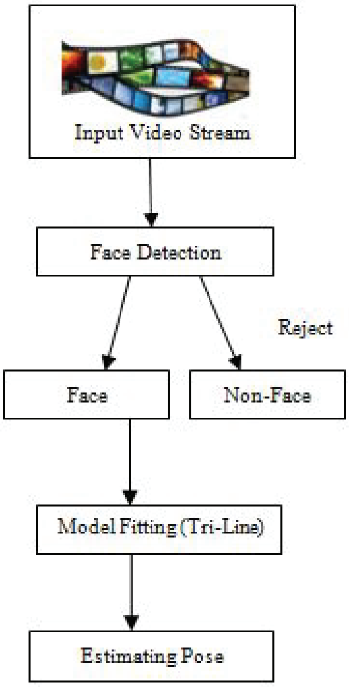

Fig. 3: Proposed Pose Estimation Architecture

The pitch component is defined as

$$pitch = (max(P_X, P_Y) - min(P_X, P_Y))/max(P_X, P_Y) \tag{4}$$

Where,

$$P_X = N_y - REL_y \tag{5}$$

$$P_Y = ML_y - N_y \tag{6}$$

N_y is Nose tip of y-coordinate,
REL_y is y-coordinate of Right Eye Left corner,
ML_y is y-coordinate of Mouth Left corner.

Face recognition in video gives maximum accuracy if the detected face has frontal view. The proposed Tri-Line method gives yaw and pitch value 0, then the detected face is frontal view.

IV. Experimental Results

The proposed Tri-Line method for pose estimation in real time videos for both indoor and outdoor environment is implemented in Open Source Computer Vision library (OpenCV) with C++ in Windows. Simulation testing is done in Intel Core i7-3632 processor system running with 2.5 GHz and 8 GB RAM.

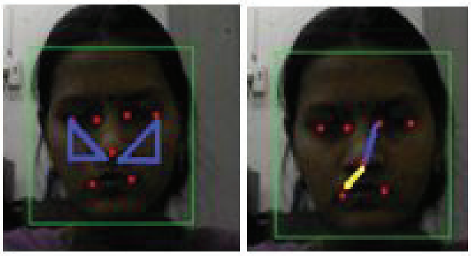

Fig. 4: Model for Pose Estimation

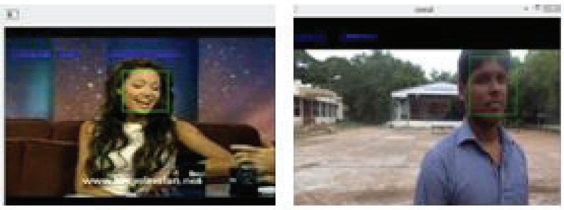

Fig. 5: Profile View (Yaw Left)

The proposed algorithm has been tested on publicly available benchmark databases like NRC-IIT Face Video database, YouTube celebrity video database and Pointing'04 database and also manually created video database for controlled and uncontrolled environment. Pointing'04 database consists of 15 sets of images. Each set contains 2 series of 93 images of the same person at different poses with vertical and horizontal angle, some persons wearing glasses and having various skin color. In vertical angle, person is looking bottom and top with negative and positive orientation. In horizontal angle, person is looking left and right with positive and negative orientation. The YouTube celebrity video dataset contains 3,425 videos of 1,595 different people and an average of 2.15 videos is available for each subject. The shortest clip duration is 48 frames, longest clip is 6,070 frames and the average length of a video clip is 181.3 frames. The NRC-IIT video database contains short video clips of users showing their face with different types of emotions while sitting in front of the monitor, the emotions are captured by an Intel Webcam mounted on computer monitor.

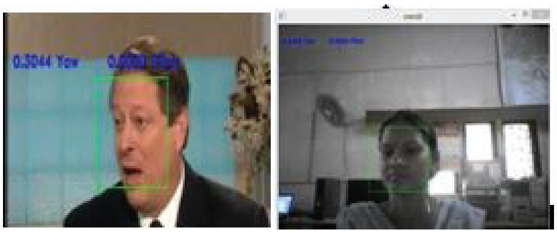

Fig. 6: Profile View (Yaw Left)

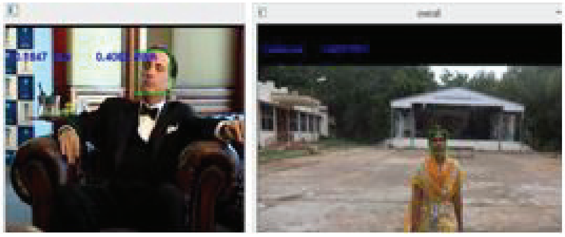

Fig. 7: Profile View (Yaw Left)

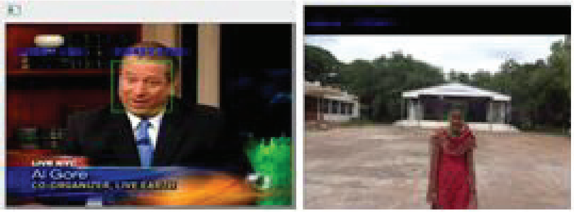

Fig. 8: Profile View (Yaw Left)

TABLE I: COMPARISON OF POSE ACCURACY WITH EXISTING WORKS

Methods	Accuracy in (%)	
	pitch	yaw
Existing Work in [13]	72.44	85.53
Existing Work in [14]	78.5	88.5
Tri-Line (proposed method)	85	90

TABLE II: POSE ESTIMATION ACCURACY FOR DIFFERENT BENCH-MARK DATABASES USING TRI-LINE METHOD

Different Databases	Accuracy in (%)	
	pitch	yaw
YouTube Celebrity Database	87	89.5
NRC-IIT face video database	88	92
Manual Dataset	93	95

TABLE III: POSE ESTIMATION F-SCORE FOR DIFFERENT BENCH-MARK DATABASES USING TRI-LINE METHOD

Different Databases	F-Score in (%)	
	pitch	yaw
YouTube Celebrity Database	90.9	94.3
NRC-IIT face video database	91.5	94.3
Manual Dataset	95.7	93.5
Pointing'04 Dataset	86.6	90.9

TABLE IV: POSE ESTIMATION PRECISION FOR DIFFERENT BENCH-MARK DATABASES USING TRI-LINE METHOD

Different Databases	Precision in (%)	
	pitch	yaw
YouTube Celebrity Database	91.3	85.1
NRC-IIT face video database	86.3	91.3
Manual Dataset	90.6	94.4
Pointing'04 Dataset	81.25	83.3

Fig. 4 shows the model for pose estimation in real time videos for both controlled and uncontrolled environment using right angle triangle formulation and straight line model. Fig. 5-8 shows the person looking from frontal to non-frontal pose.

Fig. 5 shows the result of the profile view for yaw left using manually created uncontrolled video database and NRC-IIT face video database. Fig. 6 shows the output of the yaw right for manually created uncontrolled video database and YouTube celebrity video database.

The results of the pitch up direction for manually created uncontrolled video database and YouTube celebrity video database is shown in the Fig. 7. Figure 8 shows the result of the profile view (i.e pitch down) using manually created uncontrolled video database and YouTube celebrity video database.

Table I shows the accuracy of the proposed Tri-Line method with existing methods applied to Pointing'04 database. The accuracy of different benchmark databases (i.e YouTube celebrity video database, NRC-IIT video database and manual dataset) are shown in Table II.

Table III and IV shows the result of precision and F-score value for different benchmark databases using Tri-Line method.

The proposed Tri-Line method gives better accuracy than the existing methods [13], [14] for Pointing'04 database.

Yaw estimation using Tri-Line method gives the accuracy of 90(%) whereas the existing methods [14], [13] give the accuracy of 88.5(%) and 85.53(%) respectively. The pitch component estimation of the proposed method gives the accuracy of 85(%), whereas existing methods [14], [13] give the accuracy of 78.5(%) and 72.44(%) respectively.

V. CONCLUSION

Tri-Line method for pose estimation in real time video has been proposed. The proposed system has three stages, first stage is face detection using Viola Jones algorithm, second stage is identifying the landmarks accurately and finally, a model is fitted in the face image for estimating the pose. The proposed method is tested with different benchmark video databases (YouTube video, NRC-IIT Face Video and Pointint'04). The accuracy of the proposed method is found to be 90(%) in Yaw and 85(%) in Pitch on Pointing'04 database. The proposed Tri-Line method is tested in different databases and the F-Score & Precision values are presented. In future the proposed method can be extended to address partial occlusion in complex background for indoor and outdoor environment.

ACKNOWLEDGMENT

This work is supported by DAE-BRNS, Mumbai. The authors would like to extend their sincere thanks to DAE-BRNS for funding the project.

REFERENCES

[1] X. Mou and H. Wang, "A fast and robust head pose estimation system based on depth data," Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on, Guangzhou, 2012, pp. 470-475.

[2] M. De Marsico, M. Nappi, D. Riccio and H. Wechsler, "Robust Face Recognition for Uncontrolled Pose and Illumination Changes," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 43, no. 1, pp. 149-163, Jan. 2013.

[3] A. Narayanan, R. M. Kaimal and K. Bijlani, "Yaw Estimation Using Cylindrical and Ellipsoidal Face Models," in IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 2308-2320, Oct. 2014.

[4] Tao Xu, C. Wang, Y. Wang and Z. Zhang, "Saliency model based head pose estimation by sparse optical flow," The First Asian Conference on Pattern Recognition, Beijing, 2011, pp. 575-579.

[5] G. Meydanipour and K. Faez, "Robust head pose estimation using contourletSD transform and GLCM," Machine Vision and Image Processing (MVIP), 2013 8th Iranian Conference on, Zanjan, 2013, pp. 375-380.

[6] M. Jiang, L. Deng, L. Zhang, J. Tang and C. Fan, "Head pose estimation based on Active Shape Model and Relevant Vector Machine," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, 2012, pp. 1035-1038.

[7] G. Fanelli, T. Weise, J. Gall and L. V. Gool, "Real Time Head Pose Estimation from Consumer Depth Cameras," in 33rd Annual Symp. of the German Association for Pattern Recognition (DAGM'11), Frankfurt, Germany, pp 101-110, 2011.

[8] Y. Wang, G. Liu, "Head Pose Estimation Based on Head Tracking and the Kalman Filter," in 2nd Int. Conf. on Physics Science and Technology (ICPST), Hong Kong, China, pp 420-427, 2011.

[9] A. Fernandez, R. Garcia and R. Usamentiaga "Glasses detection on real images based on robust alignment," Machine Vision and Applications 26.4, pp. 519-531, 2015.

[10] M. Uricar, V. Fran and V. Hlavac, "Detector of facial landmarks learned by the structured output SVM," In: VISAPP12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, vol. 1, pp. 547-556, 2012.

[11] J. Lu and Y. P. Tan, "Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation," in IEEE Transactions on Human-Machine Systems, vol. 43, no. 2, pp. 249-258, March 2013.

[12] A. Ranganathan, M. H. Yang and J Ho, "Online Sparse Gaussian Process Regression and Its Applications," in IEEE Transactions on Image Processing, vol. 20, no. 2, pp. 391-404, Feb. 2011.

[13] T. Wang, C. Hu, L. Gong, F. Liu, Q. Feng, "An effective head pose estimation approach using Lie Algebrized Gaussians based facerepresentation," Multimedia Tools and Applications, Springer, August 2013

[14] A. Fathima, V. Vaidehi, S. Vasuhi, M. Murali, SK. Parulkar, "Pose Invariant Face Detection in Video," Proceedings of the 2nd International Conference on Perception and Machine Intelligence. ACM, 2015.