Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018)
IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3

Face Recognition Using LBPH Descriptor and Convolution Neural Network.

V.Betcy Thanga Shoba[1]
Research Scholar, Dept. of Computer Science,
Nesamony Memorial Christian College affiliated to
Manonmaniam Sundaranar University, Abishekapatti,
Tirunelveli-627012, TamilNadu, India
shobarobertdec27@gmail.com

Dr. I. Shatheesh Sam[2]
Associate Professor, Dept. of PG Computer Science,
Nesamony Memorial Christian College affiliated to
Manonmaniam Sundaranar University, Abishekapatti,
Tirunelveli-627012, TamilNadu, India
shatheeshsam@yahoo.com

Abstract

Face Recognition is an exigent problem in Biometrics and Computer Vision. It has become a wonderful field for researchers and can be widely used in applications like surveillance and security also. To create strong and distinct features, increase the inter-personal variations and decrease the intra-personal variations simultaneously remains a demanding problem in facial recognition. In this paper, the researcher explains how to improve the ability of face recognition system using Local Binary Pattern (LBP) for feature extraction and Convolution Neural Network (CNN) for classification of the images. The correspondence between the trained images helps CNN to converge faster and achieve better accuracy. There is a great improvement compared to other traditional methods too. To evaluate the accomplishment of this new method, it is found that higher face recognition accuracy can be achieved with less computational cost. The proposed framework is tested on the Yale dataset and achieved an accuracy of 98.6%.

Keywords: **Convolution Neural Network (CNN), Local Binary Pattern Histogram (LBPH).**

I. INTRODUCTION

Face recognition has become a fascinating field for researchers. The difficulty in face recognition is the face identification during situations of different light intensity variations and invariant pose. The main task is in designing a high accurate facial recognition system at real time [1]. The problem is challenging when there are variations in facial expression, facial appearance, various poses, light illuminations, ages, makeups and occlusions [13]. Facial recognition includes comparison between two faces. The facial features reflect the same features of the compared faces by calculating the similarity scores. Using the similarity measure faces can be recognized whether it is a similar face or a dissimilar face.

Local Binary Pattern is a powerful feature extraction method to describe the nature and standard of the input facial images. In LBP, a given input facial image is split into small regions called cells and then LBP histograms are brought out from all such region and then concatenated into a single feature vector. This feature vector forms the feature representation of the face and also calculates the similarity measure between the images [2]. To enhance the face recognition, Convolution Neural Network is used for classifying the images. Classification is the assignment of features of the face images to a group that have the images of other face image objects that all have the same particular characteristics in common. CNN identifies successively larger features in a set of layers.

Face recognition system is divided into four phases as shown in Fig.1. First phase is the image pre-processing. During image pre-processing, background noise can be removed. Second phase is the post-processing. During image post-processing, the face image is cropped to the same size as 200 x150 with face features properly aligned to mainly focus on the forehead, eyes, nose, cheeks, mouth and chin. The third phase is the feature extraction stage. Various image descriptors are used in the existing methods. Methods like Local Binary Pattern (LBP) [2], Local Ternary Pattern (LTP) [12], Scale Invariant Feature Transform (SIFT) [13], Gabor Wavelets [11], Linear Discriminant Analysis (LDA) [10] etc. Final phase is the classification stage. Existing classifiers like CNN (Convolution Neural Network) [3-4] and the fully connected Neural Network [9], Mahalanobis distance classifier [8], Hidden Markov models [5], Support Vector Machine [6] etc.

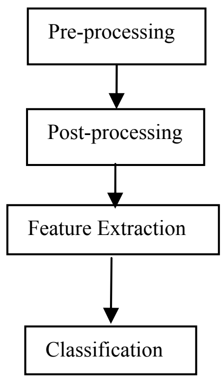

Fig.1. Face Recognition System Process.

The main contribution in this work is on face recognition using LBPH and CNN. The system starts with some pre-processing operation and post-processing operation to reduce the processing time. Feature extraction methods like LBPH can be used to compare the significant features of the facial

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.

images. Finally for classification, CNN is used to classify the face images and calculate the accuracy of this system.

The remaining of this paper is organised as follows. Section II shows the overview of the LBPH and CNN methods. The proposed face recognition system using LBPH and CNN is presented in section III. In section IV, experimental results of this face recognition system is presented and section V gives the conclusion of the system.

II. OVERVIEW

A. Median Filter

The images acquired from any source may contain noises. Some special filters are used to remove these noises. During image pre-processing, median filter can be used. Median filter is also called as order-statistics filter. The formula for median filter is

$$\hat{f}(x, y) = \underset{(s,t)\in S_{x,y}}{median} \{g(s,t)\} \tag{1}$$

where g(s, t) is the input image and S is the sample space. The median value is calculated using the intensity value of the neighbourhood of the pixel (x, y) and the median value is replaced with its neighbours. This median filter has become quite popular as they provide excellent random noise reduction capabilities, less blurring when compared with other smoothing filters. Median filters provide best results with the bipolar and unipolar noise images.

B. Cropping the input image

For post-processing, the images can be properly aligned with the forehead, two eye centers, nose tip and two mouth corners, cheeks, mouth and chin by similarity transformation. We then crop the other regions and the size of cropping should be uniform with faces aligned to ensure good correspondence between the pixels of the same person.

C. Feature Extraction Method

LBP is a generic approach that considers both the shape and texture data to represent facial images [14]. Face recognition is done using nearest neighbour classifier. LBP allows very fast feature extraction. LBP introduces discriminative feature space which is applied for face recognition. The LBP operator works on eight neighbours of the pixel. The input image divided into small regions called cells and the features are extracted from each region. The LBP operator works as follows. The face image is divided

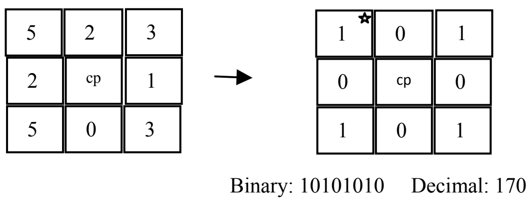

Fig.2. Circularly Symmetric Neighbour sets

where cp is the center pixel and ☆ denotes the starting point of the binary code.

into small cells. For each center pixel in the cell, compare it with its eight surrounding neighbours. Fig. 2 shows the center pixel with its surrounding circularly symmetric 8 neighbour pixels. The threshold t takes the value as

$$t = \begin{cases} 1 & \text{if pixel value} >= cp \\ 0 & \text{otherwise}. \end{cases} \tag{2}$$

It generates 8 bit binary code. Compute the histogram for each cell and normalize it. Image feature is used to reduce from 250 dimensions decimal to a 59 dimensional histogram. The value of LBP is

$$LBP_{p,r}(x_c, y_c) = \sum_{p=0}^{p=1} s(g_p - g_c)2^p \tag{3}$$

where g_c is the gray-value of the center pixel and g_p is the gray-value of the pixel value of p with the 8 neighbouring cells and S is the sample image. The overall value of LBPH can be represented as

$$LBPH(z) = \sum_{i=0}^{n} \sum_{j=0}^{m} f(LBP_{p,r}(xc, yc), z) z \in [0, z] \tag{4}$$

D. Distance Method

K-Nearest neighbour is a non-parametric classifier. Usually the distance function is calculated by comparing the distances of the testing and the training data for classification. Although many distance methods are available, the chosen distance function can affect the classification accuracy. In this proposed method, Mahalanobis distance method is used as it provides better results. If the feature extraction provides a set of values like $\vec{x} = (x_1, \ldots\ldots x_n)^T$ from a set of observations with mean $\vec{\hat{\mu}} = (\mu_1, \ldots\ldots \mu_n)^T$ and the co-variance matrix S, the Mahalanobis distance *Md(x, y)* is defined by

$$Md(x, y) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \tag{5}$$

Mahalanobis distance is an identifier for the similarity measurements between two vectors. Given a test face image, the value of Md(x, y) is calculated for each face image with its distance minimum. It is also used to detect outliers.

E. Convolution Neural Network

In Convolution Neural Network, the input is a set of images. The layers of Convolution Neural Network have neurons in three dimensions of width, height and depth. CNN layers have input layer, output layer and many hidden layers. The hidden layers have convolution layers, pooling layers and fully connected layers. CNN is a sequence of layers and every layer of a Convolution Neural Network transforms one volume of activations to another layer. The input layer contains the image and is divisible by two many times, like 32, 64, 96, 224 etc. The Convolution layer is the important building block of CNN. It accepts a volume of size W×H×D, using small filters of size 3×3 or atmost 5×5. The pooling layer lies in between two successive convolution layers and it reduces the spatial size, the amount of parameters and computation in the network. The neurons in the fully

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.

connected layers have full connections to all activations in the previous layers.

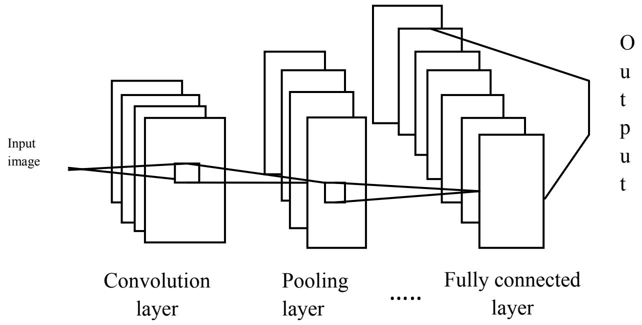

Fig.3. Typical Convolution Neural network

III. PROPOSED FACE RECOGNITION SYSTEM USING LBPH AND CNN

This proposed face recognition system uses LBPH descriptor and CNN. Fig. 4 explains the proposed architecture of the face recognition system using LBPH descriptor and CNN in detail. Raw training data is fed as input and then the image undergoes pre-processing and post-processing of the input images. Then LBPH descriptor used for feature extraction is calculated using equation 3 and equation 4. Mahalanobis distance measure is calculated using equation 5 to reduce the dissimilarities between the compared faces and then fed as input to the convolution layer. First training for 100 images is undergone and then testing is done with the remaining 65 images. Precision, Recall, F-Measure and Accuracy are also calculated.

This new proposed method is divided into five steps. In step one, pre-processing method is done by using the median filter to remove the noises and during post-processing, the input image is cropped to focus on the important features in the face image. So noise removal and cropping is done in the first phase. Fig. 5 shows how the image appears after undergoing the post-processing operations. In step two, the extracted by the LBP descriptor. Divide the images into 16 smaller cells. Thresholding is applied to the eight neighbourhood of each pixel. Then histogram method is applied to reduce the computing time.

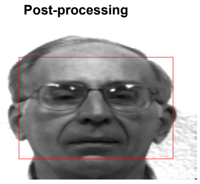

Fig.5. Post-processed image

In step three, Mahalanobis distance measure is used to calculate the distance metric between the testing data and the training data to decide the final classification output. This chosen distance function affects the classification accuracy.

In step four, the convolution parameters are set up and then the training begins. Following is the algorithm for the Convolution Neural Network.

Step 1 Initialize convolution neural network.
Step 2 Add convolution layer.
Set the size of feature maps. That is, the output will be (h-kernel_height+1) × (w-kernel_width+1) where h × w is size of input to this layer.
If this is the first layer, h × w is the size of the input image, else h × w is the size of previous layer's output.
Step 3 Add Pooling layer between the convolution layers by the size of input to this layer and it should be an integer multiple of subsampling rate (subsampling factor). The size of output of this layer is the input size /subsampling_rate.
Step 4 Add Fully connected neural network layer.
Step 5 Train the network with the training dataset
Step 6 Test the network with the testing dataset.

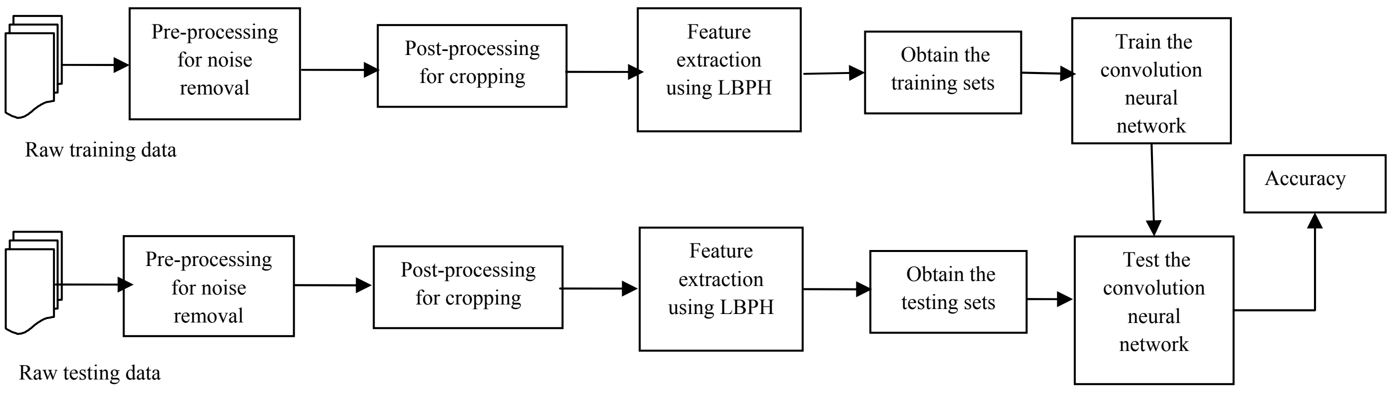

Fig.4. The proposed architecture of the Face Recognition system using LBPH and CNN.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.

Finally in step five, the accuracy is tested as follows. If this is the last layer, number of nodes should match with the number of labels in the dataset.

- Repeat steps 1 to 3 for each testing image.
- Provide the input of the testing input data to the trained CNN and obtain the expected output.
- Identify the correctness of the testing based on the image label.
- Calculate precision, recall, fmeasure and accuracy.

Precision is the ratio of the retrieved face image pattern to the input image pattern. It is the percentage of the relevant image patterns returned by the search. It is also called as positive predictive value. It is a measure of exactness. It is used with recall.

$$\Pr ecision = \frac{tp}{tp + fp} \tag{6}$$

where tp is the value for true positive, fp is false positive, tn is true negative and fn is false negative.

Recall is a fraction of the relevant image that are successfully retrieved. It is also called as sensitivity.

$$\operatorname{Re} call = \frac{tp}{tp + fn} \tag{7}$$

Fmeasure is the combination of precision and recall. It is the harmonic mean of precision and recall. It is also called F-score.

$$Fmeasure = \frac{2 \times precision \times recall}{precision + recall} \tag{8}$$

Finally, the overall system accuracy is calculated. Accuracy is a weighted arithmetic mean of the precision and inverse precision as well as the weighted arithmetic mean of recall and inverse recall. That is

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{9}$$

IV. EXPERIMENTAL RESULTS

The proposed method is tested on the small public dataset Yale to understand the benefits of the descriptors as it provides training data clearly. Fig.6 shows an example of how an input image undergoes pre-processing operation, post-processing operation, feature extraction, distance calculation and trained by convolution neural network and the output is the ten matching face images from the dataset. The Yale dataset has a total of 165 face images of 15 different persons with 11 images per person. The Yale dataset provides different facial expressions, genders, light configuration, images wearing eye glasses etc. The 165 images are in gray scale domain and images are cropped to the size of 92 × 112 pixels. Only 100 images are trained by the CNN and the remaining images are tested by the system. The percentage of accuracy is compared with the other existing methods and is listed in the table-I. The table below shows the comparison of this proposed system that has achieved the accuracy of 98.6% to the other existing methods.

TABLE. I COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD TO THE EXISTING METHODS ON THE DATASET.

Year	**Method**	**Number of training images per person**	**Accuracy %**
2015	PCA+BPNN	N/A	88
2015	LDA	5	89.5
2015	GABOR+NMF	5	95
2014	PCA,LDA,DCT,ICA	5	85.5,88.5,91.5, 87.5
2013	FKNN	5	87
2012	WT+PCA	8	95
2012	CASMM,FFNN	5	86.5,80
2011	PCA-DDCT-Corr-PIFS	N/A	86.8
2018	LBPH, multi-KNN, BPNN	5	98
2018	**Proposed: LBPH,CNN**	**11**	**98.6**

Precision is the fraction of relevant information that is retrieved over the total amount of relevant information. Suppose in the face image matching, there are 8 matching features, 5 features identified (true positive) but containing a total of 12 features, then precision is 5/8 and recall is 5/12. Recall is a measure of completeness. In FMeasure, precision and recall are evenly weighted.

Table II shows the testing set image count for the various images and the calculated precision, recall, Fmeasure and accuracy values are displayed below.

TABLE. II SHOWING THE TESTING SET IMAGE COUNT FOR PRECISION, RECALL, FMEASURE AND ACCURACY.

Precision	0.9940	0.9990	0.9750	0.9733
Recall	0.9790	0.9910	0.9790	0.9836
Fmeasure	0.9860	0.9950	0.9770	0.9784
Accuracy	0.9860	0.9950	0.9770	0.9785

In face recognition, a perfect precision score of 1.0 means that every result retrieved by a database search was relevant whereas a perfect recall score of 1.0 means that all relevant faces were retrieved by the search. Both precision and recall are used for understanding the Fmeasure of relevance. The results displayed in table-II shows values greater 0.97.

Figures 7 to 10 show the plots for the individual calculated precision values, Recall values, Fmeasure values and Accuracy. Figure 11 show the overall result of the precision values, Recall values, Fmeasure values and Accuracy values.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.

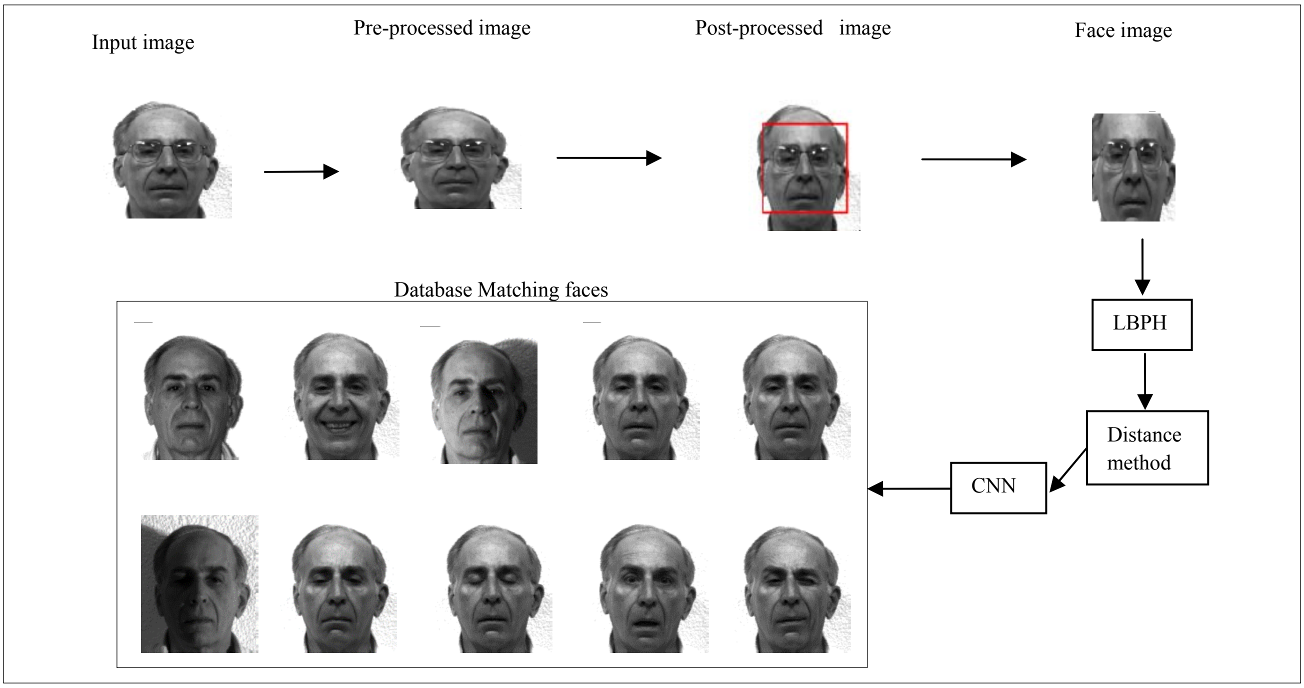

Fig.6. **An example of the input image then undergoing pre-processing by removing noise and then post-processing by cropping the face where features can be extracted and then input given to the convolution layers and the matching face images got as output from the YALE database.**

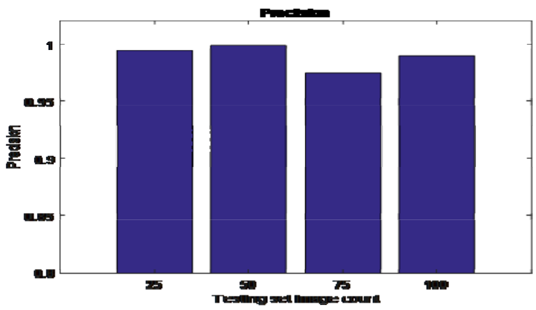

Fig.7. Comparison of Precision values.

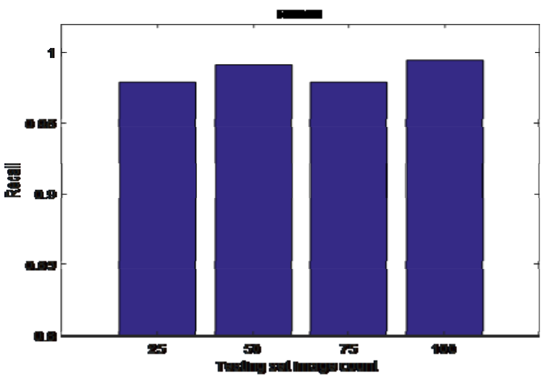

Fig.8. **Comparison of Recall values.**

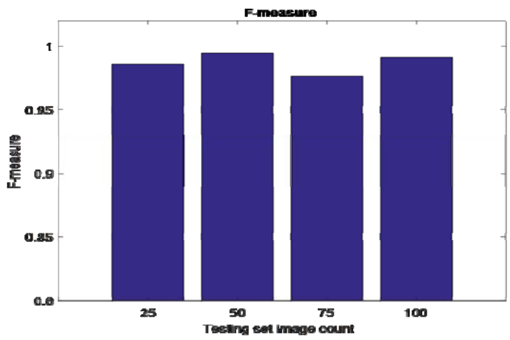

Fig.9. **Comparison of Fmeasure values**

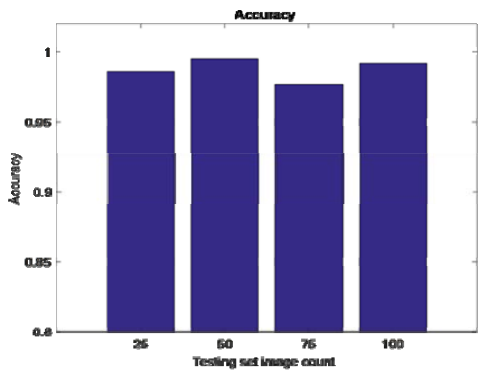

Fig.10. **Comparison of Accuracy Values**

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.

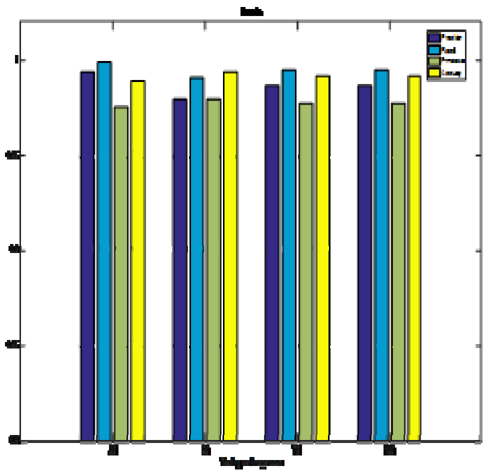

Fig.11. Comparison of Precision, Recall, Fmeasure and Accuracy Values.

V. CONCLUSION

This new proposed system provides an improved framework for face recognition system using LBPH descriptor and Convolution Neural Network. The LBPH descriptor and CNN are helpful to provide a training data set with distinction patterns based on the correspondence between the original training images. The Yale dataset helps the CNN to converge faster with greater accuracy and takes less computation time. Moreover, Precision, Recall, Fmeasure and Accuracy of the system are calculated. However, in this proposed work, when a facial image is input, it recognizes only the similar faces found in the dataset. In future, this proposed work is extended to identify different facial expressions like smile, disgust etc among the recognized faces.

References

[1] Mohannad A.Abuznied and Ausif Mahmood "Enhanced Human Face Recognition Using LBPH Descriptor, Multi-KNN, and Back-Propagation Neural Netowrk," ACCESS.2018,2825310, IEEE Access.

[2] Sarabjit Singh, Amritpal Kaur, Taqdir "A Face Recognition Technique using Local Binary Pattern Method", International Journal of Advanced Research in computer and Communication Engineering.

[3] Steve Lawrence,C.Lee Giles, Ah Chung Tsoi and Andrew D. Back " Face Recognition: A convolutional Neural-Network Approach", IEEE Transactions on Neural Networks. Vol-8 No.1, January 1997.

[4] Alwin Anuse, Vibha Vyas "A novel training algorithm for convolutional neural network" Springer, Instell Syst(2016), Cross mark.

[5] H.Othman, H.Abouhnasr," A separate low complexity 2D HMM with application to face recogniton, IEEE Transaction on pattern Analysis and Machine Intellignece 2000, vol 25,no.10,pp1229-1238.

[6] Lee K.Chang,Y.Byun, H,"SVM based face verification with feature set of samall size, Electronic letters 2002, vol 38, no.15,pp 787-789.

[7] Yale Face Dataset. Available online:http://www.vision.ucsd.edu/content/yale-face-database(accessed on 23 01 2017.

[8] Mahmon, N.A.Yaacob, N.Yusuf A.L,"Differences of image classification Techiniques for Land use and land cover classification, IEEE Trans, CSPA 2015, pp90-94

[9] Boughrara H. Chtourou,M.Amar,C.Chen,"Face Recognition based on perceived facial images and multilayer perceptron neural network using constructive training algortihm,IET computer vision 2014,vol 8,no.6,pp729-739.

[10] P.N.Belhumeur, J.Hespanha and D.J.Krienmam Eigenfaces vs fisher faces recognition using class specific linear projection TPAMI,19(7)711-720,1997.

[11] C.Liu and H.Wechsler," Gabor feature base classification uisng the enhanse fisher linear discriminant model for face recognition. 11(4) 467-476,2002.

[12] Tan. X. Triggs,"Enhanced local texture feature sets for face recogniton under difficult lighting conditions, IEEE Trans Image Process 2010, vol 19, no.6 pp 1635-1650.

[13] D.G.Lowe, "Distinctive image features from scale-invariant keypoints", Int.J.Comput, Vis, Vol 60,no.2,pp 91-110, Nov 2004.

[14] Jiwen Lu, Venice Erin Liong, Gang Wang, Pierre Moulin," Joint Feature Learning for Face Recognition", IEEE.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on June 23,2020 at 13:33:27 UTC from IEEE Xplore. Restrictions apply.