World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:4, 2021

A Structural Support Vector Machine Approach for Biometric Recognition

Vishal Awasthi, Atul Kumar Agnihotri

Abstract—Face is a non-intrusive strong biometrics for identification of original and dummy facial by different artificial means. Face recognition is extremely important in the contexts of computer vision, psychology, surveillance, pattern recognition, neural network, content based video processing. The availability of a widespread face database is crucial to test the performance of these face recognition algorithms. The openly available face databases include face images with a wide range of poses, illumination, gestures and face occlusions but there is no dummy face database accessible in public domain. This paper presents a face detection algorithm based on the image segmentation in terms of distance from a fixed point and template matching methods. This proposed work is having the most appropriate number of nodal points resulting in most appropriate outcomes in terms of face recognition and detection. The time taken to identify and extract distinctive facial features is improved in the range of 90 to 110 sec. with the increment of efficiency by 3%.

Keywords—Face recognition, Principal Component Analysis, PCA, Linear Discriminant Analysis, LDA, Improved Support Vector Machine, iSVM, elastic bunch mapping technique.

I. Introduction

WITH the innovation of recent technologies, face recognition has rapidly expanded not only for engineers but also for neuroscientists. Face detection is a key part of the first step of automatic face recognition. It has lots of deviations of image appearance (i.e. pose variation), image orientation, occlusion, illuminating condition and facial expression. Many innovative methods have been proposed to resolve each variation such as the template-matching methods that are used for face localization and detection by computing the correlation of an input image to a standard face pattern. The feature invariant approaches are employed for feature detection of eyes, ears, nose, mouth etc. The appearance-based techniques are used for face detection with Eigen face neural network and biometric-based techniques have appeared as the most promising selection for recognizing individuals. These techniques examine an individual's physiological and/or behavioural characteristics in order to resolve and/or determine his/her identity. Biometric-based techniques include identification based on physiological features (such as voice, face, fingerprints, finger and hand geometry etc.) and behavioural characteristics. Face recognition appears to offer numerous advantages over other biometric methods such as fingerprinting or hand geometry detection and can be done reflexively without any explicit participation on the part of the user since face images can be obtained from a distance by a camera. This is particularly advantageous for security and surveillance purposes. Additionally, data acquisition in general is fraught with problem for other biometrics techniques that depend on hands and fingers. Iris and retina identification entail expensive equipment and are too sensitive to any body gesture. Voice recognition is vulnerable to background noises in public areas; however, facial images can be easily achieved with a couple of inexpensive fixed cameras. Appropriate pre-processing of the images and efficient face recognition algorithms can compensate for noise and slight variations in scale, illumination and orientation. Finally, technologies that entail multiple individuals to utilize the same equipment to capture their biological characteristics potentially expose the user to spread germs and impurities from other users. However, face recognition is totally non-intrusive and does not carry any health risks.

Jain et al. [1] explained biometric identification in networked security through facial recognition system. Kim [2] developed a theory on intelligent Immigration Control System by means of Passport Recognition and Face Verification based on passport checking on regular basis daily and his work was extended by Liu et al. [3] who utilized an Internet-based intelligent robot security system using invariant face recognition against intruder and plastic surgery [4]. The work proposed by Ahonen et al. [5] used Local binary pattern to signify the face. The method was applied on FERET dataset. The results showed that the proposed approach attained 97% accuracy for images with different facial expressions but did not perform efficiently for other factors such as image orientation and pose variation. Lei et al. [6] proposed an efficacious face detection system. The system worked in two modes: detection mode and the recognition mode. In the detection mode, they used AdaBoost algorithm to sustain a low computational cost whereas an improved independent component analysis approach was used in the recognition mode. Hausdorff distance was utilized in the recognition phase to estimate the similarity measure between the face and other objects present in an image. Phillips et al. [7] established a theory by researches named as The FERET Evaluation Methodology for Face Recognition Algorithms. In [8], a comparison among AdaBoost, PCA, LDA, and elastic bunch graph matching algorithms for face recognition has been illustrated with their drawbacks, benefits, success rate and other factors.

Luo et al. [9] described an unsupervised deep adaptation method. In their method, they adopted a deep unsupervised domain adaptation neural network and by using the training

Vishal Awasthi and Atul Kumar Agnihotri are with the Department of Electronics and Communication Engineering, University Institute of Engineering & Technology, C.S.J.M.University, Kanpur, India (e-mail: awasthiv@rediffmail.com, er.atulagni@gmail.com).

dataset for supervised learning methods, they reported an improvement of approximately 17% in the performance w.r.t. the results obtained by other face recognition algorithms. Acosta et al. [10] proposed an algorithm for automatic face detection and recognition in video indexing applications. Lee et al. [11] proposed a theory of video summarization and retrieval system using face recognition and MPEG-7 descriptors in image and video Retrieval. Recently in [12], prior to CNN, probabilistic decision based neural networks (PBDNN) have been utilized for facial recognition, eye localization and face detection.

In this paper a comparison of some texture and feature based methods has been performed and also holistic face algorithms on the dummy face data have been critically analyzed.

Other paper is organized as follows. Preliminary background of different face recognition methods is briefly d escribed in Section II. Section III describes the database formation for dummy face recognition. The efficient face detection algorithm with its components is discussed in Section IV. Section V describes the proposed algorithm for biometric physiological recognition system using nodal point selection. Performance analysis and simulation results of proposed algorithm with existing algorithm are demonstrated in Section VI. Finally a quick discussion and conclusions are given in Section VII.

II. Face Recognition Methods

The key factors that degrade the accuracy of the face recognition systems are primarily: occlusions, low resolution, noise, illumination, pose variation, expressions, ageing and plastic surgery [4].

1) Principal Component Analysis (PCA): The central idea of principal component analysis is to decrease the dimensionality of a data set by transforming to a new set of non-correlated principal components i.e. variables. Computation of the principal components diminishes to the result of an eigenvalue-eigenvector problem for a positive-semi definite ymmetric matrix. One of the most primitive such attempts was done by Acosta et al. [10], who used several image processing methods to extract a vector of facial parameters which were ratios of distances, areas and angles. A simple Euclidean distance measures to achieve a high performance of 75% on a database of 20 unlike people using 2 images per person (one for reference and one for testing).

Ahonen et al. [5], building upon Acosta's approach, figured a vector of 35 geometric features from a database of 47 people (4 images per person) and reported a 90% recognition rate. Furthermore, a certain tolerance must be specified to the models since they can never perfectly incorporate the structures in the image. On the other hand, the use of a higher tolerance value tends to destroy the precision required to distinguish individuals on the basis of the model's final best-fit parameters and hence these techniques become insensitive to the minute variations required for recognition. In general, existing algorithms for automatic feature extraction do not grant a high degree of accuracy and require considerable computational capacity.

The key advantage of this technique is that it can reduce the requirement of data to recognize the individual to 1/1000th of the data presented. The basis vectors are computed from the set of training images (I). The average image in training images is evaluated and subtracted from the images under training and generates the set of data samples $i_1, i_2, i_3, \ldots\ldots, i_{n-1}$ in matrix form.

Typically, only the N eigenvectors related with the largest Eigen values are utilized to define the subspace, where N is the desired subspace dimensionality. Knowing the weights of the training images and a new test face image, the nearest neighbour approach resolves the identity of the face.

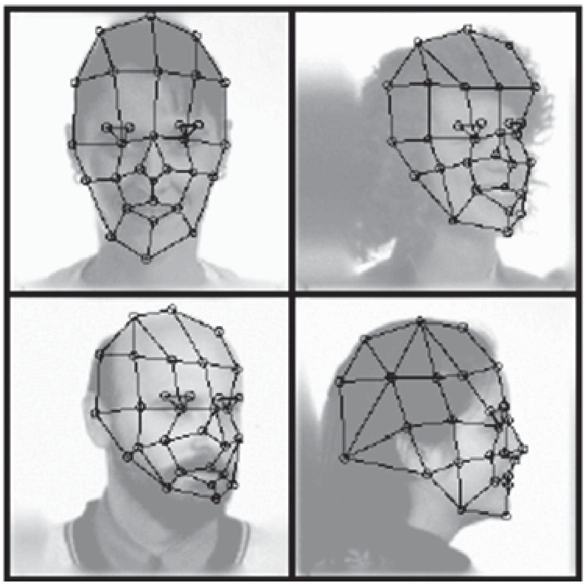

Fig. 1 Grids for face recognition

2) Linear Discriminant Analysis (LDA): Linear Discriminant Analysis (LDA) is a well-known classification technique that has been used significantly in many statistical pattern recognition problems. The approach of the LDA is to transfer all the data points into lower dimension new space. The objective of this technique is to maximize the class variance across users and minimize the class variance within user. In these techniques a block corresponds to a class, and there are large variations between blocks. Each class will structure a single dimensional Gaussian through the axis and the means will be well apart. It searches for those vectors in underlying space that pre-eminently discriminates to all the classes (rather than those that best depict the data). More formally specified a number of independent features relative to which the data is described. LDA produces a linear combination of these features which yields the largest mean difference between desire classes.

The unique t-dimensional space is projected onto an intermediary g-dimension space using PCA and then final f-dimension space LDA. Its main limitation is the inherent assumption that the true covariance matrices of each class are the similar.

3) Improved Support Vector Machine (iSVM): The structural support vector machine technique is a discriminative algorithm that allocates flexible feature construction with robust control over fitted data through structured output

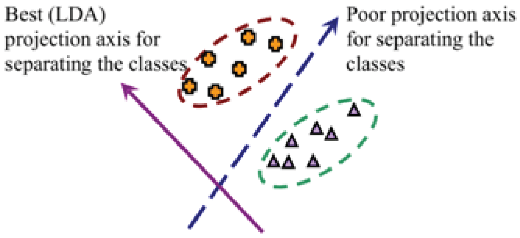

Fig. 2 LDA analysis

learning process. It gives state of art prediction accuracies for various structured output prediction tasks present in natural language processing, computational biology, and information recovery. First as a discriminative learning algorithm it permits flexible construction of features to enhance prediction accuracy and secondly it gives us the freedom to construct rich features from the input through discriminative modelling and improves prediction accuracy. Improved SVM controls over fitting features by maximizing the margin between alternative outputs like binary support vector machine.

Regularization is particularly significant when we have high dimensional feature space and small training sample sizes. Binary SVMs work by designing a large-margin hyper plane in high dimensional feature space as a separator between positive and negative points. In addition to latent variables, the utilization of nonlinear kernels in structural SVMs can also improve their articulateness and prediction accuracies. However their high computational cost during training restricts their wider application.

4) Elastic bunch mapping technique: This mapping technique is based on dynamic link structures in which a graph for an individual face is generated and a set of fiducial points on the face are selected. Each fiducial point corresponds to a node of a full connected graph and is characterized by the Gabor filters responses applied to a fiducial point window. Each arch is characterized by the distance between the corresponding fiducial points. A symbolic set of such graphs is united into a stack-like structure, designated as a face bunch graph. Graphs for new face images can then be generated automatically if the system has created a face bunch graph through elastic bunch graph matching. Recognition of a new face image is achieved by comparing its image graph to all the known face images and selecting the one with the highest similarity value. With this architecture, the recognition rate can attain 98% for the first rank and 99% for the first 10 ranks using a group of 250 individuals. The system has been improved further to deal with different poses besides the recognition performance on faces of the same orientation should remain the same. Several researchers have experimented with this technique, where they have removed the requirement of graph placement manually by utilizing parametric models which automatically locate fiducial points.

III. Database Description

Generally the dummy face images do not pursue the strictly controlled benchmark protocol of database acquirement because these images are located at various public places where any controlling restrictions cannot be imposed while acquisition.

Usually people capture outdoor photographs of various subjects (for example, say, 10 images per subject from random positions for pose variation) as shown in figure. A 12.2 megapixels, 5x optical image stabilized camera has been used for the data acquisition and images have been captured at a distance of 24 cm (approx.) from dummy faces in an unrestrained environment. They set the camera at the approximated angles which maintains the angles between the poses by $\theta = X/R$ radians as shown in the figure, where X is the arc size and R is approximated distance of camera from dummy face. Therefore the captured images are natural images without imposition of any restrictions neither on the targeted objects nor their surroundings.

A. Pose Variations

The spot photographs under uncontrolled surroundings have been taken under arbitrary condition and due to the effectiveness of proposed algorithm it automatically saves image in database as per the necessary condition required for face detection and recognition.

Fig. 3 Pose Variations

B. Camera Positioning for Pose Variation

To align the face image, the captured images have been rotated up to certain degree and then as pre-processing steps we cropped out only the dummy faces from the dynamic pictures ousting the background. Finally all cropped dummy face images have regularized to set all the subjects at standard gray level illumination with same size. Co-variation of illumination with pose is a real challenge in face recognition system. Dummy face images are captured during day time in outdoor atmosphere, but are altered by the changes in weather condition such as shadow diminishing of dummy faces due to extreme light. Moreover, excessive lighting can produce too bright images, which can change the automatic recognition process.

In recent works the face modelling, normalization, pre-processing, and invariant features extraction techniques have been developed to resolve the illumination problem up to the certain level.

In first modelling step, the illumination plane subtraction with histogram equalization has been performed and in second step, an illumination plane dummy $f(l,m)$ of an image dummy (l,m) corresponds to the best-fit plane from the image intensities that have been generated. Dummy $f(l,m)$ is a linear estimation of dummy (l,m) and is given by:

$$Dummy f(l,m) = al + bm + c \quad (1)$$

where coefficients a, b and c are expressed in multiple linear regression depending on the independence of model terms. When terms are inter-related and column of design matrix have an approximate linear dependence, the matrix becomes close to singular. Hence linear estimation becomes extremely sensitive to random errors in the observed response l, generating a large variance and a situation of multi co-linearity can occur.

In this paper normalization and pre-processing algorithms have been considered for the compensation of illumination because it doesn not require any training.

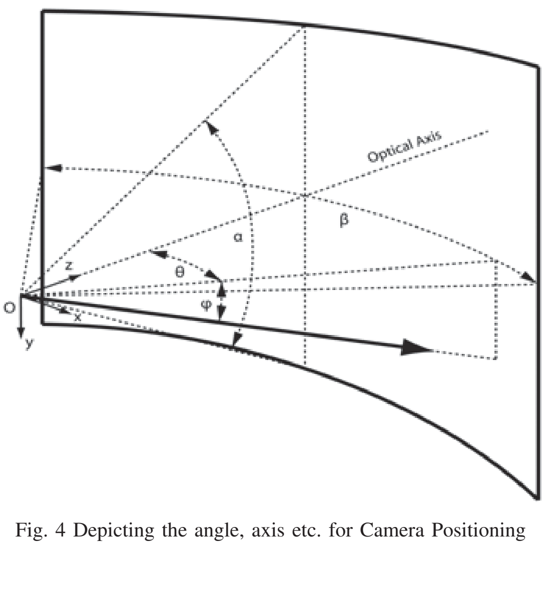

Fig. 4 Depicting the angle, axis etc. for Camera Positioning

IV. Face Detection

Skin colour detection in coloured images is very essential for face detection. Various techniques have been reported for tracing skin colour regions in the input image.

A. Color Segmentation

Generally the input colour image is characteristically in the RGB format. The techniques usually employ colour components in the colour space through HSV or YIQ formats. The RGB components are normally depends on the lighting conditions and subsequently the face detection might fail if the lighting condition varies. Among different colour spaces, YC_bC_r components are used to save the computation time. In the YC_bC_r colour space, the luminance information is enclosed in Y component and the chrominance information is in C_b and C_r respectively. Hence, the luminance information can be easily separated. The formula to convert RGB components to YC_bC_r components is given by:

$$Y = 0.299R + 0.587G + 0.114B \quad (2)$$

$$C_b = -0.169R - 0.332G + 0.500B \quad (3)$$

$$C_r = 0.500R - 0.419G - 0.081B \quad (4)$$

In the skin colour detection procedure, each pixel was classified on the basis of skin or non-skin in its colour components. The detection window for skin colour was decided based on the mean and standard deviation of C_b and C_r component, acquired using 164 training faces in 7 input images as shown in Fig. 5.

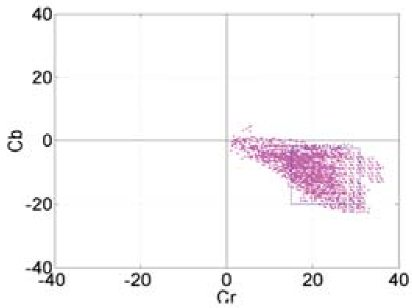

Fig. 5 Skin pixel in YC_bC_r color space

The colour segmentation technique has been applied on a training image and its result is shown in Fig. 6. Several non-skin objects are inevitably examined in the result as their colours fall into the skin colour space.

Fig. 6 Color segmentation result of a training image

B. Image Segmentation

The next step is to separate out the image blobs present in the colour filtered binary image into individual regions. This process comprises of three basic steps. Firstly the black isolated holes in image are filled up and white isolated regions are removed which are smaller than the least face area in training images. The threshold (170 pixels) is set conventionally. After this step, the filtered image followed by preliminary erosion, leaves the white regions with sensible areas only.

In the second step, the integrated regions present into individual faces are separated by using the Roberts Cross Edge detection algorithm. The Roberts Cross Operator carries out a straightforward, fast to compute, 2-D spatial gradient measurement on an image. It thus highlights areas of high spatial gradients that often correspond to edges. The highlighted portion is converted into black lines and eroded to unite crossly separated pixels.

Fig. 7 Edges detected by the Roberts cross operator

Finally in the last step, the previous images are joined together into one binary image and relatively small black and white areas are omitted. The key advantage of this process is that the edges connected to black areas remains unaltered even after filtering.

C. Image Matching

For performing image matching following steps should be performed:

1) Image generation & Building Eigen image database: Initially a set of Eigen images was generated using test images (for example: 106 test images), which were manually divided & grouped in to 7 test images and then edited to extract the exact location of faces with a square shape. The cropped test images were changed into gray scale, and then Eigen images were determined through those 106 test images. The highest 10 Eigen images in terms of their energy densities have been obtained, in order to achieve a generalized shape of a face. To save computing time, the information of Eigen images was compressed into one image which was obtained after averaging the first 9 Eigen images apart from the Eigen image 1 containing the highest energy concentration. Because of excessive energy concentration, the first image was excluded as it would have removed the details of face shapes that can be expressed from other Eigen images. The accumulated Eigen images were normalized through dividing the image matrix by its second norm and thus the effect of Eigen image size does not change the face detection algorithm.

2) Test image selection: After the process of colour-based segmentation, skin-coloured area can be taken separately. For this binary image, a set of small test images is required to be selected and forwarded to the image matching algorithm for the advance process. The results of image selection are exclusively depends on the colour information. A square box was considered and applied on each section of image with the quantified window size to meet the standard size of a face.

The faces are separated its upper and neck part by means of the erosion process. To combine these two separated parts into one area, box-merge algorithm was applied which simply combines two or more adjacent square boxes into one. Since most of the time this phenomenon takes place between face upper and neck part so the distance threshold was set large for vertical direction and small for horizontal direction respectively.

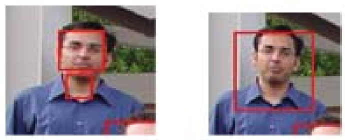

Fig. 8 Test Image Selection: (a) Merging of Adjacent Boxes before merging process

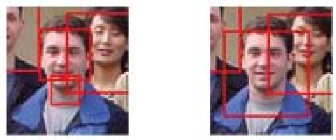

Fig. 9 Test Image Selection: (b) Merging of Adjacent Boxes after merging process

Fig. 10 Test Image Selection after Applying Box-Merge Algorithm

3) Correlation: For normalization selected test images are passed through the image matching algorithm. The image matching can be achieved by loading the corresponding data of Eigen image from the database images and then performing correlation between test image and the loaded Eigen image. The number indicated inside each window corresponds to the rank of the correlation value.

Since the figure to be analyzed will be a group image, faces are located near each other in the central area whereas the hands, arms, or legs are comparatively located distant from the faces as in Fig. 10. Hence, the mean square distance of a test image can be calculated w.r.t. other test images, and then to determine the geographical information of image component, the reciprocal mean square distance can be multiplied with the correlation value obtained primarily. A test image will achieve higher correlation value when it is located close to the other test images, while will have lower correlation value when it is far from the other group respectively.

D. Images Using Statical Information

The subsequent stage is filtering out non-facial test pictures. It was difficult to determine an absolute threshold value which can be applied to different images with various light condition and piece. After geographical consideration, by using statistical method the correlation values yield wide distribution of the output values. Among the group of the test images, the column which is having smallest correlation values will correspond to the left most columns in the test images. Out of 21 faces in the image, the algorithm has identified 19 faces within considerable error margin of the location of faces. The two unobserved faces are partly blocked by the other faces.

To detect the skin colour in colour segmentation process, generally a rectangular window was applied while the original distribution was a cone shape. As a consequence, some of actual skin colour was rejected and alternately some of non-skin tone was incorporated. More specific skin colour detection is needed if the window shape is nearer to the actual distribution, for example, triangle. In several cases, during the process of edge integration in image segmentation the redundant noises were added. To overcome this unexpected noise normally the Sobel cross filter or the Prewitt filter with pre-rejection of small clutters were identified which works more efficiently in some training images. Depending on the colour segmentation, skin-coloured areas were taken separately into small and squared test images and then after Eigen image matching process, the test images can be rearranged from the most-similar facial piece to the least-similar facial piece. At this point, an absolute condition or threshold is required to separate out the non-facial images. The correlation ranking after geographical consideration of the test image is shown in Fig. 11.

Fig. 11 Correlation Ranking after Geographical Consideration

V. Proposed Work Description

Face recognition systems identifies people by their face images using structured prediction rules learning along with the structural support vector machines (SVMs). As the first process of feature-based approaches, we have to recognize and extract the distinctive facial features of the input image such as fiducial marks other than the eyes, mouth, nose, etc. and then establish the geometric correlation among those facial points. After converting the input facial image to a vector of geometric features, a standard statistical pattern recognition technique is applied to match faces by means of these observations and measurements.

In this paper we proposed a latent structural SVM algorithm, which permits the use of latent variable structure to solve the problem of associated non-convex optimization. This system first incorporates the latent variables of face images in to testing images and then speeding up the training of structural SVMs through approximate cutting plane model with non-linear kernels. This results to an improvement in prediction accuracies with new alternative problem formulations using latent variables.

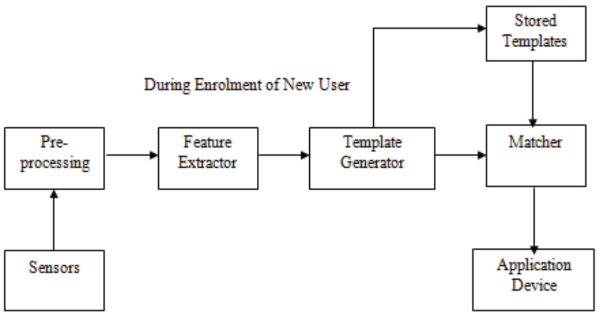

Fig. 12 Block diagram for the Face Recognition Process

The detailed block diagram demonstrates the complete process of how a picture normally identified and overview of work is shown in Figs. 12 and 13 respectively.

VI. Simulation Results

In this work a theoretical analysis with their iteration complexity in face recognition and their approximation is provided. The analysis has been done and compared with several state-of-art approximation algorithms on the basis of speed, time and efficiency to recognize the precise face with an improved accuracy.

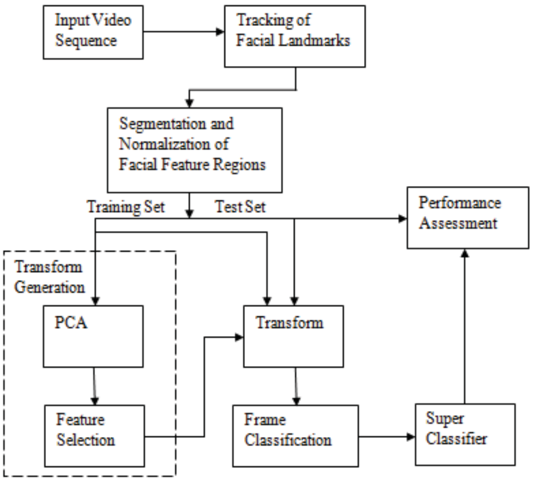

Fig. 13 Proposed Feature Based Approach using PCA

A. Speed & Time

As per the work done by Lee et al. [11], the speed of face detection improved and case timing is in the range of 90-110 seconds. In comparison to Phillips et al. [7], the proposed algorithm recognizes a person's identity within 30 seconds (except when the information in the database is not too large to analyze and computing).

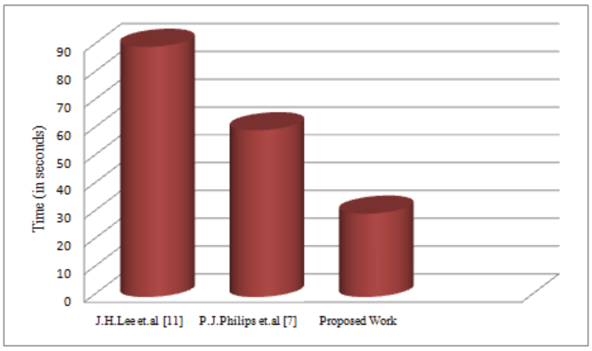

Fig. 14 For worst case timing

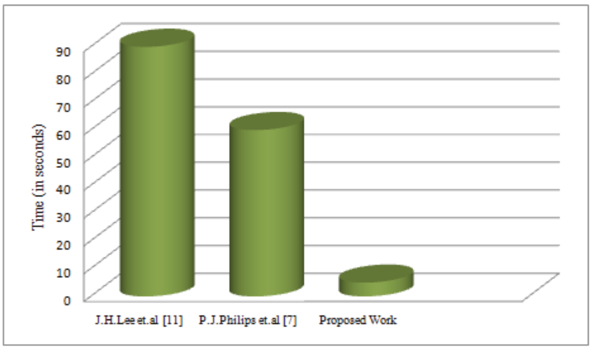

Fig. 15 For average case timing

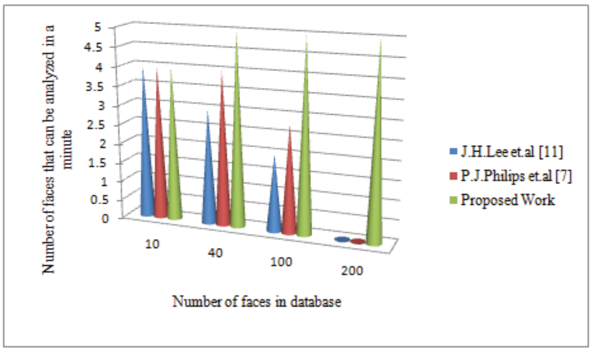

Fig. 16 Speed analysis

B. Efficiency

The efficiency of this designed system of facial recognition is improved by 1.2% to 3% as compared to Phillips et al. [7] and Lee et al. [11] algorithm respectively. This is so, because we considered a total of 80 nodal points instead of 50-60 nodal points.

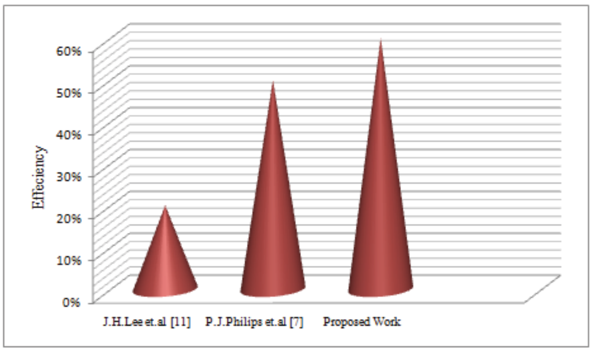

Fig. 17 Efficiency for worst case

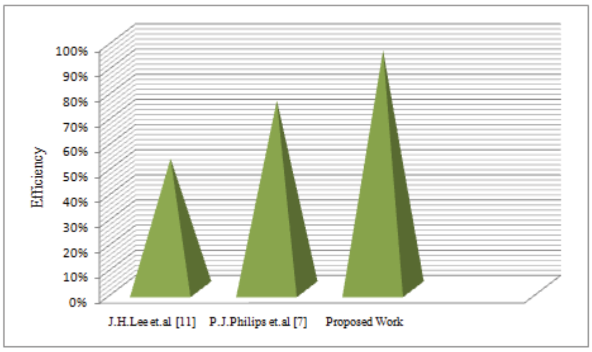

Fig. 18 Efficiency for an average case

VII. Conclusion

Face recognition and detection has lots of discrepancy of image appearance, like variation in pose, image orientation, illumination and facial expression. PCA technique is selected to recognize and extract distinctive facial features. The analysis is done on the basis of speed, time and efficiency and it was found that the propose algorithm improved the speed of face detection in the range of 90 to 110 seconds (recognition rate within 30 seconds) with efficiency improvement of 1.2% to 3% as compared to Phillips et al. [7] and Lee et al. [11] algorithm respectively.

References

[1] A. K. Jain, R. Bolle, and S. Pankanti, "Biometrics: Personal Identification in Networked Security," A. K. Jain, R. Bolle, and S. Pankanti, Eds.: Kluwer Academic Publishers, 1999.

[2] K. Kim, "Intelligent Immigration Control System by Using Passport Recognition and Face Verification," in International Symposium on Neural Networks. Changing, China, 2005, pp.147-156.

[3] J. N. K. Liu, M. Wang, and B. Feng, "iBotGuard: an Internet-based intelligent robot security system using invariant face recognition against intruder," IEEE Transactions on Systems Man And Cybernetics Part C-Applications And Reviews, Vol.35, pp.97-105, 2005.

[4] Sharif,M., Naz, F., Yasmin, M., Shahid, M.A., Rehman, A.: Face recognition: a survey. J. Eng. Sci. Technol. Rev. 10(2), 166–177 (2017)

[5] Ahonen, T., Hadid, A., Pietika, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2037–2041 (2006)

[6] Lei, Z., Wang, C., Wang, Q., Huang, Y.: Real-time face detection and recognition for video surveillance applications. In: 2009 World Congress on Computer Science and Information Engineering Real-time, pp. 168–172 (2009)

[7] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi, "The FERET Evaluation Methodology for Face Recognition Algorithms," IEEE Transactions onPattern Analysis and Machine Intelligence, Vol.22, pp.1090-1104, 2000.

[8] Chawla, D., Trivedi, M.C.: A comparative study on face detection techniques for security surveillance. In: Advances in Computer and Computational Sciences, pp. 531–541 (2018)

[9] Luo, Z., Hu, J., Deng,W., Shen, H.: Deep unsupervised domain adaptation for face recognition.In: 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 453–457 (2018)

[10] E. Acosta, L. Torres, A. Albiol, and E. J. Delp, "An automatic face detection and recognition system for video indexing applications," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.4. Orlando, Florida, 2002, pp.3644-3647.

[11] J.H. Lee and W.Y. Kim, "Video Summarization and Retrieval System Using Face Recognition and MPEG-7 Descriptors," in Image and Video Retrieval, Vol.3115, Lecture Notes in Computer Science: Springer Berlin / Heidelberg, 2004, pp.179-188.

[12] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," Neurocomputing, vol. 323, no. 61176031, pp. 37–51, 2019.

Dr. Vishal Awasthi received his B.E. and M. Tech. degree in the field of Electronics & Communication Engineering from Mumbai University and HBTI, Kanpur in 1999 and 2007 respectively. He has completed his Ph.D. in the field of VLSI-Digital Signal Processing. His area of interest is Digital Signal Processing & Control system. Presently he is working as Chairperson, Academic Performance Evaluation Committee (APEC), U.I.E.T., C.S.J.M. University, Kanpur (UP), India.

Mr. Atul Kumar Agnihotri received his B.Tech. and M. Tech. degree in the field of Electronics Instrumentation & Control from U.P.Technical University (UPTU), Lucknow in 2006 and 2014 respectively. His area of interest is Signal Processing & Wireless sensor network.