JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FACE RECOGNITION SYSTEM USING MACHINE LEARNING ALGORITHMS (HAAR-CASCADE & PCA) AND THEIR EFFICIENCY COMPARISON

[1]Irfan Ayoub, [2]Dr. Bhawna Sharma,
[3]Er. Sheetal Gandotra, [4]Imtiyaz Hussain,[5]Neelesh Singh Katoch, [6]Manthan Kaushal
[1]Student, [2]Professor, [3]Associate Professor, [4]Student, [5]Student, [6]student
[1]Department of Computer Engineering,
[1]Govt. College of Engineering And Technology Chak Bhalwal, Jammu, India

***Abstract*— This paper mainly addresses the building of face recognition system by using Haar Cascade and Principal Component Analysis (PCA) and their percentage efficiency in different cases. Haar Cascade is a machine learning algorithm used in object detection and to determine the object in an image. It uses function called Haar wavelet to determine features called Haar features. In haar-features we find the difference in pixel intensities of the regions inside the detection window. Though PCA is a measurable methodology utilized for lessening the quantity of factors in face acknowledgment. Each picture in the preparation set is addressed as a straight blend of weighted eigen vectors. These eigen vectors are called Eigen faces. The weights are obtained by selecting most relevant vectors from covariance matrix. The recognition is done by comparing a projected test image with the subspace of Eigen faces. The minimum Elucidean distance is measured for classification of faces. Various investigations were performed to assess the presentation of the face recognition system. In this thesis, we used a training database of students of Computer Engineering department, Batch 2016, Govt. College of Engineering and Technology, Chak Bhalwal Jammu.**

***Index Terms*— Face detection, Face recognition, Pattern recognition, Haar Cascade, Cascade Classifier, Haar Features, Local Binary Patterns Histograms(LBPH), Principle component analysis (PCA), Histogram of Oriented Gradients(HOG), Percentage Efficiency Comparison .**

1. INTRODUCTION

Face recognition is one of the important challenges in appearance-based pattern recognition field. Face recognitionis widely used in many devices and applications as an identity verification mechanism. It can be used in numerous applications such as law enforcement, surveillance, public safety, personal identification etc. Due to this wide ranging applications and future potential of being a reliable identification system it has received attention from a lot of researchers. Face recognition systems face many challenges because of a large number of factors that affect the face recognition performance like illumination, pose, background noise, face expressions or occlusion.

The process of face recognition involves mainly two stages Face detection and Face recognition. First the test image (image to be recognized) undergoes face detection so that the face is extracted from a large image. Then the image is normalized so that the format of image matches the images in the database. Then the normalized face is used for recognition.

Face acknowledgment can regularly be utilized for confirmation or distinguishing proof. In confirmation an individual is as of now signed up for the reference data set or exhibition for example it is a coordinated matching errand while in ID, a test picture is coordinated with a biometric reference in the exhibition i.e. it represents a one-to-many problem. Face recognition involves four main phases: preprocessing phase, segmentation or localization, feature extraction phase and recognition phase.

The result of face recognition can be known face i.e. the face is recognized or unknown face i.e. the face is not recognized. Beside these results there can also be false results such as false reject which means that a known face isnot recognized or false accept which means that a unknown face is incorrectly identified as some known face.

The success achieved by the past work done on the face recognition resulted in the implementation of the system in wide variety of real time applications. The algorithms can bedivided into two main categories based on the different approaches they use: two dimensional (2D) approaches and three dimensional (3D) approaches. Mainly, the traditional 2D approaches are divided into six algorithms: eigenfaces (PCA), fisher faces or linear discriminant analysis (LDA)independent component analysis (ICA), support vector machine (SVM), neural network and hidden markov model (HMM). The 3D face recognition approaches uses a different technique. This procedure utilizes 3D sensors to catch data about the state of a face. It tends to be separated in to two principal classes: 3D face recreation and 3D posture assessment. Another methodology utilizes joins these to approaches is called Hybrid methodology.
In this paper the authors attempt to compare the performance of the PCA and Haar Cascade which results the algorithm having decreased computational time without greatly affect the recognition accuracy. Experiments are based on face database. The performance of the proposed algorithm is tested on PyCharm , and the obtained results show the most efficient and suitable in performance among the proposed algorithms. The rest of this paper is organized as follows: in Section 2 we give a background illustrating the popular face recognition algorithms Haar Cascade and PCA. Section 3, demonstrates the recent and related work in face recognition field. Section 4, we present our work. In Section 5 we display the conducted experiments; Section 6 demonstrates the experiment results, Section 7 Comparison and Section 8 Concludes this work.

1.1 HAAR CASCADE

Haar Cascade classifier utilizes Haar Wavelet method to dissect pixels in the image. The grayscale image comprises of predominantly dull and light pixels. The distinction in the pixel force decides the different Haar highlights in the picture. These elements are utilized to address most significant highlights of a human face. Haar overflow classifier depends on Viola Jones recognition calculation which is prepared in given a few information faces and non-faces and preparing a classifier which recognizes a face.
In automatic face recognition meaningful features from an image are extracted and put together in a useful representation and performing some classifications on them.The process can be divided in three major steps:

First, creating a good database of faces with multiple imagesfor each individual. The subsequent stage is to distinguish faces in the data set pictures and use them to prepare the face recognizer (the recognizer utilized here is Local Binary Patterns Histogram (LBPH)) and the last step is to test the face recognizer to perceive faces it was prepared for.

- **Import the required modules**

The Modules expected to play out the facial acknowledgment are cv2, os, picture module and numpy. CV2 is the OpenCV module and contains the capacities for face ID and affirmation. Operating system will be utilized to work with document index and pictures names and so forth. In the first place, we utilize this module to extricate the picture names in the data set catalog and afterward from these names individual number is removed, which is utilized as a mark for the face in that picture. The dataset of the pictures of a face is in BGR design and to permit OpenCV to work with the data set Image module from PIL is utilized to peruse the pictures in grayscale design. Numpy exhibits are used to store the photos.

- **Load the face detection Cascade**

The most vital phase in this cycle is to track down the face in the picture. From that point, face discovery overflow is stacked. The face is grouped and the district containing the face in the picture is found. When we get the area of interest containing the face in the picture, we use it for preparing the recognizer. With the end goal of face identification, we will utilize the Haar Cascade given by OpenCV. OpenCV establishment catalog

contains the haar overflow classifier. Overflow is stacked utilizing the cv2 Cascade Classifier capability which follows the way to the cascade xml record.

- **Create the Face Recognizer Object**

A face recognizer object is created to provide the functionality to train the recognizer and recognize a face (real time recognition). Face Recognizer train and Face Recognizer predict is used for training the recognizer and predicting/recognizing the face respectively. OpenCV provides recognizers such as Eigen face Recognizer, Fisher face Recognizer and Local Binary Patterns Histograms Face Recognizer. Here Local Binary Patterns Histograms Face Recognizer is used to perform face recognition. With Local Binary Patterns depicting the surface and state of a computerized image is conceivable. This is finished by separating a picture into a few little locales from which the elements are extricated that can be utilized to get an action for the similitude between the pictures.

- **Prepare the training set and Perform the training**

To make the capability to set up the preparation set, we will characterize a capability that follows the outright way to the picture information base as info contention and returns tuple of 2 list, one containing the identified countenances and the other containing the relating name for that face. For instance, if the ith index in the rundown of appearances addresses the fourth person in the data set, then, at that point, the relating ith area in the rundown of marks has esteem equivalent to 4.Presently to play out the planning or preparing using the Face Recognizer. Train capability. It requires 2 contentions, the elements which for this situation are the pictures of countenances and the relating marks relegated to these appearances which for this situation are the singular number that we extracted from the picture names.

2.0 Testing

For testing the label of the images in database i.e. the actual label is compared with the predicted label. This involves identifying if the results of recognition of images produced by Face Recognizer are desirable/correct or not. The actual label is extracted utilizing the string tasks from the name of the picture. The confidence score or percentage accuracy is also displayed so that the accuracy of the algorithm can be determined. Also it helps in better understanding the working of the algorithm in different conditions.

2.1 PRINCIPLE COMPONENT ANALYSIS(PCA) USING HOG

Real time face recognition is an extremely complex task because a large number of variables that needs to be processed collected from the image. A machine learning algorithm works more efficiently if the number of variables are within a certain limit and also if the redundant data is removed from database. This helps machine learning algorithms to learn better and produce more desirable results. Principle component analysis (PCA) is a statistical procedure that uses a symmetrical change to change over a bunch of perceptions of potentially connected factors (elements every one of which takes on different mathematical qualities) into a bunch of upsides of straightly uncorrelated factors called head parts.

The dimensionality of the large datasets are reduced by PCAusing vector space transform. The redundant data in theoriginal dataset can be removed by using mathematical project. A large number of correlated variables can be reduced to a few variables. Thus, using PCA the process of machine learning can be enhanced by allowing useful data and patterns and removing the redundant data.

For facial features Histogram of Oriented Gradient (HOG) features are used. HOG highlights are extremely famous in numerous grouping issues. The histogram of situated slopes (HOG) is an element descriptor utilized in computer vision and picture processing with the end goal of item identification. The HOG feature is that where it is combine with an SVM classifier to perform people detection. This technique is used to count the gradient orientation of pixels in localized portion of the image. The image is divided in to cells and the histogram of gradient orientation is compiled. The HOG descriptor is computed on the overlapping grid of cells.

The process of generating HOG Descriptor for an image is shown in Figure 1.

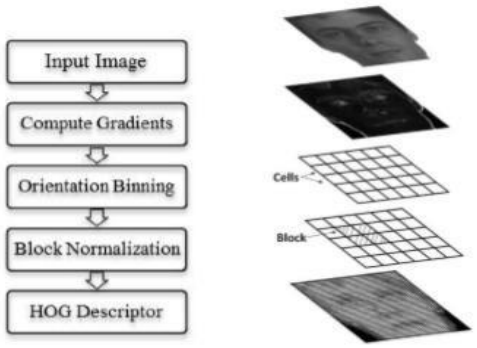

Figure 1. process of formation of histogram of oriented gradient features descriptor.

To obtain the HOG features, the gradient of the image is computed by applying an edge mask. A simple 1-D mask ofthe form [-1, 0, 1] works well. More complex masks, such as Sober operator and others, were tested in [6], but the simple, centered, 1-D derivative mask works the best. The resulting gradient image is divided into smaller no overlapping spatial regions, called cells. These cells can be rectangular, which results in the R-HOG descriptor, or circular, which results in the C-HOG descriptor, as shown in Figure 2.

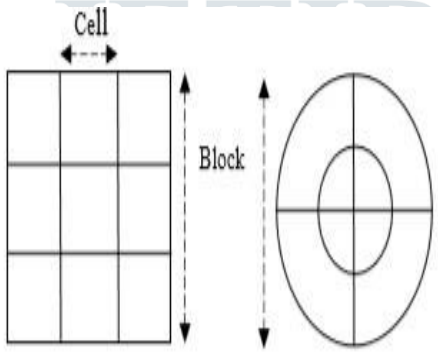

Figure 2. rectangular and circular cells used in histogram of oriented gradient.

The HOG descriptor has proven effective for classification purposes and has been applied on a variety of problems. Its main disadvantage is the associated computational load, and we utilize dimensionality reduction to ease the computational burden.

3.0 RECENT WORK

Many existing face acknowledgment explores use Haar Cascade and PCA for face acknowledgment, we show the absolute latest work.

There has been a gigantic measure of exploration on face acknowledgment innovation. Specialists have attempted and tried different which can be utilized to identify and perceive faces really. Consequently, making this innovation more helpful to in certifiable applications. Face identification is a PC innovation that decides the area and size of human face in erratic (digital) picture. The facial highlights are identified and some other items like trees, structures and bodies and so forth are disregarded from the advanced picture. The two sorts of approaches talked about before which are utilized to recognize facial part in the given picture are highlight base and picture base methodology. Highlight base methodology attempts to remove elements of the picture and match it against the information on the face highlights. While image base approach tries to get best match between training and testing images.

3.1 LOW LEVEL ANALYSIS:

Many existing face acknowledgment explores use Haar Cascade and PCA for face acknowledgment, we show the absolute latest work. There has been a gigantic measure of exploration on face acknowledgment innovation. Specialists have attempted and tried different which can be utilized to identify and perceive faces really. Consequently, making this innovation more helpful to in certifiable applications. Face identification is a PC innovation that decides the area and size of human face in erratic (digital) picture. The facial highlights are identified and some other items like trees, structures and bodies and so forth are disregarded from the

advanced picture. The two sorts of approaches talked about before which are utilized to recognize facial part in the given picture are highlight base and picture base methodology. Highlight base methodology attempts to remove elements of the picture and match it against the information on the face highlights.

Majorly three different face detection algorithms are available based on RGB, YCbCr, and HIS color space models. In the implementation of the algorithms there are three main steps viz.

(1) Classify the skin region in the color space,

(2) Apply threshold to mask the skin region and

(3) Draw bounding box to extract the face image.

Crowley and Coutaz recommended easiest skin variety calculations for recognizing skin pixels. The apparent human variety differs as a component of the general heading to the brightening. The pixels for skin locale can be distinguished utilizing a standardized variety histogram, and can be standardized for changes in power on partitioning by luminance. Changed over a [R, G, B] vector is changed over into a [r, g] vector of standardized variety which gives a quick method for skin discovery. This calculation falls flat when there are some more skin locale like legs, arms, and so on. Cahi and Ngan proposed skin variety characterization calculation with YCbCr variety space. Research found that pixels having a place with skin district having comparable Cb and Cr values.

4.0 RESEARCH METHODOLOGY

Our work aims to compare the performance of Haar Cascade and PCA by comparing their percentage efficiency and the time of computation. We conduct the experiments using PyCharm which ensures one value for face recognition. The experiment is used to compare the two algorithms by using the required number of images for each individual to be used in the training set that gives a highest percentage of recognition. It is expected that the highest matching ratio is made when light conditions are adequate and the image used for training the algorithm is taken from the same device that is being used for real time face recognition.

In PCA we test 28 persons in the test database with one images for each person in the training database .We change the exposure of light trying to make a decision of the best matching. In this experiment we observe the false acceptance rate for the algorithm so, we decrease the exposure of light to the images in the training set which results consequently increase in the percentage accuracy of the face recognition.

In Haar Cascade we test same persons creating live database with 30 images for each person in the training database .The algorithm detects and recognizes a single face very accurately. Although the match percentage of the LBPH(*Local Binary Patterns Histograms)* recognizer is not as high as PCA. The match percentage ranges from 60-70% in most cases. The results of this experiment give the same recognition results as the second experiment but with less percentage accuracy.

5.0 EXPERIMENTS

We conduct two experiments each answer a question that helps to complete our research**.**

5.1 The First Experiment: For PCA

Case1 : A single face in normal condition which results in accurate face matching with higher percentage accuracy that is 92% and 91% and we change the angle of face results 87% accuracy of the same image as shown in the screenshot below :

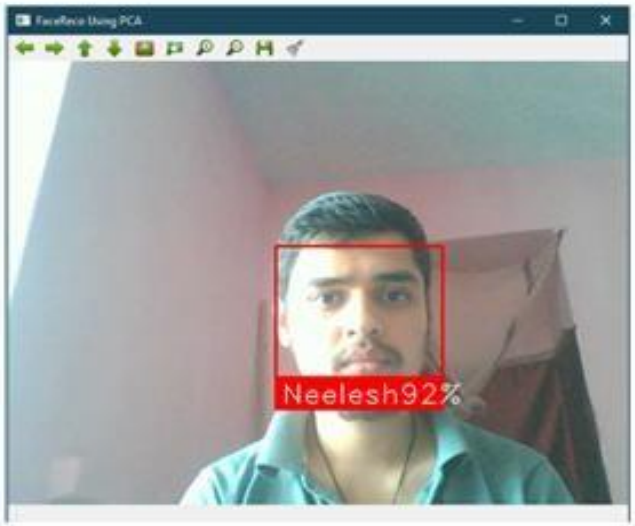

Figure 3. normal condition

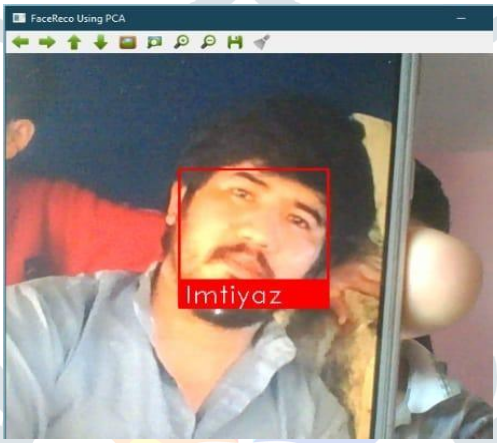

Figure 4. normal condition

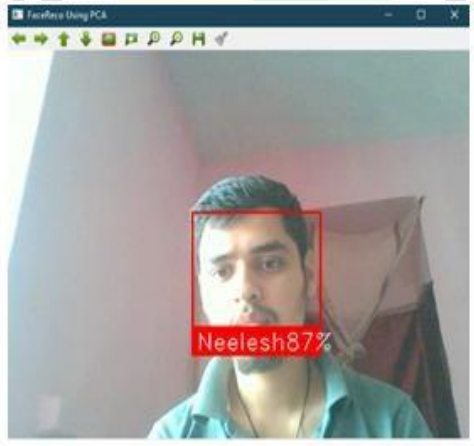

Figure 5. normal condition

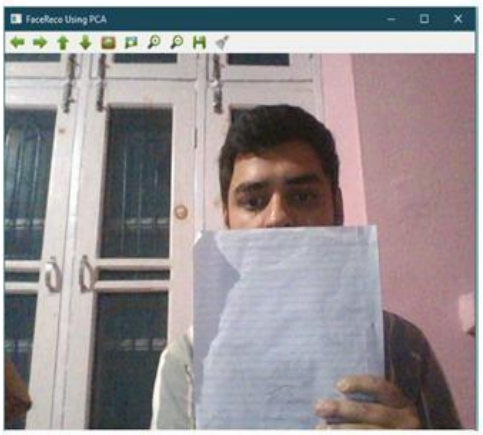

Figure 6. face angle changed

Case2 : A single face which is partially covered with a paper which results no face detection or recognition and when we allowed the exposure of face from paper it results in 85.4% accuracy of the image as shown in the screenshot below :

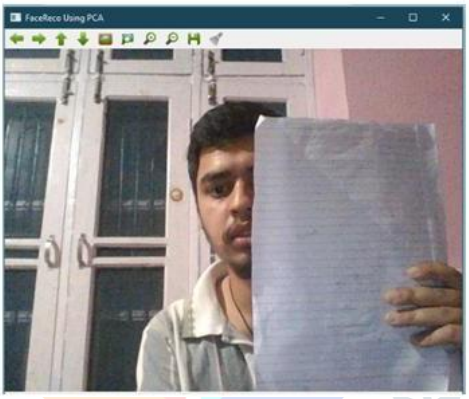

Figure 6. vertically partially covered

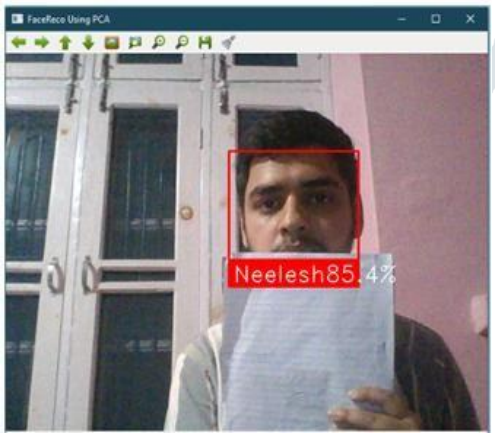

figure 7. horizontally partially covered

Case3 : Checking if the algorithm is capable of differentiating between drawing of human faces and actual human face which results no detection of drawing faces as shown in the screenshot below :

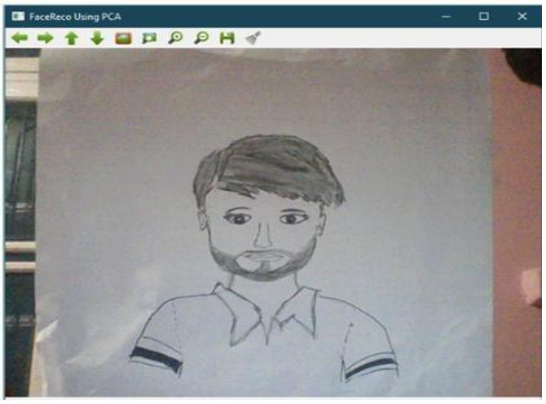

Figure 8. no detection of drawing faces

Case4 : Multiple faces to detect and recognize the Known and the Unknown faces which shows it can recognize multiple faces of a single frame as shown in the screenshotbelow :

Figure 8. multiple faces for detection and recognize

Figure 9. multiple faces for detection and recognize

Figure 10. multiple faces for detection and recognize

5.1 The Second Experiment: For Haar Cascade

Case1 : A single face in normal condition which results accurate face matching with percentage accuracy that is 70% 61% and 63% of the images which is comparatively very lesser than PCA as shown in the screenshot below :

figure 11. single faces to detect and recognize in normal condition

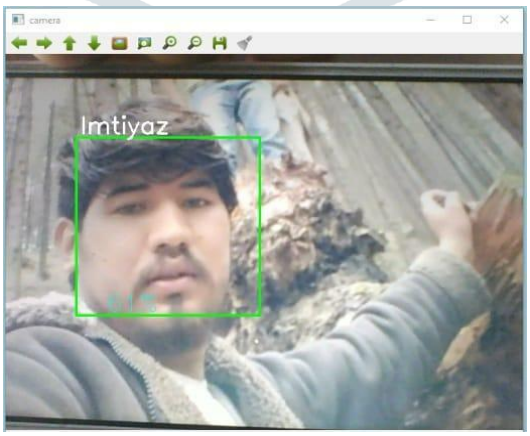

figure 12. single faces to detect and recognize in normal condition

Case2 : A single face for recognition which is partially covered with a paper which results in no face recognition but the algorithm is able to detect the face as shown in the screenshot below:

figure 13. single face partially covered

Case3 : Checking if the algorithm is capable of differentiating between drawing of human faces and actual human face which results in no detection of drawing faces as shown in the screenshot below :

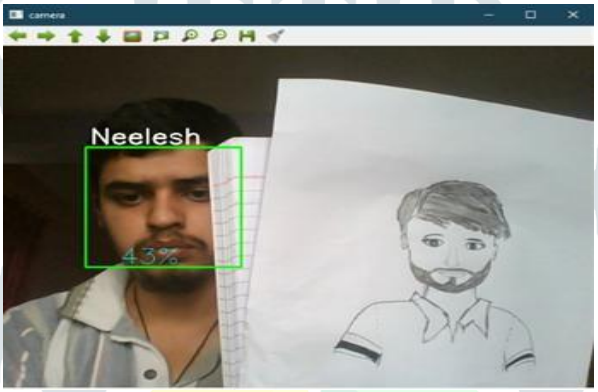

figure 14. no detection of drawing face

Case4 : Multiple faces to detect and recognize the Known and the Unknown faces which showa that it can recognize multiple faces in a single frame as shown in the screenshot below :

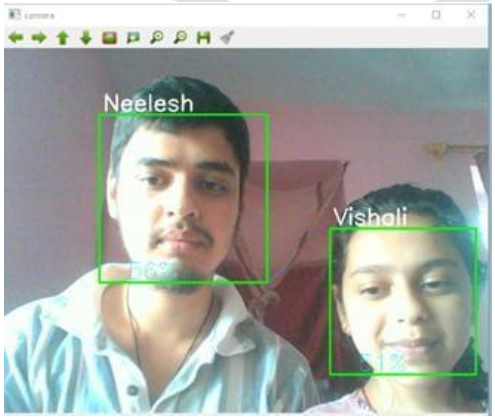

figure 15. multiple faces for detection and recognize

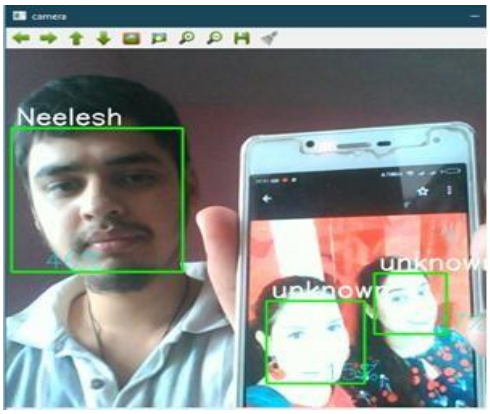

figure 16. multiple faces for detection and recognize

6.0 RESULTS

After conducting a number of tests in different condition following observations are made:

6.1 FOR PCA

1. The algorithm recognizes a single face with high accuracy. The percentage accuracy with ideal conditions is greater than 90% in almost every case.

But due to the approach used for this algorithm, if the imageprovided for creating face-encodings differs highly fromthat of the image used in real-time detection the percentage accuracy drops significantly. This problem arises because the images used for creating face-encodings are not in large numbers.

Nevertheless, the algorithm manages to recognize the faces every time.

2. The algorithm is unable to recognize or even detect a facewhen it is partially covered with an object. It only recognizes a face if the important features such as eyes, noseand lips are fully or partially visible.

3. The algorithm is capable of differentiating between adrawing and a real face. The occurrence of false positive in this case is very less.

4. The accuracy of recognizing multiple faces is similar to that of the accuracy of recognizing a single face.

5. Negative and false positive results are very less in this algorithm. Most of the negative or false positive resultsare observed when the faces are exposed to high intensity of light and the device camera was unable to control the exposure.

Thus, more accuracy can be achieved using a better camera and providing more images for creating face-encodings .but due to approach of this algorithm it is not possible in this algorithm.

6.2 FOR HAARCASCADE

1. The algorithm detects and recognizes a single face very accurately. Although the match percentage of the LBPH recognizer is not as high as PCA. The match percentage ranges from 60-70% in most cases.

2. The detection of the face when partially covered is successfully done. The algorithm wasn't able to recognize the face in some cases such as, if the lower half of the faceis covered. But it recognizes the face if the left or right half of the face is partially covered.

3. The algorithm differentiates between the real face and a drawing successfully. The false positives are rare.

4 The algorithm detects and recognizes multiple faces in the image accurately. Though, it is observed that the match percentage decreases when more than one face is detected.

5. Again, as observed in PCA high exposure to the light decreases the efficiency of the algorithm.

6. The false positives and negative results are also observedin this algorithm. The faulty results may occur due to highor very low intensity of light while detecting the face and also at the time of creating the database.

7. COMPARISON

-The Match percentage accuracy of PCA is higher than the recognizer used in Haar Cascade i.e. LBPH.

-PCA is faster than the Haar Cascade but the time efficiencygap becomes less as more faces are added.

-Haar Cascade detects faces better than the PCA (Histogramof oriented gradients). Note, the face detection should not beconfused with face recognition.

-The number of false positive and negative results in PCA are less than that of Haar Cascade.

NOTE: The approach used for both the algorithms are different from each other. In PCA, the image for creating face-encoding is taken from the device storage. Whereas, in Haar Cascade real-time images are taken for creating the database. The two different approaches are used to cover different aspects of face-recognition

8. CONCLUSIONS

In this work we aim to determine the most efficient face detection algorithm among PCA and Haar Cascade by comparing their percentage efficiency of the recognized images taken in different scenarios as described in experiment section. The experiments shows that PCA has greater face recognition efficiency as compared to Haar Cascade. The face detection part of the project was made using an OpenCV and face recognition Library. OpenCV (Open Source Computer Vision) is a library of programming functions for real-time computer vision the algorithm is coded using PyCharm and the language used is Python.

Our research methodology is based on exercising the most efficient algorithm among PCA and Haar Cascade. This work on face recognition had given us a greater extend about many popular methods used in the field of face recognition system and the trade-off associated with them. We also came to know that combining two or more techniques can improve the accuracy of the system greatly.

8.1 FUTURE WORK

For future the experiment can be repeated on a larger databases and with different algorithm. The efficiency of these algorithm can be compared with other algorithms as well.

The same experiment can be conducted using different face database and compare the results with our current experiment to ensure the validity of these algorithm over different types and sizes of database.

Other techniques to enhance the accuracy of the face recognition and decrease the false acceptance rate need to becompared.

REFRENCES

[1] International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012 .

[2] A. S. Tolba; A.H. El-Baz ; A.A. El-Harby :"Face Recognition: A Literature Review," World Academy of Science, Engineering and Technology 2006,

[3] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86

[4] http://cswww.essex.ac.uk/mv/allfaces/

[5] Harguess, J.; Aggarwal, J.K.: "Is there a connectionbetween face symmetry and face recognition?," Computer Vision and Pattern Recognition Workshops (CVPRW),2011 IEEE Computer Society Conference on , vol., no.,pp.66-73, 20-25 June 2011

[6] Ibrahim, R.; Zin, Z.M., "Study of automated face recognition system for office door access control application," Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference, vol.,no., pp.132-136, 27-29 May 2011

[7] Junying Gan; Peng Wang; , "A novel model for face recognition," System Science and Engineering (ICSSE), 2011 International Conference on , vol., no., pp.482-486, 8- 10 June 2011

[8] Kshirsagar, V.P.; Baviskar, M.R.; Gaikwad, M.E., "Face recognition using Eigenfaces," Computer Researchand Development (ICCRD), 2011 3rd InternationalConference, vol.2, no., pp.302-306, 1113 March 2011

[9] PyCharm,(R14),theMath Works. Inc, Access date, June22,2010,from: http://www.mathworks.com/products/PyCharm/

[10] Martin A.; Doddington G.; Kamm T.; Ordowski M.; Przybocki M.: "The DET Curve in Assessment of Detection TaskPerformance," http://www.itl.nist.gov/iad/mig/publications/storage_paper/d et.pdf