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

A novel technique to facial recognition based on coupled learning of convolutional neural networks and local phase quantization

[1]Raveendra K, [2]Ravi J
[1] Assistant Professor, [2]Professor
[1]Electronics and Communication Engineering,
[1]Government Engineering College, K R Pet, India

Abstract: The success rate of face recognition is influenced by illumination, expression, posture change and other variables thus use of Convolutional Neural Network (CNN) alone has a weak generalization capability. A novel technique to facial recognition based on coupled learning of convolutional neural networks and Local Phase Quantization (LPQ) is proposed as a solution to this problem. First, face texture features are extracted using the LPQ operator. Once the layer is fully connected, two convolutional neural networks are utilized to extract more training data, improve network parameters, and deliver classification results using the softmax function. Using this method, the recognition rates in the NIR and Yale face datasets are enhanced to 90% and 93.333%, respectively. The results of the studies demonstrate that the suggested method not only increases the model's tolerance for illumination, expression, and posture but also increases the accuracy of face recognition.

I. INTRODUCTION

Face information processing has a number of advantages, including a high safety coefficient, quick data collection, and easy popularisation in biometric feature recognition. Face recognition technology may be used for a variety of purposes in finance and security, including video surveillance and intelligent payment. It is presently one of the most popular research topics in machine learning and computer vision.

Face recognition is often used in unconstrained environments. Several environmental factors, including as illumination, emotion, posture, and so on, have an effect on the images collected. As a result, the standard feature extraction approach is ineffective for face recognition. Because Histogram of Oriented Gradient (HOG) [1] has great invariance to optical and geometric image distortion, and local phase quantization (LPQ) [2] has properties such as grayscale invariance, light insensitivity, and so on, they are utilized to get facial features for face identification.

These days, there are many face training data sets available, and deep learning has captured the interest of many. In order to perform face recognition, the neural network model may evaluate two-dimensional images, learn image features from a large number of samples, and categorize the two-dimensional images based on the learned characteristics. The abuse of subjective components can be avoided with the aid of machine feature extraction. Convolutional neural networks are frequently employed in face recognition applications due to its excellent error tolerance, parallel processing capability, self-learning capabilities, and generalization performance.

The convolutional neural network can recognize face characteristics directly from images, but it must first learn the image's noise. Both HOG and LPQ can analyze face photos and decrease noise interference, making image details more visible. Because HOG is primarily concerned with extracting the target's presentational and form characteristics, whereas LPQ is capable of extracting the target's texture features, LPQ has a stronger influence on extracting face features.

In this research, LPQ is used to extract facial texture features while compensating for illumination and expression. Meanwhile, the parallel ensemble learning technique and the parallel connection of two convolutional neural networks with different topologies for face recognition boost the diversity of the ensemble members as well as the network's generalization ability. This study provides a face recognition method based on simultaneous ensemble learning of LPQ and CNN that significantly improves recognition rate.

To evaluate the performance of our proposed CNN model in the face recognition issue, two datasets, NIR [3] and Yale [1], are employed. The experimental findings are contrasted with earlier methods in detail, and the results show that our approach works better than the others.

II. LITERATURE REVIEW

Xiaoyu Sun et al [4] introduced a unique method for detecting facial photos under uncontrolled conditions such as expressions, blur, occlusion, and illumination changes by combining the Zernike moment with multiscale patch-based local phase quantization. The Zernike moments are calculated around each pixel, and then the real and imaginary components are combined to generate double moment pictures. Texture information is then derived for non-overlapping patches of moment pictures using a multiscale patch-based local phase quantization descriptor. For picture categorization, a support vector machine is utilized. The novel's performance is evaluated on different databases.

Huu-Tuan Nguyen and Alice Caplier [5] developed a patch-based Local Phase Quantization (LPQ) of Monogenic components facial recognition model. The input image is used to generate directional monogenic bandpass elements, and each pixel of these bandpass elements is updated by the mean value of its neighborhood. On these images, LPQ histogram sequences are created, which are then concatenated to provide a global representation of the face image. To minimize dimensionality, it utilizes whitened principal component analysis, a k-nearest neighbor classifier, and weighted angle distance for classification. Experiments on the FERET and SCface databases show that the model is successful despite changes in expression, illumination, time lapse, and low resolution.

Borui Zhang et al [6] suggested a facial expression recognition system that makes use of the Gabor Wavelet Transform, the Local Binary Pattern (LBP), and LPQ. The Gabor filter is used to extract the significant aspects of the facial image across five scales and eight orientations. The LBP and LPQ operators are then used to encode the Gabor image. A two-stage principal component analysis and linear discriminant analysis approach is used to reduce the dimension of the fusion feature formed of the Gabor LBP and Gabor LPQ features. Performance is evaluated on the JAFFE database using SVM classifier in terms of recognition accuracy.

Qin-Qin Tao et al [7] developed a model for face detection utilizing a local Convolution Neural Network (CNN) and a Support Vector Machine (SVM) to handle the problem of detection under considerable appearance variations caused by perspective, intense lighting, and expression variations. Locality-sensitive SVM is used to generate a local model for each local area, making the classification work easier owing to lower within-class variance. Because local features are more robust than global features, discriminative local face features are learned by applying several local CNNs. The robustness and efficiency of this model are examined using the CMU+MIT dataset and the FDDB dataset.

Amin Jalali et al. [8] proposed a sensitive convolutional neural network with a sensitivity element in the cost function of a CNN to emphasize tiny changes and high frequency components in significantly blurred input picture samples. In CNN, the proposed cost function contains a sensitivity component in which the conventional error is divided by the activation function's derivative. This highlights the minute differences in the very blurred pictures, allowing for enhanced feature extraction and generalization and classification performance. The suggested sensitivity element's usefulness is validated using low illumination, quality deteriorated LDHF images obtained in long standoffs.

Shreyas Saxena and Jakob Verbeek [9] presented a heterogeneous face recognition system that uses CNNs to recognize facial pictures from many sensor modalities. In most cases, gallery images are conventional visible images, whereas probe images are infrared images. To decrease inequalities across disparate modalities, face characteristics from a CNN pre-trained on visible spectrum face photos are used for recognition with various learning approaches. Experiment results show that utilizing CNNs trained on visible spectrum images can outperform the state of the art for heterogeneous recognition using near-infrared images.

Xi Yin and Xiaoming Liu [10] presented the Multitask CNN for recognizing pose invariant faces with identity classification as main task and pose, illumination and expression estimates as sub tasks. Dynamic-weighting technique is deployed to dynamically apply loss weights to each sub task to eliminate the problem of balancing between different tasks during training, where higher loss weight is assigned to a simpler sub task. CNN learns definite identity features during training as well as a stochastic routing method for feature fusion during testing. On wild datasets, the efficacy of the method is tested, and the results show that it achieves superior performance than the current state of the art on LFW, CFP, and IJB-A datasets.

Yanan Wang et al [11] suggested a bi-directional Collaborative representation-based classification system for identifying facial images using CNN features. To extract face attributes from the original gallery and query sets, deep convolutional neural networks were utilised, and an efficient reverse representation model was used to retrieve the auxiliary residual information between each training sample and test sample. To eliminate the bi-directional optimization problem, the input sample is described by a forward linear combination and a backward one. Furthermore, for robust classification, the residuals from forward and reverse modelling are integrated at the lowest possible cost. The algorithm's validity is supported by experimental results obtained from well-known face databases such as AR, FERET, and ORL in terms of robustness to the small sample size problem.

Hana Ben Fredj et al. [12] developed a CNN framework to recognizing faces in unconstrained environment. In the framework, aggressive data augmentation is used for learning. Further adaptive fusion of softmax loss and center loss as supervised signals is added to improve the performance. The effectiveness of the model was tested on the LFW and You Tube datasets.

M. Chandrakala and P. Durga Devi [13] proposed a two stage classifier using HOG features to recognize face images. The desired features from the preprocessed images are extracted using HOG. At first, the k-NN classifier was used. Then, unrecognized face images were tested with the SVM classifier, which led to more accurate face recognition.

Htwe Pa Pa Win et al. [14] proposed method for face recognition using CNN. The deep learning strategies of the CNN are used for detecting face, extracting face features and for recognition. The effectiveness of this method was tested by experimentation on FEI dataset resulting in better accuracy and reduced time complexity.

Peng Lu et al. [15] developed a method for face recognition that combines CNN with augmented dataset. The small dataset is augmented to a larger one by transformations of the face images, such as flip, shift, scaling, and rotation. Face features extracted from these augmented datasets resulted in higher recognition rate. This method was tested on the ORL dataset.

Walid Hariri [16] introduced an efficient technique for recognizing masked faces. In this technique, first the masked face region is removed, then three pre-trained deep CNN namelyVGG-16, AlexNet, and ResNet-50 is applied to extract deep features. The classification is performed by multilayer perceptron. Experiment was conducted on Real-World-Masked-Face-Dataset to evaluate the performance.

III. METHODOLOGY

This research focuses on face recognition utilizing parallel ensemble learning with LPQ and CNN. In this study, CNN is used to first extract the facial features from the images processed by LPQ, then LPQ is used to analyse the texture of the input images, and finally, the parallel ensemble learning method is used to improve CNN's subpar generalisation performance brought on by the learning algorithm getting stuck in a local minimum, enhancing the effect of distinguishing different faces. Fig. 1 depicts the proposed approach's implementation flow, and the exact implementation technique is described in the following sections.

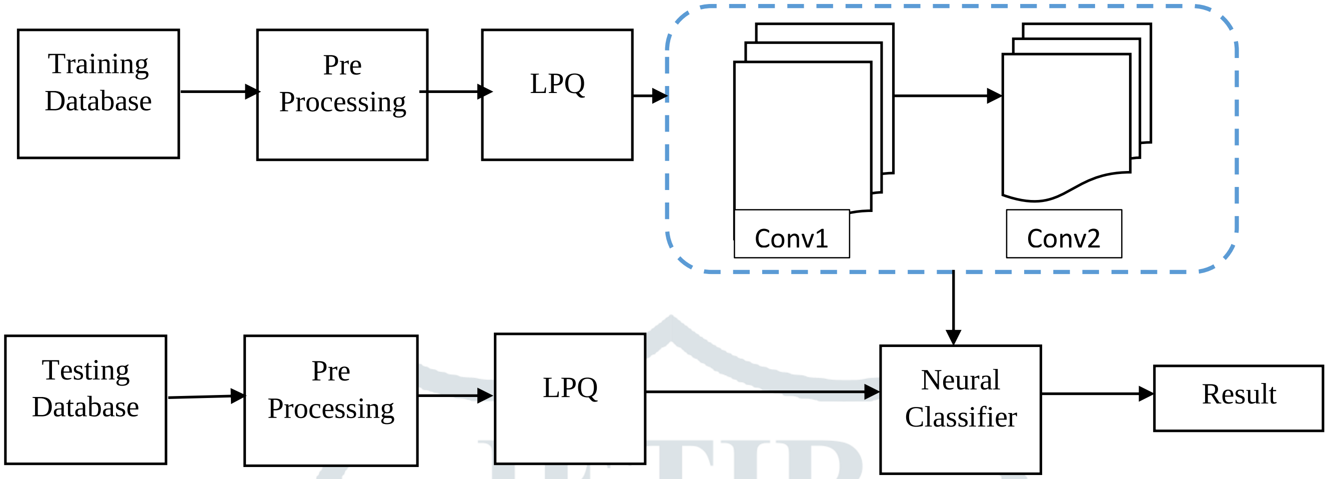

Fig.1.The block diagram of the proposed model

A. Training Face Database

Each database comprises a variety of various face images, each with its unique set of features. The YALE database is used to investigate the efficacy of the proposed algorithm under changes in position and facial expressions. It is made up of 15 participants, each with 11 samples. By selecting the first 9 samples from each participant, a total of 135 images are examined for training. The recognition rate is calculated using the 10th face image of the same 15 individuals. To investigate the efficacy of the proposed method under different conditions of intensity, light, position, blurring effect, and expressions, an expanded NIR database is used. It is made up of 41 participants, each with six samples. Out of 41 subjects, the first 35 are selected for training by selecting the first 5 samples of each topic, for a total of 175 images. The recognition rate is calculated using the sixth face image of the same 35 individuals.

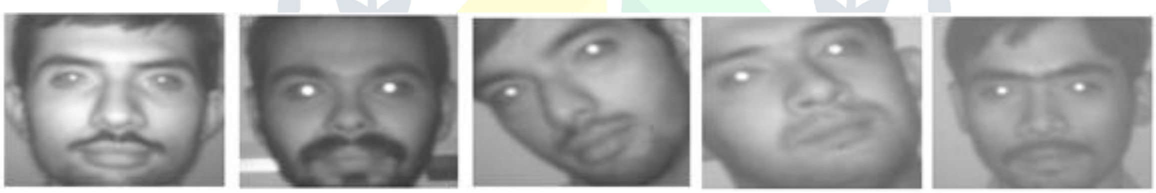

Fig 2. Sample Images from NIR database

Fig 3. Sample Images from Yale database

B. Preprocessing

Image preprocessing is a critical step since it removes unnecessary information. The input face picture is preprocessed to remove the face region from the complicated background using the viola jones technique [17]. Because the retrieved facial images have varied dimensions, they are scaled to 100×100 to ensure uniformity.

C. Local Phase Quantization (LPQ)

LPQ is introduced by Ojansivu, is a spatial blurring approach based on quantizing the Fourier transform phase in local areas. It is commonly understood that spatial blurring may be represented by a convolution of image intensity and a point spread function (PSF). This leads to a multiplication in the frequency domain.

$G(k) = F(k) * H(k)$ (1)

Where $G(k)$, $F(k)$ and $H(k)$ are the Fourier transforms of the blurred, original, and PSF images, respectively. When only the phase of the spectrum is considered, the relationship becomes a sum.

$\angle G(k) = \angle F(k) + \angle H(k)$ (2)

Where$\angle G(k)$,$\angle F(k)$ and $\angle H(k)$ denote the phase angles of$G(k)$, $F(k)$ and $H(k)$ If the PSF is centrally symmetric, i.e. $h(y) = h(-y)$the transform $H(k)$ becomes real valued, and the phase angle $\angle H(k)$ is represented by a two-valued function:

$$\angle H(k) = \begin{cases} 0 \; if \; H(k) > 0 \\ 1 \; if \; H(k) < 0 \end{cases} \qquad (3)$$

$\angle G(k) = \angle F(k)$ Exists for every$\angle H(k) \geq 0$. (u). Furthermore, the form of $H(k)$ for a regular PSF is similar to that of a Gaussian or a sinc-function, implying that at least the low frequency values of $\angle H(k)$are non-negative. $\angle G(k) = \angle F(k)$ At these frequencies, making $\angle F(k)$ a blur invariant characteristic.

To investigate the phase in LPQ, local M-by-M neighborhoods N_x at each pixel position of the picture $f(x)$ are used. A short term Fourier transform is used to get the local spectrum.

$F(k,x) = \sum_{y \epsilon N_x} f(y-x) e^{j2\pi k T_y}$ (4)

The transform is effectively assessed for all picture coordinates $x \epsilon \{\boldsymbol{x_1}, \boldsymbol{x_2} \dots \dots \boldsymbol{x_N}\}$ using simple 1-D convolutions for the rows and columns in a row-by-row fashion. The local Fourier coefficients are computed at four frequency locations,$k_1 = [a,0]^T$, $k_2 = [0,a]^T$, $k_3 = [a,a]^T$ and $k_4 = [a,-a]^T$, where a is a small enough scalar to fulfil $H(k_i) \geq 0$. This yields a vector for each pixel position:

$$F(x) = \genfrac{}{}{0pt}{}{([Re\{F(k_1,x), Im\{F((k_1,x)\}}{\dots\dots[Re\{F(k_4,x), Im\{F(F(k_4,x)\}])} \qquad (5)$$

By monitoring the signs of the real and imaginary portions of each component in $F(x)$, the phase information in the Fourier coefficients is recorded. This is accomplished by employing a simple scalar quantizer.

$$q_j = \begin{cases} 1 \; if \; f_j > 0 \\ 0 \; if \; f_j < 0 \end{cases}$$

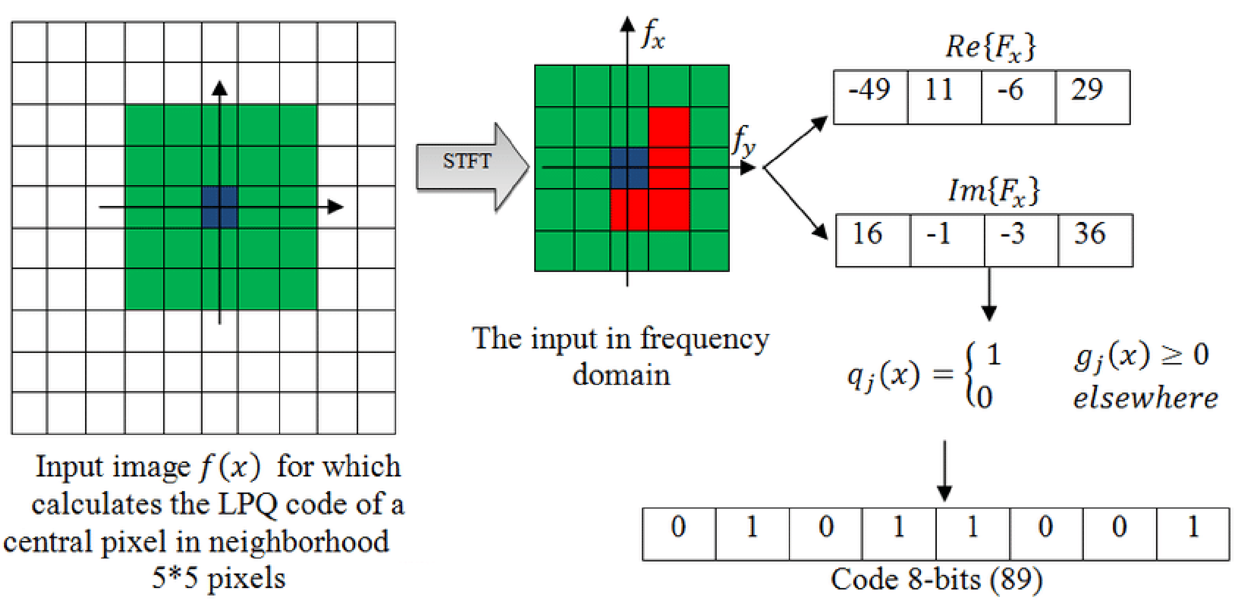

Fig 4. An example of LPQ operator

Where f_j is the vector j^{th} component $f(x)$. Using binary coding, the resulting eight binary coefficients q_j are represented as integer values ranging from 0 to 255.

$f_{LPQ}(x) = \sum_{i=0}^{7} q_j 2^i$ (6)

As a consequence, for face description, the label image f_{LPQ}whose values are the blur invariant LPQ labels is obtained. Fig 4 is an example of the technique used to compute the LPQ operator.

D. Convolutional Neural Network

CNNs are a type of deep neural network that is particularly successful for pattern recognition and classification. These CNNs are feed forward networks made up of multiple hidden layers. CNNs are made up of kernels with taught weights and biases. Each kernel takes some input and conducts a non-linear convolution function.

After LPQ texture extraction, CNN is used to extract face characteristics in the work. The network structure shown in Fig. 5 is one of the CNN models utilized in this article, and it is made up of two convolutional layers, two maximum pooling layers, and a final fully connected layer.

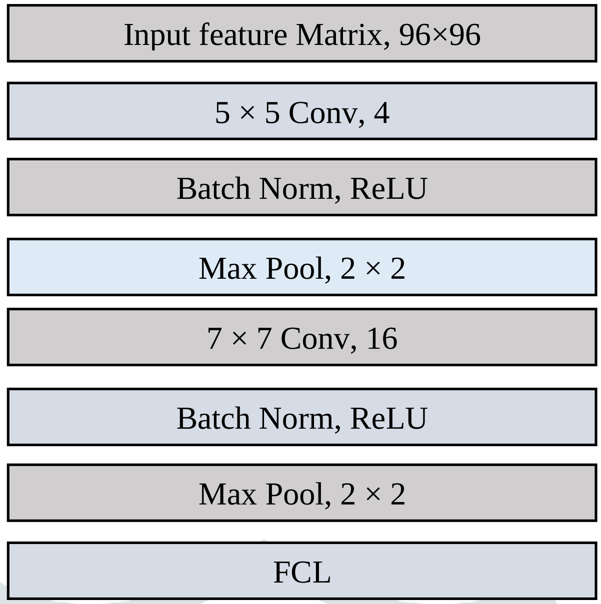

Fig 5. Proposed Convolutional Neural Network architecture

IV. Results and Discussions

The size of the each image feature matrix is modified to 96×96×1 after extracting the LPQ features from the input face image. We carried out several experiments by selecting different batch sizes. Because MATLAB 2018a has a CNN library, the proposed approach is implemented on it. The minimal batch sizes are 16, 32, and 64, and the recognition rate is calculated for each batch size. The network's learning rate is set at 0.001. It is the gradient decent optimization algorithm that is being employed. The face recognition system's performance was assessed using the NIR, and YALE databases.

YALE dataset

Experiment is carried out using batch sizes of 16, 32, and 64. It has been discovered that a batch size of 64 results in greater recognition accuracy. The Accuracy and loss (error) graphs produced for both training and testing (validation) features are shown in Fig 6 and Fig 7. Table 1 tabulates the data for validation accuracy and validation loss for various epochs. From epoch 12, the system accuracy exceeds 90% and achieves 93.33 percent with a minimal validation loss of 0.1493 and minimal batch loss of 0.0022 over 18 epochs.

NIR dataset

The experiment is carried out using batch sizes of 16, 32, and 64. It is discovered that a batch size of 16 results in greater recognition accuracy. The Accuracy and Loss (Error) Graphs acquired for both training and testing (validation) features are shown in Fig 8 and Fig 9. The numbers calculated for validation accuracy and validation loss for different epochs are shown in Table 2. From epoch 8 onwards, the system accuracy exceeds 90%, peaking at 90.00 percent with a minimal validation loss of 0.5961 for 11 epochs.

Table 1. Shows the values tabulated for YALE database for different epochs

Epoch	Iteration	Validation Accuracy in %	Validation Loss	Mini-Batch Accuracy in %	Mini-Batch Loss
1	1	6.67	4.4160	6.25	3.5973
1	2	20.00	5.7585	4.69	5.4045
2	4	20.00	5.6948	20.31	4.9251
3	6	46.67	4.3700	34.38	4.7479
4	8	53.33	4.9414	70.31	3.8493
5	10	60.00	3.7552	67.19	3.0359
6	12	66.67	2.0801	87.50	1.4859
7	14	73.33	1.0447	92.19	0.5495
8	16	80.00	0.8714	95.31	0.1703
9	18	80.00	0.6580	95.31	0.1396
10	20	80.00	0.3142	96.88	0.0889
11	22	80.00	0.1632	96.88	0.0529
12	24	93.33	0.1282	98.44	0.0183
13	26	100.00	0.1192	100.00	0.0183
14	28	100.00	0.1195	98.44	0.0464
15	30	93.33	0.1291	100.00	0.0080
16	32	93.33	0.1368	100.00	0.0036
17	34	93.33	0.1434	100.00	0.0047
18	**36**	**93.33**	**0.1493**	**100.00**	**0.0022**

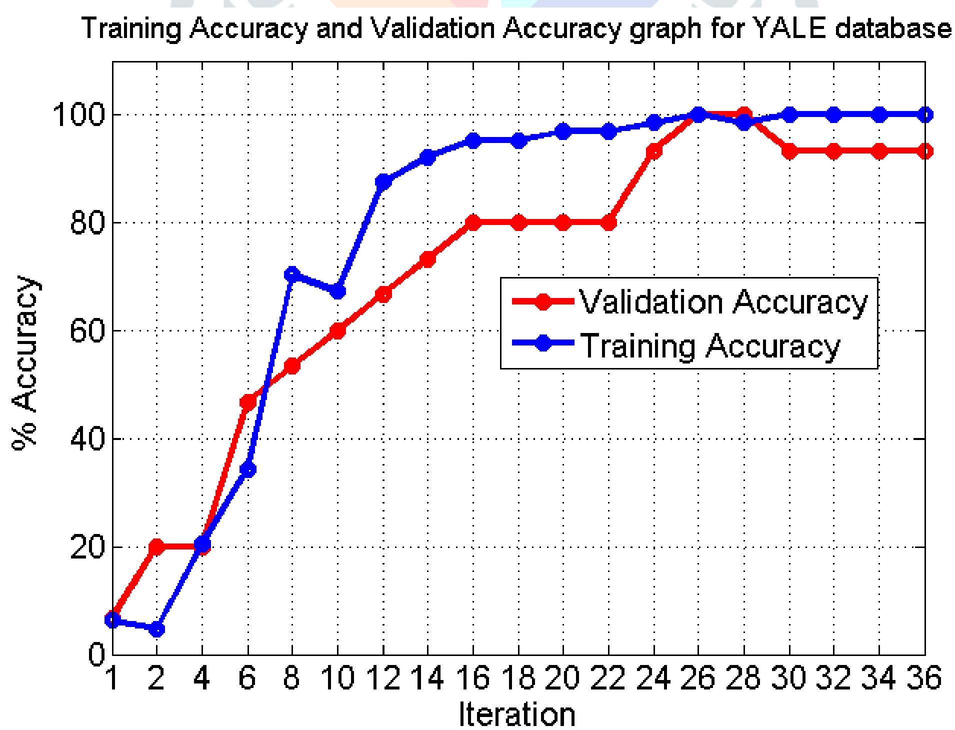

Figure 6. Shows Training and Testing Accuracy graphs for Yale database

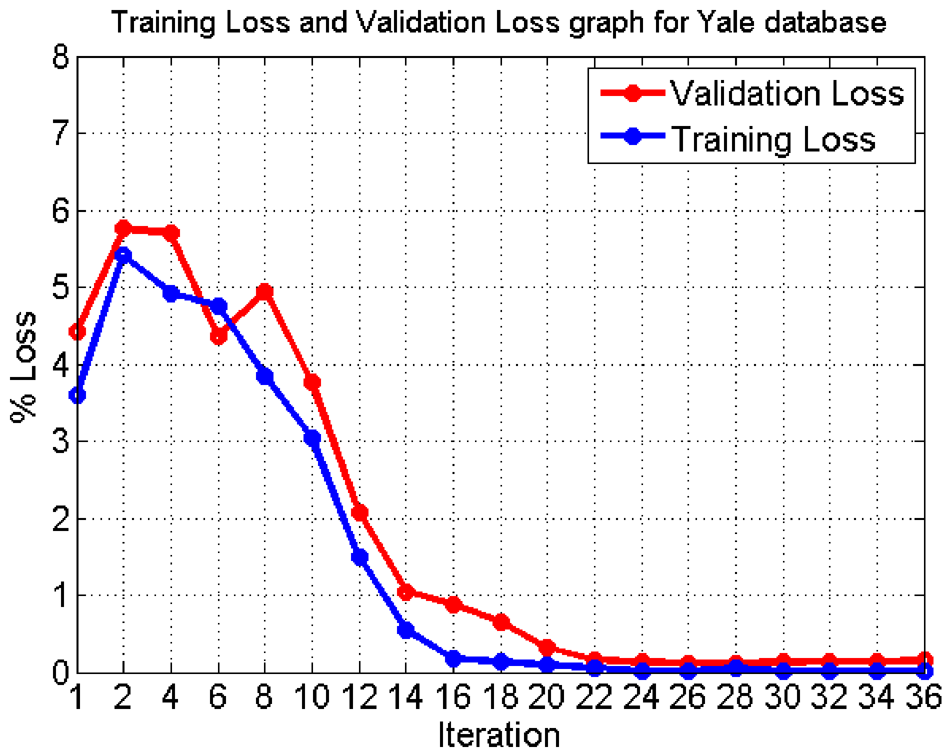

Figure 7. Shows Training and Testing Error (Loss) graphs for Yale database

Table 2. Shows the values tabulated for NIR database for different epochs.

Epoch	Iteration	Validation Accuracy in %	Validation Loss	Mini-Batch Accuracy in %	Mini-Batch Loss
1	1	0.00	7.8543	0.00	4.4893
1	12	32.50	5.2099	12.50	5.2604
2	24	65.00	2.2228	62.50	2.2169
3	36	87.50	0.9761	81.25	0.7669
4	48	87.50	0.7629	93.75	0.2337
5	60	87.50	0.6082	100.00	0.0861
6	72	90.00	0.5218	93.75	0.3059
7	84	85.00	0.6648	100.00	0.0670
8	96	90.00	0.7603	100.00	0.0405
9	108	90.00	0.7431	100.00	0.0983
10	120	90.00	0.6258	93.75	0.6258
11	**132**	**90.00**	**0.5961**	**100.00**	**0.5961**

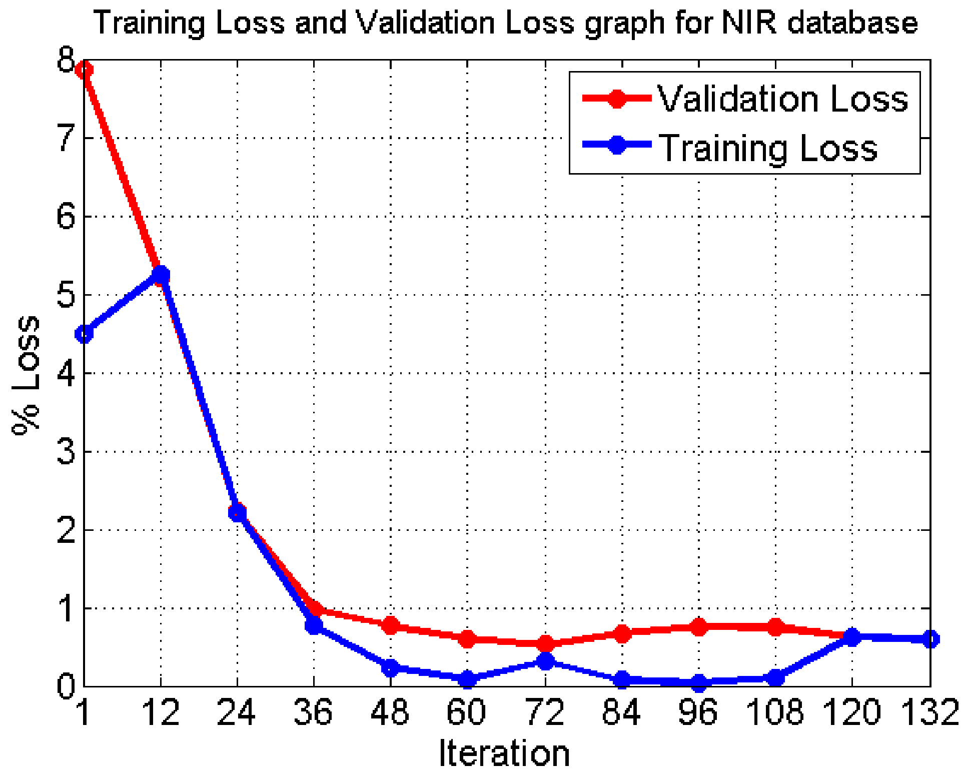

Figure 8. Shows Training and Testing Accuracy graphs for NIR database

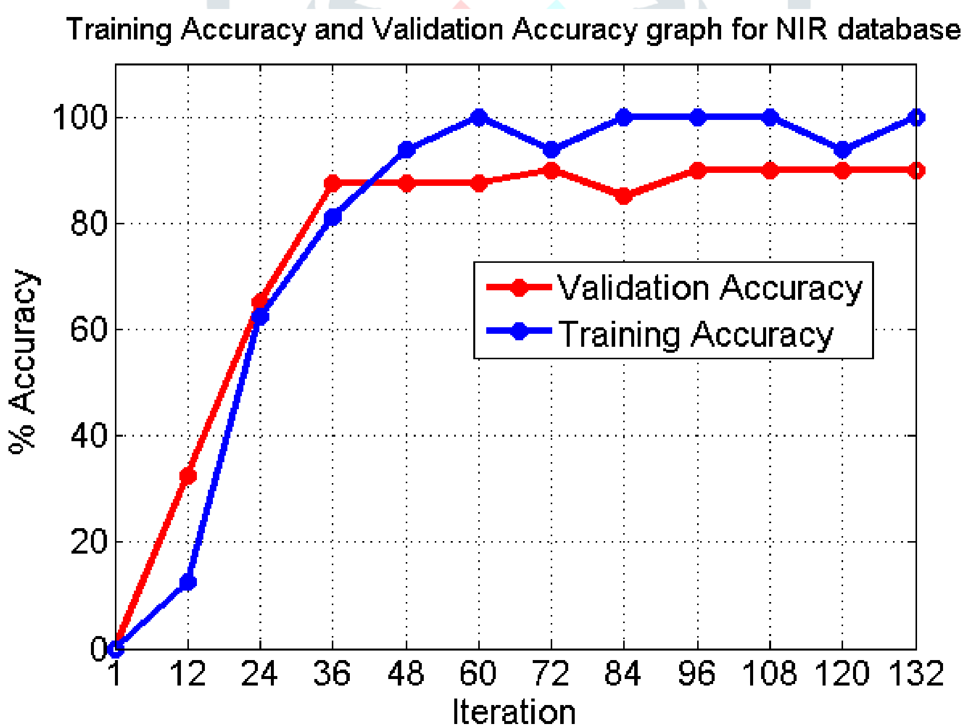

Figure 9. Shows Training and Testing Error (Loss) graphs for NIR database

Table 3. Shows the comparison of proposed work with existing work.

Author-Year	Database	Technique	Classifier	Accuracy
Farid Ayeche et al., 2020 [18]	YALE	HDGG	SVM	92.12%
Srinivas Halvi et al [19]	NIR	1DTD	FFBP NN	72%
Shreyas Saxena et al.,2016 [20]	NIR	CNN	Softmax	85.9%
Sateesh Kumar H C et al., 2017 [21]	NIR YALE	STWT DTCWT STWT DTCWT	ED	86.67% 90.24% 87.5% 87.5%
Proposed	**YALE** **NIR**	**LPQ**	**CNN**	**93.33%** **90.00%**

Experimental results obtained for the proposed work is compared with previous methods and it is shown Table 3.It is found that the combination of LPQ and CNN yields better recognition rate for NIR and YALE database when compared other existing techniques.

V. Conclusion

The research introduces a face recognition strategy based on ensemble learning of LPQ and CNN that uses LPQ to extract texture characteristics as training data for the CNN and is then applicable to face identification. Because LPQ may lessen the impact of illumination on facial characteristics and enhance face recognition accuracy, LPQ features are mostly used to extract facial texture features. The collected features are scaled as a 96×96 matrix and transferred to the four layer CNN structure for training, which consists of convolutional, max pooling, convolutional, max pooling layers. Face image categorization is performed by the softmax layer. Performance analysis is measured in terms of recognition accuracy. It is found that YALE and NIR datasets yields better recognition rate of 93.33% and 90.00% respectively.

References

[1] Navneet Dalal and Bill Triggs.2015. "Histograms of oriented gradients for human detection," Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893.

[2] Wang Zhe and Ying Zilu, "Facial Expression Recognition Based on Local Phase Quantization and Sparse Representation," International Conference on Natural Computation, pp. 222-225.

[3] Ran He, Xiang Wu and Tieniu Tan.2019. "Wasserstein CNN: Learning Invariant Features for NIR-VIS Face Recognition," Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1761-1773.

[4] Xiaoyu Sun, Xiaoyan Fu, Zhuhong Shao, Yuanyuan Shang and Hui Ding.2016. "Local Zernike Moment and Multiscale Patch-Based LPQ for Face Recognition," Proceedings of 2016 Chinese Intelligent Systems Conference, Vol. 405, pp. 19-27.

[5] H. Nguyen and A. Caplier.2014. "Patch based local phase quantization of monogenic components for face recognition," International Conference on Image Processing, pp. 229-233.

[6] Borui Zhang, Guangyuan Liu and Guoqiang Xie.2016. "Facial expression recognition using LBP and LPQ based on Gabor wavelet transform," International Conference on Computer and Communications, pp. 365-369.

[7] Qin-Qin Tao, Shu Zhan, Xiao-Hong Li and Toru Kurihara.2016. "Robust face detection using local CNN and SVM based on kernel combination," Neurocomputing, vol. 211, pp. 98-105.

[8] Amin Jalali, Rammohan Mallipeddi and Minho Lee.2017. "Sensitive deep convolutional neural network for face recognition at large standoffs with small dataset," Expert Systems With Applications, vol. 87, pp. 304-315.

[9] Shreyas Saxena and Jakob Verbeek.2016. "Heterogeneous Face Recognition with CNNs," Computer Vision ECCV 2016 Workshops, Lecture Notes in Computer Science, Springer, Cham, vol. 9915. pp. 483-491.

[10] Xi Yin and Xiaoming Liu.2018. "Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition," IEEE Transactions on Image Processing, vol. 27, no. 2, pp. 964-975.

[11] Yanan Wang, Tian Na, Xiaoning Song and Guosheng Hu.2018. "Bi-directional CRC algorithm using CNN based features for face classification," Asian Conference on Artificial Intelligence Technology, Vol.2018 ,Issue 16 p. 1457 – 1462.

[12] Hana Ben Fredj, Safa Bouguezzi and Chokri Souani.2021. "Face Recognition in Unconstrained Environment with CNN," The Visual Computer, vol. 37, pp. 217–226.

[13] M. Chandrakala and P. Durga Devi.2021. "Two-stage Classifier for Face Recognition using HOG Features," Materials Today: Proceedings, vol. 47, pp. 5771-5775.

[14] Htwe Pa Pa Win, Phyo Thu Thu Khine and Khin Nwe Ni Tun.2021. "Face Recognition System based on Convolution Neural Networks," International Journal on Image, Graphics and Signal Processing, vol. 6, pp. 23-29.

[15] Peng Lu, Baoye Song and Lin Xu.2021. "Human face recognition based on convolutional neural network and augmented dataset," Systems Science & Control Engineering, vol. 9, pp. 29-37.

[16] Walid Hariri, "Efficient masked face recognition method during the COVID-19 pandemic.2021." Signal, Image and Video Processing, vol. 16, pp. 605–612.

[17] Bobby R Bruce, Jonathan M. Aitken and Justyna Petke.2016. "Deep Parameter Optimization for Face Detection Using the Viola-Jones Algorithm in Open CV," Springer International Publishing, pp. 238–243.

[18] Farid Ayeche and Adel Alti.2021. "HDG and HDGG: an extensible feature extraction descriptor for effective face and facial expressions," Pattern Analysis and Applications, vol. 24, pp. 1-16.

[19] Halvi, S, Ramapur, N, Raja, K.B and Prasad, S.2017. "Fusion Based Face Recognition System Using 1D Transform Domains," International Conference on Advances in Computing and Communications, Procedia Computer Science, vol. 115, pp.383–390.

[20] Shreyas Saxena and Jakob Verbeek.2016. "Heterogeneous Face Recognition with CNNs," ECCV Workshops, pp. 483-491.

[21] Sateesh Kumar H C, C Chowda reddy, Raja K B and Venugopal K R.2017. "Face Recognition based on STWT and DTCWT using two dimensional Q-shift Filters," International Journal of Engineering Research and Application, vol. 7, pp. 64-69.