Real time Face Recognition and Detection for Industrial Application

MUHAMMAD USMAN GHANI KHAN[1,2*], ALI FAROOQ[1], OMER IRSHAD[1,2]

[1]*Al-Khawarizmi Institute of Computer Sciences*
[2]*Department of Computer Science and Engineering, University of Engineering and Technology Lahore, Pakistan*
**Corresponding author's e-mail: usman.ghani@uet.edu.pk*

Abstract

In the recent era, face recognition has procured the most important application of image understanding and analysis which is quite evident by some of the recent researches. Computer vision, pattern recognition and biometrics are emerging field for the researchers in the recent past with the focus in the area of the human recognition. Face recognition is an important factor keeping different issues such as robotics, security, attendance system. This paper concentrates on the industrial utility of Computer Vision techniques, namely face detection and recognition module. The present study mainly focuses on the Real Time Face Recognition System (RTFRS) and its applications. This scenario provides the opportunity to treat the face as a biometric identifier, that the person using information about the person or other biological organism, and, that how to deal this identifier for different services.

Keywords: Real time Face recognition; Face detection; Computer Vision technology; Pattern Recognition; Artificial Intelligence

INTRODUCTION

Humans have been using different characteristics to recognize each other through eyes, face, fingerprints, voice and gait since ancient times. After the emergence of the Computer Science, the replication of human perception is being one of the most growing tasks. Researchers and Scientists have been working to recreate the human intelligence. This generation is known as the generation of Artificial Intelligence. A lot of work has been done to enable human's interaction with the electronic devices and to make those devices sensitive so that they can respond back to the human. This area of artificially intelligent systems allows us to create such computer aided security systems [1]. Biometric systems are basically automated system to detect the human's presence. This can be done on the basis of their behavior and their psyche [2]. As described earlier, this paper is divided into two parts, first face detection and recognition and then its application

The Real time face detection and recognition [3] module mainly takes the raw image as an input afterwards face detection using haar cascade classifier is done [4], then extracted facial coordinates are sent to the pre-processing module where orientation of the image is processed and at last it is sent to the face recognition module which is linked with the database and returns the identity of the sample as shown in Figure 1. This methodology is explained in detail in forthcoming sections.

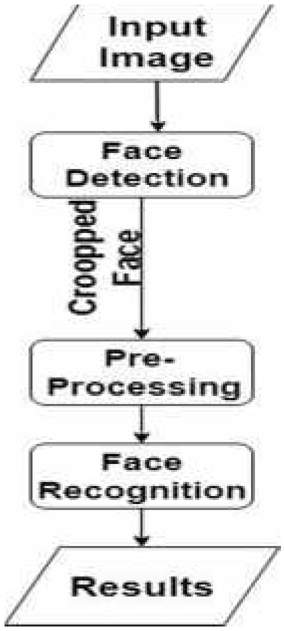

Figure 1: Face Detection System and Recognition

Literature Survey

Wiskott *et al*. [5] presented human face recognition system using the labeled graph comparison. This system produced graphs of concise facial descriptions in which descriptive fiducial points for describing face like eyes, mouth, nose, were illustrated by group of wavelet components. The construction of bunch graph was achieved by utilizing graph images from samples. Recognition was done through straight-forward comparison. The system was tested on FERET database alongwith self-generated Bochum database and achieved satisfactory results.

Ahonen *et al*. [6] proposed a novel approach in face recognition in which the images were split into small patches that are further employed for computing Local Binary Pattern (LBP) to describe texture information. Face recognition was achieved using nearest neighbor classifier on extracted features. This system claimed 79% accuracy which was more than the state-of-the-art approaches like PCA, BIC and EBGM on FERET database.

Zhu and Ramanan [7] used the pose and landmark estimation to implement a face detection system. They used tree-structured models which was more convenient to optimize than dense graph structure. They used self-collected and annotated dataset and achieved the accuracy of 76.7 % to detect the face from the frontal (45 to 45) side and the partial side.

Yang *et al*. [8] presented Principal Component Analysis in two-dimensional (2D-PCA) model for image representation. This model was based on image matrices of two-dimensions for feature extraction in which a covariance matrix for the image is established directly utilizing the actual photo matrices, furthermore its eigenvectors were extracted for feature computation from image. The proposed system was evaluated on three different face datasets: ORL, AR and Yale face dataset. It was claimed that the throughput of face recognition was higher in case of 2DPCA rather than PCA.

Wright *et al*. [9] presented an algorithm for recognizing the potential objects in the image of a scene. This algorithm asserts that feature selection process can be improved by considering the sparsity in the problem of facial identification. Unusual representations such as resized photographs and random projections were extracted by choosing accurate threshold, which was computed on the basis of sparse representation theory. This framework can handle problems because of occluded region and corruption quite efficiently. They performed different experiments on globally accessible databases and achieved 95% accuracy.

He *et al*. [10] proposed a face recognition method based on appearance features by employing Locality Preserving Projections (LPP) and named their approach the Laplacian-Face Approach. They generated face subspace from face images for further analysis. LPP keeps unchanged local features and attains a face subspace for deciphering manifold structure. A comparison was made among Laplacian-face approach, Fisher-Face and Eigenface methods on three multiple face datasets and better accuracy and low errors in recognition were pro claimed.

Tan and Triggs [11] proposed combination of methodologies for reliable face recognition under un-restrained illumination circumstances. They combined illumination normalization, filter-based feature computation, multiple feature merging, texture oriented facial features and distance transform based matching. Specially, they introduced local ternary patterns (LTP), principal component analysis (PCA) for computing features and Gabor wavelets based local representation cues. It was proposed that feature set alone is less accurate than its combination. The resulting system achieved state-of-the-art results on three distinct datasets by reaching a facial verification rate of 88.1.

Neto *et al*. [12] used Kinect sensor for an RBG-D image especially, after acquiring the image, Histogram of oriented gradients (HOG) and Principal component analysis (PCA) was applied for face recognition with a combination of K-NN used as a classifier by Euclidean distance metric and at-last 3-D audio was used at the detected position. They used Kinect-based dataset. The developed system achieved 94.26% accuracy.

Kumar *et al*. [13] detected the face from the live camera and extracted features by using open-cv and passed these features to Neural Network to train the model. After summarizing these features, they were stored in the database. The developed system was controlled by a cell-phone while accuracy of 90.0% was achieved.

Rehman *et al*. [14] captured the image, then preprocessing was applied on the required image and Bayesian classifier for face skin detection was employed. Calculated histogram using Gray Level Co-occurrence matrix (GLCM) and extracted features by Haralick classifier and Local binary pattern (LBP). 85.39 % of accuracy in Ethnicity was achieved.

Khan *et al*. [15] input the image to framework, pre-processing was applied on the captured image and then facial feature extraction and facial landmarks point was done. The optical flow of these feature points were tracked and passed to the classifier to detect the facial expression. Self-generated dataset was used and accuracy of 93.3% was achieved.

Ding and Tao [17] recognized face even from the blurred video stream. First, the face was detected, and then its image was restored to the original dimension by using the Trunk-Branch Ensemble CNN model (TBE-CNN). Video-based face (VFR) dataset was used and accuracy of 94.96% was achieved.

Methodology

There are a series of steps which are followed in real time face recognition and detection. A system is created which takes an image as an input, that is simple raw image of any type (.png,.jpg). Haar classification is applied on that image for the detection of eyes and mouth which returned the reduced image of the sample. This reduced image is then passed to the pre-processing module where image rescaling, resizing and bit-reduction is accomplished i.e. converted into gray-scale. Afterwards image is passed to the face recognition module on which different algorithms are employed e.g. Eigen faces, Fisher faces, LBPH and, CNN and then the sample is identified as shown in Figure 2.

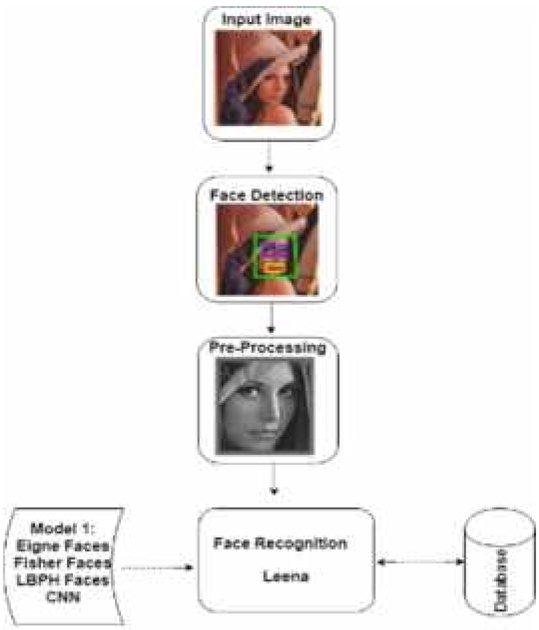

Figure 2: System Implementation

Face Identification using Haar Classification: Viola and Jones *et al.* proposed a method of Haar like features [13]. It sums up the intensities of the neighboring periodic areas at a specific location and calculates the variation between the sums which is used to categorically divide the image in the sub sections. Rectangular features of an image are calculated via some intermediate expression. This representation is called integral image. For face detection, the local area of eyes is darker than the other part of the face, same is true for lips [5]. If original image is f(x;y), then integral of an image is f[i(x; y)] that can be computed as shown below (Equation 1):

$$f_I\,(x; y) = P_x 0x; x0x0 - j_y - y_0 jA(x_0; y_0) \quad (1)$$

Face Recognition Module: For face recognition, the prerequisite is face detection. The detected model is sent to the face recognition model and then identity is matched with the database, if the sample is identified then its result is shown, else it ignores the sample and waits for the new sample, as shown in the Figure 3. As discussed earlier, for face detection, Haar Cascade is applied. For validation, below points are followed in the order.

1. *ROI of image is set for face detection.*
2. *Eyes and mouth are detected using eyes cascade and mouth cascade.*
3. *If any mouth is detected in upper portion of face, it is discarded.*
4. *If more than one mouth are detected, the largest one is kept.*
5. *If any eye is detected in the lower portion of face, it is discarded.*
6. *If more than 2 eyes are detected, eyes having the largest distance between them w.r.t x-axis, the left most eye and the right most eye rectangle are kept while other eye rectangle are discarded.*
7. *At this point, all the rectangles of mouth more than one and all the rectangles of eyes more than two are discarded. If eyes are less than two and mouth is less than one then the faces are discarded.*
8. *Above steps are repeated for all detected faces.*
9. *After that weight patterns or Eigen-faces are updated, both are optional.*
10. *If the same face appeared more than x times then its features are calculated and it was added to the face space optional.*

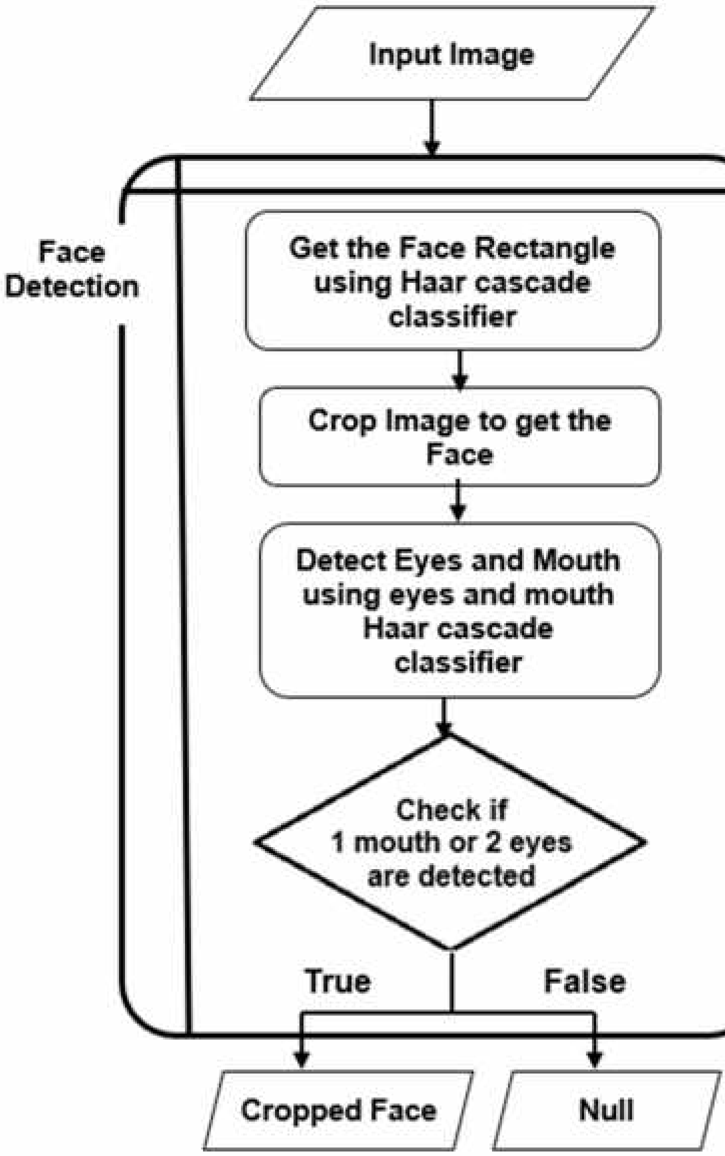

Figure 3: Face Detection Module

Eigen Faces Algorithm: Face is mainly composed of the following features set eyes, nose, hair and lips. In information theory, we need to get the pertinent knowledge in a face image and then judging according to the local database we own [18]. The face recognition methods consist of the following initialization steps:

1. *The face images set for the training are acquired.*
2. *The Eigen Faces are computed on the basis of training images, M photos that represent the largest Eigen value are kept only. These representative images depict the face space. As current faces are experienced, the value of Eigen Faces can be modified or recalled.*

3. *Concordant distribution is computed in M dimensions for every known individual in the form of weight space by mapping their facial images onto the face space. A set of weights based on the input frame is calculated and the M Eigen Faces by projecting the input frame onto each of the Eigen Faces.*
4. *Figured out whether the input image contains a face after all (whether known or unknown) by confirming if the image is quite close to the faces space.*
5. *Classification of the weight pattern is made to determine whether it is a face, and of a known person or an unknown.*

Fisher Faces Algorithm (FLD): Fisher Linear Discriminant is also well familiar as Linear Discriminant Analysis (LDA) [19] which draws a difference between different classes as a large scatter and within classes as a small scatter [18] resulting into a very compact and well defined clusters. FLD is behind many face recognition methods like Eigen-faces [20],[21], discriminant analysis [22] and robust coding scheme [23]. Mostly techniques utilize PCA to reduce the dimensionality of high dimension image space of original input images. After that FLD transformation extracts further discriminant features (MDF) space for classification. It needs large training set for proper generalization that is its drawback, and generally there are large training faces for face recognition but per class they are less. And for that artificial generation of data is a way out.

Local Binary Patterns Histograms (LBPH): As a whole, Fisher-faces and Eigen-faces use a mathematical elaboration of the most prominent features of the training. LBPH overviews each face in the set of training data independently and separately. In LBPH every image is checked separately, while the Eigen-faces computes features on the complete dataset at the same time. The LBPH algorithm is somehow simpler, in the way that we distinguish every image in the dataset locally and when a different unknown image is given, similar analysis is performed on it and the results are compared in the data-set. The manner in which we localize the image patterns for every location f the image.

Convolutional Neural Network: Face recognition system includes local image sampling, then passing these samples self-organizing maps(SOM) for dimensionality reduction and then reduced set of feature scaled maps are passed to convolutional neural network to train the network [24]. Convolutional Networks architectures make the clear suppositions that the inputs are images. CNNs take advantage over ordinary neural network as the input comprises of images and the architecture is restricted in a more sensible way. Specially, unlike a regular Neural Network, the layers of a neural network which is convolutional in nature have neurons patterned in 3 dimensions: height, width, depth. Here depth can be taken as the augmented number of layers in a given network. Proposed convolutional neural network architecture comprises five layers and they work repeatedly for training of a neural network.

I. INPUT raw pixels of images.
II. CONV output that is coming from the neurons which are connected to local regions of the input.
III. RELU element wise activation function e.g. thresholding at zero.

IV. POOL perform down sampling along spatial dimensions.
V. FC (Fully Connected) computes class scores from the number of classes.

Our system of face recognition based on convolutional neural networks performs the following steps.

1. *The face image are taken as input for face detection and validation as discussed earlier.*
2. *They are resized to 256 x 256.*
3. *They are passed over by a window of 5 x 5 with a shift of 4 to get 64 x 64 a total of 4096, 5 x 5 matrices.*
4. *These 5x5 matrices are then flattened to pass through SOM, so finally three-dimensional array of 64 x 64 x 25 is generated for every image.*
5. *Each row of 25 samples is assigned a label by SOM. Therefore, the image are reduced from 256 x 256 to 64 x 64.*
6. *Then they are passed to Convolutional Neural Networks for training.*
7. *As a comparison of dimensionality reduction, CNN is trained on images directly i.e. without dimensionality reduction.*

For implementation, an online attendance system is created using the LBPH as described below: graphical user Interfaces include two windows forms: Admin and Client

- Admin module provides an interface for initial configuration, capture image from any of the two IP cameras i.e front and rear camera to get the samples of any person. The interface of admin module is shown in Figure 4. These samples will then be used for training the system for face recognition. Name and designation of a person is also required for entering samples to database. Another part of admin module is reporting. This application is also capable of generating reports of attendance record. Customized report is also supported based on name, date and designation

- Client Interface provides the visual information of front and rear cameras. Every 6th frame of live video is processed for face detection. The interface of client module is shown in Figure 5. If face is detected, then face recognition API is called for results based on image. Image captured is re sized and gray scaled before sending for recognition to get better results. For accuracy improvement, if same person is recognized in a 5-6 of frames only then attendance is marked. This system generates different type of reports.

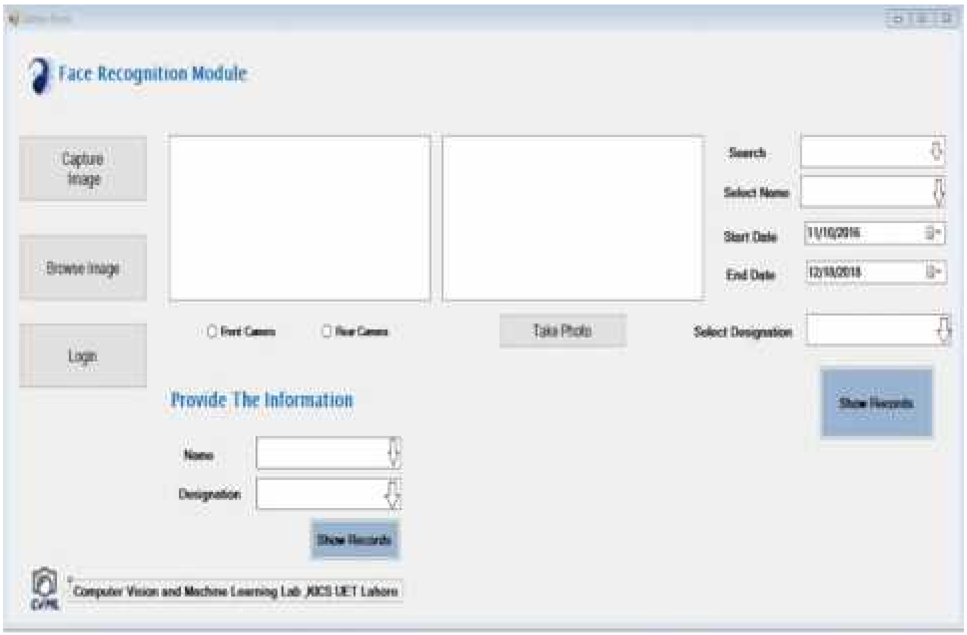

Figure 4: Admin Interface

Figure 5: Client Interface

Attendance Report: Attendance report is generated using crystal reports as shown in Figure 6. Crystal reports uses different formulas to display the results based on a query. Crystal reports relates two database tables and displays the person name across the login and logout details. Multiple filters can be applied to the reports which are provided in the admin interface. You can get the report of all persons from start date till today (all entries in the database). Filters based on name of the person, designation of the person and date are also available. Reports are available for printing as well as saving for future reference.

1/08/2016 Attendance Report

Al-Khwarizmi Institute of Computer Science

ID	Person Name	Designation	Login Time	Logout Time
1	Ayaan	Research Officer	1/08/2016 9.00 Am	1/08/2016 5.30 pm
11	DR.Usman	Ro	1/08/2016 9.00 Am	1/08/2016 7.15 Pm
24	Gulraiz	Ro	1/08/2016 9.05 Am	1/08/2016 5.15 Pm
57	Arooj	R o	1/08/2016 9.30 Am	

Figure 6: Attendance Report

Database: Amazon SQL Server database is currently being used by the software to maintain the attendance record, person's data and images path as well as the initial configurations of software. The architecture of database is shown in Figure 7. The database is located at cloud storage, so internet connection is needed to use this software. Twenty GB of space is allocated for this database by Amazon server.

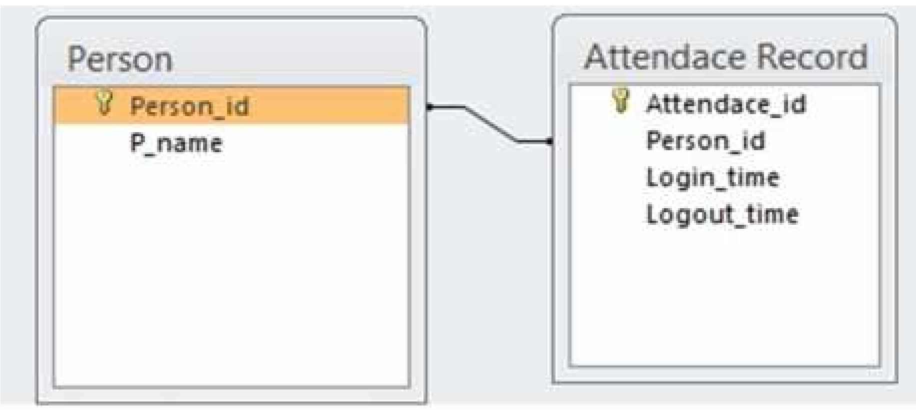

Figure 7: Data schema

Working: For initial configurations, admin module has to be run. Selection of path for storage of images and training is made. After the initial configuration, admin module's interface appears. <Capture> button can be clicked to start the live view from the camera. When the person's face is detected, a red rectangle is drawn on the person's face. <Take Picture> button is clicked to capture the image of that person. After entering Name and Designation of person, <Submit> button is clicked to submit the record. On same pattern, 5 samples may be taken. After that <Train> button is clicked to train the system with captured images. Then client module is run. Two views of cameras: one front and one rear appear over the interface for recognition. The source path of these cameras is stored into the database which can be edited later on based upon the requirements.

RESULTS AND DISCUSSIONS

SOM and CNN are computationally expensive algorithms because they require training recursively. Hidden layers, convolution and many other computations take more resources in case of hardware as well as time. The training time of CNN with 55 image as samples and 11 images per subject of 256x256 resolution, was around 30 minutes on the Intel core i5 processor. This CNN is restricted to 20 epochs only. The accuracy of CNN is 91 percent with 44 images for training and 11 images for testing. After increasing the number of samples to 200 samples per subject for 14 subjects, resulted in a total of 2800 images with same resolution of 256x256 and thus training time of CNN with these inputs increased from 30 minutes to 30 hours (Table 1). Although CNN is using more resources as compared to Eigen, Fisher and LBPH but results are outstanding as compared to these algorithms. The GPU therefore uses dedicated memory and are made for heavy computation; there will be huge difference in the training time of the system. Table 2 shows that CNN is giving the best precision, accuracy and F Score while the recall is dependent upon the false positive sample, but due to the limited data, at the moment we are not able to get the recall value improved. We make the confusion matrix for evaluating system as shown in Table 3.

Table 1: Details of methodologies

No of Images	Resolution	Methodology	Training Time (min)	Images / Subject	Accuracy (%)
1400	256 x 256	Eigen	10	200	85 (119 out of 140 correct)
1400	256 x 256	Fisher	15	200	77 (108 out of 140 correct)
1400	256 x 256	LBPH	10	200	85 (119 out of 140 correct)
55	256 x 256	CNN	30	11	91 (10 out of 11 correct)
55	256 x 256	SOM+CNN	3600 (60hrs)	11	54 (6 out of 11 correct)
2800	256 x 256	CNN	1800 (30hrs)	200	99 (555 out of 560 correct)
2800	256 x 256	SOM+CNN	In Progress	200	In Progress

Table 2: Measure of Relevance

Methodology	Precision	Recall	Accuracy	F Score
CNN	99.10	100	99	99.55
Eigen Faces	85	100	85	91.89
Fisher Faces	77.14	100	77.14	87.09
LBPH Faces	85	100	85	91.89

Table 3: Confusion Matrix

N=560	Predict No	Predict Yes
Actual No	3	2
Actual Yes	1	555

CONCLUSION

Eigen faces in live environment are not so successful because variation of angle of face images produces a lot of change in accuracy of Eigen faces, Fisher accuracy is not much for it to implement in live environment. As per observation, LBPH performs better in live environment as compared to Fisher and Eigen, because it is working on the basis of the locality of the sample and then calculating its probability. CNN uses more hardware resources for training, since it works in deep learning so it takes the data into layers, because of epochs. But its accuracy is far better than Eigen, BPH and Fisher. CNN is a better approach as compared to accuracy and prediction. Prediction time as per results of 560 images is 360 seconds almost half a second for CNN.

FUTURE WORK

The current module is working on the LBPH, as we are getting better and more accurate results we will move our interface to CNN as deep learning seems to be a positive field to dig out better and fast results. Moreover we have some flaws in our dataset too like we are unable to generate false positives and, ground trothing for enhancement in our data.

REFERENCES

[1] A. F. *et al.*, “3D Face Recognition in a Ambient Intelligence Environment Scenario,” in Face Recognition, I-Tech Education and Publishing, 2012.

[2] R. Bhatia, “Biometrics and Face Recognition Techniques,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 5, pp. 93–99, 2013.

[3] W.-Y. Zhao *et al.*, “Face Recognition: A Literature Survey,” *ACM Comput. Surv.*, vol. 35, pp. 399–458, 2003.

[4] P. Ian Wilson and D. John Fernandez, “Facial feature detection using Haar classifiers,” *J. Comput. Sci. Coll.*, vol. 21, 2006.

[5] L. Wiskott *et al.*, “Face recognition by elastic bunch graph matching,” in *Int. Conf. Comput. Anal. Images Patterns*, 1997, pp. 456–463.

[6] T. Ahonen *et al.*, “Face recognition with local binary patterns,” in *Eur. Conf. Comput. Vis.*, 2004, pp. 469–481.

[7] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2879–2886.

[8] J. Yang *et al.*, “Two-dimensional PCA: a new approach to appearance-based face representation and recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.

[9] J. Wright *et al.*, “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[10] X. He *et al.*, “Face recognition using laplacian faces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, 2005.

[11] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[12] L. B. Neto *et al.*, “A Kinect-based wearable face recognition system to aid visually impaired users,” *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 1, pp. 52–64, 2017.

[13] P. M. Kumar *et al.*, “Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things,” *Cluster Comput.*, pp. 1–12, 2017.

[14] A. Rehman *et al.*, “Modified Texture Features from Histogram and Gray Level Co-occurence Matrix of Facial Data for Ethnicity Detection,” in *5th Int. Multi-Topic ICT Conf.*, 2018, pp. 1–6.

[15] G. Khan *et al.*, “Geometric positions and optical flow based emotion detection using MLP and reduced dimensions,” *IET Image Process.*, vol. 13, no. 4, pp. 634–643, Mar. 2019.

[16] C. Ding and D. Tao, “Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 1002–1014, Apr. 2018.

[17] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition.* CVPR, 2001, vol. 1, pp. 511–518.

[18] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, 2002.

[19] R. A. Fisher, "The use of Multiple Measurements in Taxonomic Problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, Sep. 1936.

[20] P. N. Belhumeur *et al.*, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[21] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[22] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Am. A*, vol. 14, no. 8, p. 1724, Aug. 1997.

[23] C. Liu and H. Wechsler, "Robust coding schemes for indexing and retrieval from large face databases," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 132–137, 2000.

[24] S. Lawrence *et al.*, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.