Local Binary Pattern based Resolution Variation Video-based Face Recognition

SHAILAJA ARJUN PATIL[1], P. J. DEORE[2]
[1] E & Tc department, R. C. Patel Institute of Technology, Shirpur, Dist: Dhule, North Maharashtra University, Jalgaon, INDIA. shailajadp@gmail.com
[2] E & Tc department, R. C. Patel Institute of Technology, Shirpur, Dist: Dhule, North Maharashtra University, Jalgaon, INDIA. pjdeore@yahoo.com

Abstract: - Video-based face recognition is a very challenging problem as there is a variation in resolution, illumination, pose, facial expressions and occlusion. In this paper, we have presented an approach for resolution variation video-based face recognition system using the combination of local binary pattern (LBP), principal component analysis (PCA) and feed forward neural network (FFNN). We used, standard as well as created database. The main purpose of this paper is to evaluate the performance of the system. To the best of our knowledge this is the first work addressing the issue of resolution variation for video-based face recognition with this approach. We have experimented with three different video face databases (Created database, NRC_IIT & HONDA/UCSD) and compared with benchmark methods. Experimental results show that our system achieves better performance than other video-based face recognition algorithms on challenging resolution variation video face databases and thus advancing the state-of-the-art.

Key-Words: - Video-based face recognition, Local Binary Pattern, Principal Component Analysis, Feed forward Neural Network.

1 Introduction

In many organizations, universities, shops, institutes, schools, banks, day care centers and many more places, surveillance system is very important for security. In surveillance system video-based face recognition plays a very important role. In face recognition, the face of the person is his identity. In face verification system, it verifies or authenticates the person's face; it is one to one matching while face recognition or identification means one to many matching. The performance of face recognition systems has improved significantly since the first automatic face recognition system was developed by Kanade [1]. In video based face recognition system, it consists of video frames, identification or recognition of a person's face from video. The video consists of abundant information. The video is a sequence of time varying images. The video signal is treated as a series of images called video frames. Most of video formats use temporal sampling rates of 24 frames per second and above [2]. Up till now many different algorithms and methods are used for video-based face recognition but still there are many challenges like resolution variation, pose variation, illumination variation, facial expressions and occlusion.

In this paper the local binary pattern (LBP) for feature extraction, principal component analysis (PCA) for feature reduction and feed forward neural network (FFNN) for feature matching. The combination of LBP, PCA and FFNN gives a better accuracy.

The concept of face recognition is in use since 1960, this is the first semi-automated system for face recognition. In 1970, Goldstein and Harmon used 21 specific subjective markers for face recognition. The standard linear algebra was used for face recognition by Sirovich and Kirhy in 1988.

Y. Tahata et. al investigated and evaluated the recognition characteristics of the real-time 2D-SAN net system [3]. D. Gong proposed a novel semantic based subspace model to improve the performance of video based face recognition [4]. Chia-Te Liao et al proposed a novel face recognition system with only one single image for each individual in the training dataset [5].

M. Shamim Hossain and Ghulam Muhammad proposed an emotion recognition with high performance for mobile applications, in this system, facial video is captured by an embedded camera of a smart phone [6]. N. M. Khan, X. Nan, A. Quddus and E. Rosales proposed a method which improves the sparse representation framework. The key contribution is an intelligent and adaptive sparse dictionary that updates the current probe image into the training matrix based on continuously monitoring the probe video through a novel confidence criterion and a Bayesian inference scheme [7]. A. Mohammedan, H. Again and F.

Towhidkhah a model has been proposed to generate diverse sequences of virtual samples for a new subject. These sequences are thought to enrich the training set in order to increase robustness of recognition with respect to individual variations and improve generalization to the new person [8].

In our approach, we are using local binary pattern (LBP) for feature extraction, as it is the best texture operator for resolution variation video databases. These features are large in size, feature reduction is necessary for reducing the processing time, principal component analysis (PCA) is used for feature reduction and then for feature matching feed forward neural network (FFNN) is used.

The steps of our approach are,

1. To perform a face detection using Viola-Jones algorithm and face cropping.
2. To create the mapping pattern for resolution variation local binary pattern of video frames.
3. To extract the features using local binary pattern.
4. To reduce the features using principal component analysis.
5. To classify the features using a feed forward neural network.
6. To verify the accuracy of standard as well as created video-databases.

2 Video-based Face Recognition

A Video-based face recognition system generally consists of four modules as depicted in the block diagram of Fig. 1, detection, alignment, feature extraction and matching, where localization and normalization (face detection and alignment) are processing steps, before face recognition (facial feature extraction and matching) is performed. Face detection segments the face areas from the background. In the case of video, the detected faces may need to be tracked using a face tracking component. Face alignment is aimed at achieving more accurate localization and at normalizing faces, thereby whereas face detection provides coarse estimates of the location and scale of each detected face. Facial components are located such as eyes, nose, mouth and facial outline. The input face image is normalized with respect to geometrical properties such as size and pose, using geometrical transforms or morphing. The face is usually further normalized with respect to photo-metrical properties such illumination and gray scale. After a face is normalized geometrically and photo-metrically, feature extraction is performed to provide effective information that is useful for distinguishing faces of different individuals and stable with respect to the geometrical and photo-metrical variations. For face matching, the extracted feature vector of the input face is matched against those of enrolled faces in the database; it outputs the identity of the face when a match is found with sufficient confidence or indicates an unknown face otherwise.

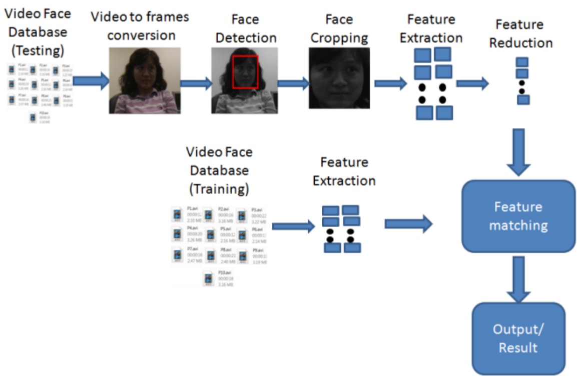

Fig. 1: Video-based Face Recognition

2.1 Face detection using Viola-Jones algorithm

As our system is video-based, the input is a video of a person; from this video our algorithm converts the video in frames. Video is a combination of images with respect to time. One video is a combination of average 300 frames. Consecutive frames have minor changes, we have taken five random frames, which are having major variations. Test video dataset is used to detect faces and crop them across the frames. The median filter is used to remove the noise. After face detection, it obtained facial features using local binary pattern (LBP). The obtained features are in large dimensions [30,18900 dimensions]. These dimensions are reduced to 100 dimensions for each feature vector using PCA. These features were given as a test input to the algorithm. The training features are available from a gallery of the training data set. The outcome of this algorithm is to recognize the person of interest from the test video database.

In this paper, the algorithm provides better recognition accuracy for the resolution variation challenge. In this algorithm, we have considered the five random frames of each person video with major variations. For face detection, an object detector is used to detect the location of the face in a frame. The Cascade Object detector is a function in MATLAB which is used for face detection. The cascade object detector uses viola-jones algorithm as a classifier for detecting faces in the video [9]. As it starts detecting faces in a single frame, then it starts detecting same face from successive frames, in which step function is used. As the face gets cropped then the next step is feature extraction.

2.2 Local Binary Patterns

In previous approaches, local binary pattern (LBP) was used for many challenges like illumination variation, facial expressions and pose variations, here we use LBP for resolution variation, as LBP is the best texture operator.

In local binary pattern (LBP) the texture analysis operator is defined as a grayscale invariant texture measure, derived from a g eneral definition of texture in a local neighborhood. The basic local binary pattern operator, introduced by Ojala et al. [10], was based on the assumption that texture has locally two complementary aspects, a pattern and its strength. In that work, the LBP was proposed as a two-level version of the texture unit to describe the local textural patterns. In local binary pattern there are three pattern labels like uniform, rotation invariant and uniform rotation invariant. We can consider any one pattern label; we have considered the uniform pattern label. An example of an LBP image and histogram are shown in Fig. 2.

Fig. 2: Face Detection, Face Cropping, LBP Histogram

2.2.1 Derivation of the Generic LBP Operator

The local binary pattern is the best texture operator and after many years its original publication was presented by Ojala et al. [11]. In contrast to the basic LBP using 8 pixels in a 3×3 pixel block, this generic formulation of the operator puts no limitations to the size of the neighborhood or to the number of sampling points. The derivation of the generic LBP presented below follows that of [11]. Consider a monochrome image $I(x,y)$ and let g_c denote the gray level of an arbitrary pixel (x,y), i.e. $g_c = I(x,y)$. Moreover, let g_p denote the gray value of a sampling point in an evenly spaced circular neighborhood of P sampling points and radius R around point (x,y):

$$g_p = I(x_p, y_p),\ p = 0.....p-1.........(1)$$

$$x_p = x + R\cos(2\pi p / p),..........(2)$$

$$y_p = y - R\cos(2\pi p / p),.........(3)$$

The learning vector quantization based approach still has certain unfortunate properties that make its use difficult. First, the differences $g_{p-}g_c$ are invariant to changes of the mean gray value of the image, but not to other changes in gray levels. Second, in order to use it for texture classification the codebook must be trained similar to the other texton-based methods. In order to alleviate these challenges, only the signs of the differences are considered:

$$t(\mathrm{s}\ (g_0 - g_c), s(g_1 - g_c),......s(g_{p-1} - g_c)),........(4)$$

S(z) is the thresholding function

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} ....(5)$$

The generic local binary pattern operator is derived from this joint distribution. As in the case of basic LBP, it is obtained by summing the thresholded differences weighted by powers of two. The LBP P,R operator is defined as

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(g_p - g_c)2^p......(6)$$

In practice, this equation means that the signs of the differences in a neighborhood are interpreted as a P -bit binary number, resulting in 2^p distinct values in the LBP code. The local grayscale distribution, i.e. texture, can thus be approximately described with a 2^p -bin discrete distribution of LBP codes.

2.3 Principal Component Analysis

The extracted features from local binary pattern (LBP) are in large dimensions. This computation of features in large dimensions produces redundant features and reduces the prediction accuracy and increases computation time. These large dimension feature vectors are reduced by principal component analysis (PCA). Principal component analysis (PCA) method used for extracting global features from high-dimensional data set. It can also be used to identify patterns in data, and expressing the data in such a way that it highlights their similarities and differences.

Principal Component Analysis (PCA) [12], [13] is a statistical dimensionality-reduction method, which produces the linear least-squares subspace of a training set. The s-dimensional vector representation of each face image in a training set of M images, the

PCA could find a t-dimensional subspace whose basis vectors correspond to the maximum variance directions in the original image space. New t-dimensional subspace is normally a lower dimensional space than the original space (t<<s). All images of a training set are projected onto the subspace, and then sets of weights W^{train}, which describe the contribution of each vector, are calculated.
To recognize an identity of a test image, the image is projected onto the same subspace, then a set of weights W^{test} is figured out. After the distance measures between W^{test} and each set of W^{train}, the identity of W^{train} is determined to that of a minimum-distanced set in W^{train}. The PCA basis vectors are defined as eigenvectors of the scatter matrix S_T defined as:

$$S_T = \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T .....(7)$$

where μ is the mean of all images in the training set and x_i is the i_{th} image in the training set.

2.4 Feed Forward Neural Network

The reduced feature vectors are given to the feed forward neural network for classification and matching purposes. FFNN is a biologically inspired classification algorithm. It consists of a (possibly large) number of simple neuron-like processing units, organized in layers. Every unit in a layer is connected with all the units in the previous layer. These connections are not all equal: each connection may have a different strength or weight. The weights on these connections encode the knowledge of a network. Often the units in a neural network are also called nodes. Neural network consist of input, hidden layers, output layer and output. Here we have taken 25 hidden layers and 1 output layer. 25 epochs are used in FFNN. Data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs. During normal operation, that is when it acts as a classifier, there is no feedback between layers. This is why they are called feed forward neural networks.

2.5 Resolution in Video-based face recognition

In video based face recognition for better accuracy, image quality is very important and for image quality, resolution is the major parameter. For Video-based face recognition here we use, the local binary pattern for resolution variation. Resolution is the sharpness and clarity of an image, measured in terms of ppi and dpi. PPI (pixels per inch) Measurement used for images displayed on Screen. DPI (dots per inch), Measurement used in printing images. A Pixel is a word invented from "picture element" the basic unit of programmable color on a computer display or in a computer image. A megapixel (that is, a million pixels) is a unit of image sensing capacity in a digital camera. In general, the more megapixels in a camera, the better the resolution when printing an image at a given size.
We have used three different databases with three different resolutions, created database resolution is 320×240, in NRC_IIT database resolution is 160×120 and in HONDA/UCSD database is 640×480. As the resolution is variable, still our approach achieves better performance.

3 Results and Discussion

3.1 Databases

We have used the created as well as standard databases. The Databases consist of video clips of male and female. Database explanation is as follows.

3.1.1 NRC IIT Facial Video standard Database

This database consists of pairs of short video clips captured by an Intel webcam mounted on the computer monitor. It shows a wide range of facial expressions and orientations. This database is downloaded from the FRiV technical website. The details of the database are as follows: The video capture resolution is 160 ×120. Resolution: 160 ×120. Average file size: 1.5 MB (Video : 900 KB, Audio: 600 KB - not used), Average duration: 10 - 20 secs. Average total number of frames in a clip: 300. Video type and compression: colour AVI, 20.0 fps (unless indicated otherwise). Intel web cam-provided codec compressed at 481 Kbps [16].

3.1.2 Honda/UCSD Video standard Database

The standard Honda/UCSD database, consists of male as well as female videos. Each video sequence is recorded in an indoor environment at 15 frames per second, and each lasted for at least 15 seconds, all the video sequences contain significant 2-D (in-plane) and 3-D (out-of-plane) head rotations means variations of pose, occlusion and facial expressions. The video clips taken by the Sony digital camera resolution of each video sequence are 640×480. The details of database are as follows: frame width is

640, frame height 480, data rate 113133 kbps, total bit rate 113133 kbps, frame length 27 seconds and frame rate is 14 frames per second. This database consists of two video clips of each person, one for training and one for testing purposes [17].

3.1.3 Created Video Face Database

The created database consists of male as well as female videos with variations of resolution, occlusion and facial expressions. The video clips taken by the mobile camera of 8 mega pixels. The details of database are as follows: frame width is 320, frame height 240, data rate 683 kbps, total bit rate 797 kbps, frame length 6 seconds and frame rate is 25 frames per second. This database consists of two video clips of each person, one for training and one for testing purpose.

3.2 Algorithm

Input: Video clip $I(x, y)$ with a set of frame sequences.

Output: Recognized face $I'(x', y')$

Step 1. Preprocessing steps

- Video clips to frame conversion.
- Random selection of frames or images.
- Conversion from RGB to grayscale.
- Initialize median filter

Step 2: Face detection using vision.CascadeObjectDetector

Step 3: Face cropping from the frame.

Step 4: Apply LBP texture classifier on cropped face $I(x, y)$.

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \text{....(8)}$$

where s(z) is the thresholding (step) function.

Step 5: Block per image are 20×26, sampling points 8 per circle and radius of circle 2, 3, 4.

Step 6: Apply PCA for feature reduction.

$$S_T = \sum_{i=1}^{M} (x_i - \mu)(x_i - \mu)^T \text{....(9)}$$

Step 7: Apply feed forward neural network for matching the face frame $I'(x', y')$ from the original input frame $I(x, y)$.

3.3 Tools used

Matlab is a matrix laboratory. It is fourth generation programming language. This language is developed by math works. In this paper, MATLAB 2013a version has been used which is also compatible for next versions. It is used for a range of applications, including signal processing and communications, image and video processing, test and measurement, computational finance and computational biology.

3.4 Results

For video-based face recognition there are some formulae to verify the result parameters as follows.

3.4.1 Formulae

a) Sensitivity or True Positive Rate (TPR) or Recall

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \text{.....(10)}$$

b) Specificity (SPC) or True Negative Rate (TNR).

$$TNR = \frac{TN}{N} = \frac{TN}{FP + TN} \text{.....(11)}$$

c) Precision or Positive Predictive Value (PPV)

$$PPV = \frac{TP}{TP + FP} \text{......(12)}$$

d) Negative Predictive Value (NPV)

$$NPV = \frac{TN}{TN + FN} \text{......(13)}$$

e) Fall-out or False Positive Rate (FPR) or False Match Rate (FMR).

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC \text{.....(14)}$$

f) False Discovery Rate (FDR)

$$FDR = \frac{FP}{FP + TP} = 1 - PPV \text{.....(15)}$$

g) Miss rate or False Negative Rate (FNR) or False Non Match Rate (FNMR)

$$FNR = \frac{FN}{P} = \frac{FN}{TP + FN} \text{.....(16)}$$

h) Accuracy (Acc)

$$Acc = \frac{TP + TN}{P + N} \text{......(17)}$$

i) F1 score is the harmonic mean of precision and sensitivity.

$$F_1 = \frac{2TP}{2TP + FP + FN} \text{.....(18)}$$

j) Matthews Correlation Coefficient (MCC)

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{(TP+FP)(TP+TN)(TN+FP)(TN+FN)}} \quad ....(19)$$

3.4.2 Result Tables

In this paper, we have applied different methods and combinations to verify the results. In table 1, comparison is done for different parameters, for three different databases having different resolutions for the combination of LBP+PCA+FFNN.

Table 1: Comparative results for three databases for the combination of LBP+PCA+FFNN

Parameters	Created Database	NRC_IIT	Honda UCSD
True Positive (TP)	20	24	24
True Negative (TN)	22	22	5
False Positive (FP)	3	1	1
False Negative (FN)	5	3	0
True Positive rate/Recall/ Sensitivity	0.8	0.8888	1
True Negative rate (TNR)/Specificity	0.88	0.9565	0.8333
Positive Predictive Value(PPV)/ Precision	0.8695	0.96	0.96
Negative Predictive Value(NPV)	0.8148	0.88	1
False Positive Rate(FPR)/ Fall out	0.12	0.0434	0.16
False Discovery Rate (FDR)	0.1304	0.04	0.04
False Negative Rate (FNR)/ Miss Rate	0.2	0.1111	0
Accuracy (Acc)	0.84	0.92	0.9666
F1 Score	0.8333	0.923	0.9795
Matthew's Correlation Coefficient	0.6821	0.4863	0.8944
Informedness	0.68	0.8453	0.8333
Markedness	0.6843	0.84	0.96

In Video-based face recognition, the performance of the system is depending on some parameters. In table 1, the comparison is done by parameters, on three different databases. From the above table, it shows that our method gives better results.

Table 2: Comparative results of accuracy for three databases for the three different combinations for three different devices with different resolutions.

Database	Device	Resolution	LBP + ICA+ ED %	LBP+ ICA+ FFNN %	LBP+ PCA+ FFNN %
Created Database	Mobile Camera	320×240	72	73	**84**
NRC_ IIT	Web Camera	160 ×120	69	70	**92**
Honda UCSD	Digital Camera	640 ×480	78	80	**96.66**

In the table 2, comparison is done with the results of accuracy for three databases for the three different combinations for three different devices with different resolutions. As per the results, our approach LBP+PCA+FFNN gives better results for resolution variation as compare to other combinations. The resolution of the created video database taken by Mobile Camera is 320×240 which is the medium resolution; our approach gives a better accuracy of 84%. For NRC_IIT video database taken by web Camera, its resolution is 160×120 which is the low resolution; our approach gives a better accuracy of 92%. For HONDA/UCSD video database taken by a Digital Camera, its resolution is 640×480 which is higher resolution than previous two resolutions; our approach gives the best accuracy of 96.66%. There is a variation in resolution, still our approach gives the best results for all three databases. In future we may apply this approach for CCTV footages, which is having a very low resolution.

Table 3: Basic Parameters of Video-based face recognition

Parameter	Input	Output
True Positive (TP)	Positive	Positive
True Negative (TN)	Negative	Negative
False Positive (FP)	Negative	Positive
False Negative (FN)	Positive	Negative

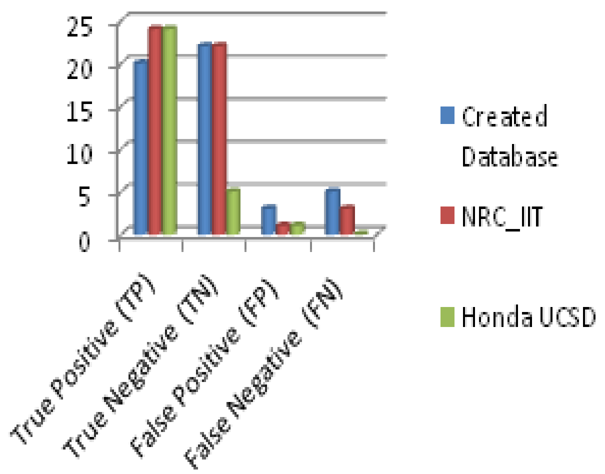

Fig. 3: Comparative result parameters for three databases.

Table 3 shows the basic parameters of video-based face recognition for analysis purpose. From this table it defines that True positive (TP) means if positive input, then the output is also positive. True negative (TN) means if negative input, then the output is also negative. False positive (FP) means if negative input, then the output is positive. False negative (FN) means if positive input, the output is negative. Fig. 3 s hows the comparative result of basic parameters for three databases.

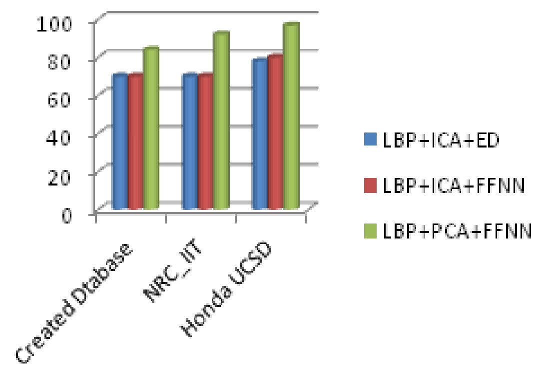

Fig. 4: Comparative results of percentage accuracy for three different combinations for three databases.

Fig. 4 shows that the percentage accuracy of our approach, which is more for all three databases than other two combinations.

Table 4: Comparative results for three databases for Different methods.

Method	Results on Created Database	Results on NRC_IIT	Results on HONDA/UCSD
PCA	65%	66.8%	69.6%
LDA	70.2%	73.5%	74.5%
ICA	70.4%	74.2%	75%
LBP	77.5%	78.3%	79.6%
HMM	82.2%	83.5%	84.2%
ARMA	82.7%	84%	84.9%
LBP+ICA+ED	70%	70%	78%
LBP+ICA+FFNN	70%	70%	80%
LBP+PCA+ FFNN (our approach)	**84%**	**92%**	**96.66%**

3.4.3 Benchmark Methods

In table 4, for comparison, we have presented different algorithms, including PCA, LDA, ICA, LBP, Hidden Markov models (HMMs) [14], Auto-Regressive and Moving Average (ARMA) models [15] and the combination of LBP+ICA+ED, LBP+ICA+FFNN as benchmark methods. Our approach is the combination of LBP+PCA+FFNN for video based face recognition. In fig. 5 we have presented the graphical comparison between different methods for three different databases.

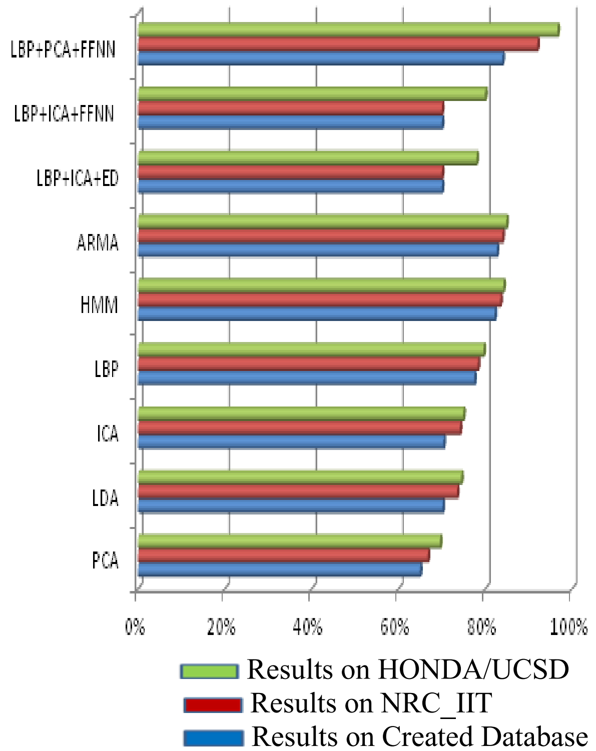

Fig. 5: Comparative results for three databases for Different methods.

3.4.4 Receiver Operating Characteristics (ROC)

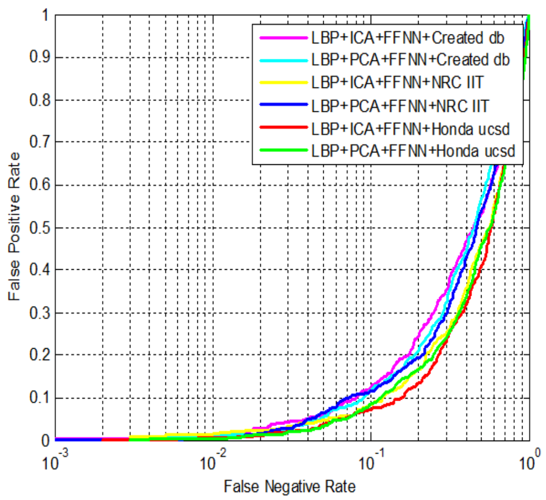

Fig. 6: Comparative results of Receiver Operating Characteristics (ROC) for two different combinations for three databases.

Fig. 6 shows comparative results of the receiver operating characteristics (ROC) for two different combinations for three databases. In statistics, a receiver operating characteristic curve or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as i ts discrimination threshold is varied. The curve is created by plotting the false positive rate (FPR) against the false negative rate (FNR) at various threshold settings. The False negative rate is also known as miss rate. The False-positive rate is also known as the fall-out or probability of false alarm. The ROC curve is thus the sensitivity as a function of fall-out. The ROC curve can be generated by plotting the False positive rate in the y-axis versus the cumulative distribution function of the false negative rate on the x-axis. The graph shows that our approach shows better performance.

3.4.5 Expected performance curves (EPC)

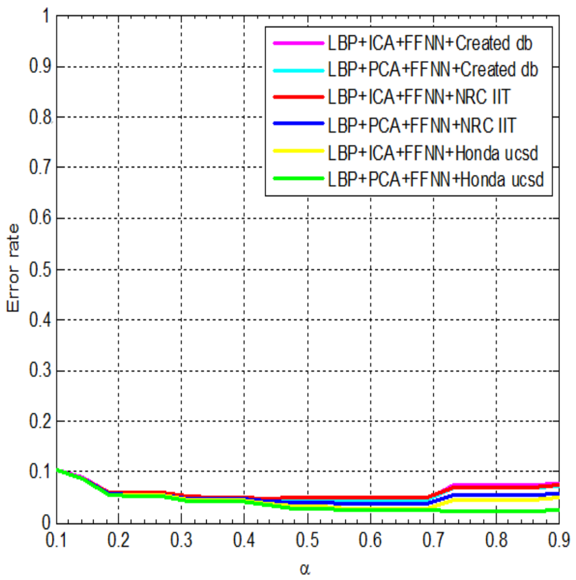

Fig. 7: Comparative results of Expected performance curves (EPC) for two different combinations.

Fig. 7 s hows comparative results of expected performance curves (EPC) for two different combinations for three databases. In statistics, an expected performance curves or EPC curve, is a graphical plot that illustrates the performance in between error rate and alpha. The graph shows that the error rate of our approach is very less as compared to other methods.

3.4.6 Cumulative match scores curves (CMC)

Fig. 8 shows comparative results of cumulative match score curves (CMC) for two different combinations for three databases. The cumulative match score curves (CMC), is a graphical plot that illustrates the performance in between recognition rate and rank. The graph shows that the recognition rate of our approach is better as compared to other methods.

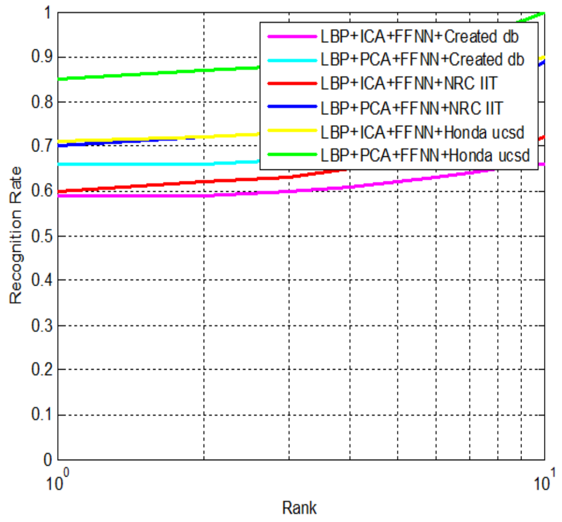

Fig. 8: Comparative results of cumulative match score curves (CMC) for two different combinations.

4 Conclusion

In this paper, we have presented the local binary pattern based resolution variation video-based face recognition system. Our experimental results show that our approach (LBP+PCA+FFNN) achieves better results than other previous methods. In addition, our system reduces the processing time, as the five frames from each video for training purpose, which consist of more information. The local binary pattern is used for feature extraction as well as it supports for feature reduction, PCA is used for dimensionality reduction, so it saves memory space as well as processing time. For future work, we have plans to use databases with pose variations and illumination variations challenges for video-based face recognition.

ACKNOWLEDGMENT

This project is supported by University Grant Commission, Western Regional Office, Ganeshkhind, Pune-411007, Maharashtra, India. (No. F.: 47-1110/14 (WRO), Dated: February 25, 2016)

References:

[1] T. Kanade,"Picture Processing by Computer Complex and Recognition of Human Faces", PhD thesis, Kyoto University, 1973.

[2] A. Murat Tekalp, "Digital Video Processing", Prentice Hall Signal Processing Series, Upper Saddle River, 1995.

[3] Y. Tahata, K. Nakamura, H. Kanayama, "Real time recognition of face shape and position by a two-D spreading associative neural network", SICE 2002, P roceedings of the 41st SICE Annual Conference, vol. 4, Print ISBN: 0-7803-7631-5, DOI: 10.1109/SICE.2002.1195760, pp: 2289-2294, 2002.

[4] D. Gong, K. Zhu, Z. Li and Y. Qiao, "A semantic model for video based face recognition", Information and Automation (ICIA), 2013 I EEE International Conference, IEEE DOI: 10.1109/ICInfA.2013.6720507, pp. 1369-1374, 2013.

[5] Chia-Te Liao, Shu-Fang Wang, Yun-Jen Lu and Shang-Hong Lai, "Video-based face recognition based on view synthesis from 3D face model reconstructed from a single image", Multimedia and Expo, 2008 IEEE International Conference, IEEE DOI: 10.1109/ICME.2008.4607753, pp. 1589-1592, 2008.

[6] M. Shamim Hossain and Ghulam Muhammad, "An Emotion Recognition System for Mobile applications", IEEE journal, special section on emotion-aware mobile computing, vol. 5, pp. 2283-2287, 2017.

[7] N. M. Khan, X. Nan, A. Quddus and E. Rosales, "On video based face recognition through adaptive sparse dictionary", Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops, IEEE DOI: 10.1109/FG.2015.7163134 vol. 1, pp. 1-6, 2015.

[8] A. Mohammadian, H. Aghaeinia and F. Towhidkhah, "Diverse videos synthesis using manifold-based parametric motion model for facial understanding", IET Image Processing, IEEE DOI: 10.1049/iet-ipr.2014.0905, Vol. 10, issue 4, pp. 253-260, 2016.

[9] Paul Viola and Michael Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision, Kluwer Academic Publishers. Manufactured in The Netherlands, 57 (2), pp. 137-154, 2004.

[10] T. Ojala, M. Pietikäinen and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, p p. 582-585,1994.

[11] T. Ojala, M. Pietikäinen and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, pp. 51-59, 1996.

[12] L. Sirovich and M. Kirby, "Low-dimensional Procedure for the Characterization of Human Faces", Journal of Optical Society of America, vol. 4, no. 3, pp. 519-524, March 1987.

[13] Matthew Turk and Alex Paul Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71 - 86, 1991.

[14] Liu X. and Chen T, "Video-based face recognition using adaptive hidden markov models", IEEE International conference on computer vision and Pattern recognition, pp. 340-345, 2003.

[15] Aggarwal G., Chowdhury A. R. and Chellappa R., "A system identification approach for video-based face recognition", 17th International conference on pattern recognition, vol. 4, pp. 175-178, 2004.

[16] Aapo Hyvarinen, Juha Karhunen and Erkki Oja, "Independent Component Analysis", A Wiley-inter science Publication, John Wiley and Sons, 2001.

[17] Dmitry O. Gorodnichy, "Video-based framework for face recognition in video", Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330-338, Victoria, BC, Canada, ISBN 0-7695-2319-6, 9-11 May, 2005.

[18] Hongyu Xu, Jingjing Zheng, Azadeh Alavi and Rama Chellappa, "Learning a structured dictionary for video-based face recognition", 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), DOI:10.1109/WACV.2016.7477689, pp. 1-9, 2016.

[19] Yuanyi Zhong, Jiansheng Chen and Bo Huang, "Towards End-to-End Face Recognition through Alignment Learning", Department of Electronic Engineering, Tsinghua University, Jan 2017.

[20] Yandong Wen, Kaipeng Zhang, Zhifeng Li and Yu Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition", The Chinese University of Hong Kong, Hong Kong, 2016.

[21] Xi Peng, Xiang Yu, Kihyuk Sohn, Dimitris Metaxas and Manmohan Chandraker, "Reconstruction for Feature Disentanglement in Pose-invariant Face Recognition" Department of Computer Science, Rutgers University, NJ, US, Feb 2017.