Face recognition from image patches using an ensemble of CNN-local mesh pattern networks

Renjith Thomas
Electronics and Communication Engineering Departement.
Musaliar College of Engineering and Technology
Pathanamthitta, India
renjiththomasdr@gmail.com

2021 6th International Conference for Convergence in Technology (I2CT) | 978-1-7281-8876-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I2CT51068.2021.9418138

***Abstract*—Face recognition is one of the superior methods among the biometric identification system. Demand for face recognition is increased recently because of the availability of data from security and surveillance cameras. This work proposes a face recognition method using an ensemble of convolutional neural network trained on original and the textural feature maps from image patches. The original face image divided into smaller patches and a textural descriptor is applied on them. The textural feature map image of all patches and the original patches are input to two pre-trained ResNet50 networks. Input images are used to fine-tune the networks. The predictions of the two networks are combined by majority voting ensemble technique to get the final labels. The proposed system achieves an accuracy of 98.05%. The system performance is also evaluated using different measures like precision, recall, and F-measure.**

***Index Terms*—Face recognition, Image patches, Local Mesh Pattern, CNN, Ensemble.**

I. Introduction

Biometric identification includes iris recognition, voice, face, and thumb identification. Among biometric monitoring, face identification is more significant because of its uniqueness in identification. Face images can be obtained from videos from security and surveillance cameras. These images are applied to different pattern recognition algorithms to obtain unique features for identification. One of the main difficulty in the face recognition process is to match an unfamiliar face to a collection of known faces.

The effect of Covid-19 pandemic in the world has forced to use face masks and hand gloves. Hence public biometric recognition is difficult in this scenario. There is an urge to identify a person from a masked face where a small region of the face is only exposed. This lead researchers to develop systems that can recognize a person from a small patch of the face image. Ghorbani et al. proposed a face recognition system using HOG and LBP features [1]. The final performance is evaluated on the reduced feature set obtained by a feature selection technique. Skin region segmentation for face recognition using RGB-YCbCr colour model with kNN classifier is proposed by Chelali et al. [2]. Abhishree et al. proposed a system using features extracted from Gabor filters for face identification [3]. Binary particle swarm optimization algorithm is used for feature selection, and the selected optimal set is used for the recognition task.

With the advancement in deep learning techniques, many kinds of research are reported in the face recognition task using CNN. A deep framework with a set of CNNs and stacked auto-encoders is proposed by Ding et al. [4]. The auto-encoder compresses feature vector from CNN, and results show that the architecture gives an accuracy of 99.0%. A hybrid face recognition system is proposed by Rikhtegar et al., where optimal CNN architecture is selected using genetic algorithm [5]. An ensemble of SVM classifiers does the final classification. In [6], an eight layer CNN network is proposed for face recognition. A softmax layer is used as the final classification layer in the network. A CNN-SVM system for face recognition is proposed by Guo et al. [7]. Features from CNN are extracted and are classified using SVM classifier for the process.

Combination of different feature descriptors with CNN has been applied in different fields as it shows better accuracy. CNN and LBP fusion-based face recognition is proposed by Yang et al. [8]. Facial expression features from CNN are fused with the rotation-invariant features from LBP for the recognition. LBP feature map is applied to CNN for face recognition in [9]. Results show that by the combination of LBP with CNN, the performance is improved.

The important contribution of the work is that the proposed system achieves high accuracy in face recognition from patches. The network used here is ResNet50 pre-trained on ImageNet [10]. The patches are obtained from original images, and its textural features are enhanced using a local mesh pattern (LMeP). Both the original patch and the textural feature map patches are applied to two ResNet50 models and are ensembled for classification. The system performance is also compared with the ensemble of CNN with standard LBP based model [11].

Paper is arranged as follows: section 2 shows the proposed system, section 3 explains the experimental results and the paper concludes in section 4.

II. Proposed System

The proposed system demonstrates the performance of an ensemble of pre-trained ResNet50 using original and textural face images for face recognition. The face images are split into smaller patches which are selected randomly. Each patch is then applied to LMeP for textural analysis. Original patch

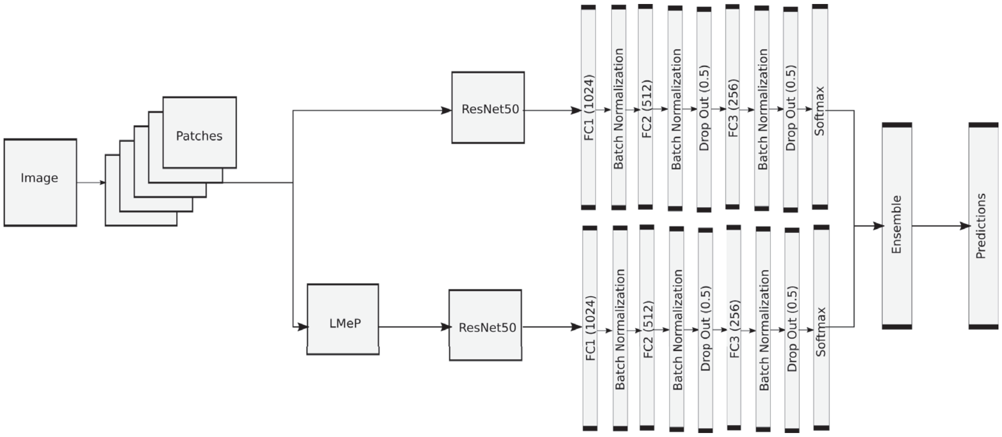


Fig. 1: Proposed system block diagram

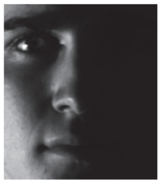


Fig. 2: Sample face image in the extended Yale-B face dataset.

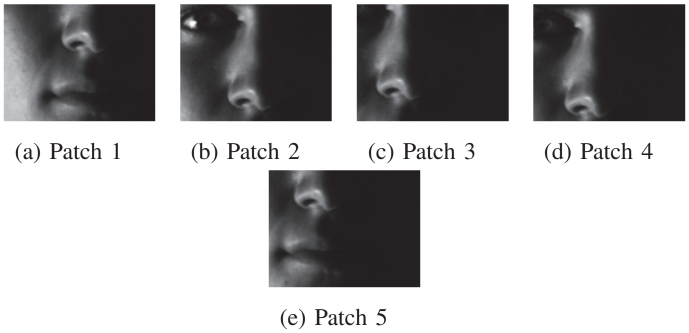
images and the textural images are given to two ResNet50 models with modified fully connected layers for face recognition. The output of these networks is then ensembled to get the final prediction. Fig. 1 presents the block diagram of the proposed system.

A. Image data set

The extended Yale-B face dataset [12] is adopted in this work to carry operations with different lighting conditions. The dataset has images of 38 individuals which includes ten cases of original Yale B dataset. Each class has images with 9 poses and 64 different lighting conditions. The dataset with images cropped to a dimension of 192 X 168 pixels showing only the face region is available for download. This dataset is used for the identification process in this work. Fig. 2 displays a sample face image in the extended Yale-B face dataset.

B. Image patch generation

One of the main issue faced in face recognition is the existence of occlusions. The work uses patches selected randomly from the face image to handle this condition. An image selected from the extended Yale-B dataset is split into five patches with resolution 100 X 130 pixels. These five patches will help the machine to learn five times more. The patches of the original image shown in Fig. 2 are shown in Fig. 3. In a face image, more information is present at the centre of the region of interest. As every patch has a resolution of 100 X 130, more details of the face come in every patch.



(a) Patch 1
(b) Patch 2
(c) Patch 3
(d) Patch 4
(e) Patch 5

Fig. 3: Patches of sample face image in the extended Yale-B face dataset.

C. Local mesh pattern (LMeP)

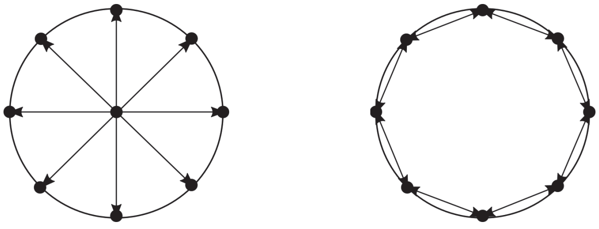
LBP was proposed by Ojala et al. [11] for the classification of texture in images. It is an excellent textural descriptor where the gray value of the centre pixel of a neighbourhood in an image is compared with its neighbours. If the centre pixel value is higher than the value of the neighbour pixel, it is taken as ’0’, otherwise as ’1’. LBP gives an 8-bit number and is converted to its corresponding decimal value which replaces the centre pixel. The equations for LBP are shown in equations 1 and 2.

$$LBP_{(M,N)} = \sum_{i=1}^{M} 2^{(i-1)} F(X_{i|N} - X_c) \tag{1}$$

$$F(t) = \begin{cases} 1, & t \geq \quad 0 \\ 0, & else \end{cases} \tag{2}$$

where X_c is the centre pixel of the neighbourhood, $X_{i|N}$ is the neighbouring pixels of radius N, and M is the number of neighbours from the centre pixel at a distance N.

The modified LBP feature descriptor is the LMeP descriptor which is used for face recognition in this work [13]. The value of LMeP is calculated based on the correlation between the neighbours of a neighbourhood. LBP and LMeP patterns used in this work are as shown in Fig. 4. From the pattern, it can be



(a) (b)

Fig. 4: Circular pattern of (a) LBP (b) LMeP

7	5	62
56	19	42
15	8	23

-12	-14	43
37		23
-4	-11	4

0	0	1
1		1
0	0	1

8	4	2
16		1
32	64	128

	147	

(a) LBP

7	5	62
56	19	42
15	8	23

-49	-2	57
41		-20
7	-15	-19

0	0	1
1		0
1	0	0

8	4	2
16		1
32	64	128

	50	

(b) LMeP

Fig. 5: Example of LBP and LMeP

seen that LBP finds the correlation between the centre pixel and the surrounding pixels, while LMeP finds the correlation between neighbouring pixels. An example for LBP and LMeP computation for a 3 X 3 pattern is illustrated in Fig. 5a and Fig. 5b respectively.

The equations for LMeP is shown in equation 3.

$$LMeP_{(M,N)} = \sum_{i=1}^{M} 2^{(i-1)} F(X_{\beta|N} - X_{i|N}) \tag{3}$$

$$\beta = 1 + mod((i + M + j - 1), M) \tag{4}$$

where $mod(x, y)$ gives the remainder of $\frac{x}{y}$ operation and j gives the LMeP index. In the work, j is taken as 1.

M=8 is used for the experiment in this work. As shown in the figure, the correlation between pixels in the neighbourhood is taken for texture analysis. Experimental outcomes illustrate that the LMeP provides more reliable features as compared to LBP, showing that it can gain more edge data than LBP for face image recognition.

D. CNN models

ResNet50 and MobileNet models pre-trained on ImageNet are used in this work. ResNet50 architecture is stacked with 50 layers. To reduce vanishing gradient and degradation problem, ResNet uses residual learning block [14]. MobileNet uses depthwise separable convolutional blocks [15]. Each block have 3 X 3 convolutional filters. Flattened output from the network is applied to fully connected layers. Three Fully Connected (FC) layers are used in the work with 1024, 512, and 256 neurons respectively. All dense layers are followed by a batch normalization layer. Batch normalization standardizes the information on a layer to all mini-batch. It stabilizes the learning method and reduces the number of epochs needed to train the networks. To avoid overfitting of the network, drop out layers with drop out rate of 0.5 is included between the dense layers. In the training period, weights of all layers of the pre-trained network is changed with respect to the input images. Lastly, during the examination stage, to classify the face images, a softmax activation layer is used.

E. Ensemble

There are many different ensemble techniques such as majority voting, bagging, boosting, and stacking. Majority voting technique [16] is used in the work. In this, the predictions from the models are counted and the label that is voted maximum is assigned as the final label of the test set.

III. Results and discussion

Extended Yale-B face dataset is used for this work. The original cropped image is of pixel size 192 X 168. These images are divided randomly into five overlapping patches of pixel size 100 X 130. In this work, 33 individuals with 65 images are considered for recognition. A total of 10725 patches are used for training and testing purposes. Image patches are split into training and testing set in the ratio of 0.8 and 0.2, respectively. The training set is further split in a ratio 0.2 for getting the validation dataset. Experiments are done with Google Colab [17] with 25 GB RAM and GPU.

The image patches are applied on two networks: MobileNet and ResNet50. The network is retrained with all layers kept trainable for 10 epochs. The network shows good training and validation accuracy with 10 epochs, above which the system overfits. Adam optimizer with a learning rate of 0.0001, β_1 and β_2 of 0.9 and 0.999 respectively are used for training. Batch size of 8 is used in work. Softmax classifier and categorical cross-entropy loss function are used in the last layer.

Training, validation, and testing accuracies of ResNet50 and MobileNet are shown in Table I. Figure 6 shows the training accuracy and loss of both the networks with epochs. From the results, it is seen that ResNet50 show high accuracy of 95.27% with the patches. ResNet50 is considered for further analysis as it shows better accuracy than MobileNet on original patch images.

For analyze the performance of textural images in the classification process, LBP and LMeP descriptors are applied on the patches individually. Fig.7 demonstrates the feature diagrams obtained by applying the LBP and LMeP operator on a sample face image. The performance of ResNet50 model with LBP and LMeP feature map as input is analyzed and these are denoted as LBP-CNN and LMeP-CNN respectively. For improve the accuracy of the recognition process, an ensemble of two CNN is done in this work. ResNet50 model trained on original patches and LMeP-CNN model is used to get the ensemble model. Predictions from both the models are counted to get the maximum counting label. The label that is counted maximum is taken as the final prediction of the test set. For

TABLE I: Training, validation, and testing accuracies of CNN models with original patches

Model	**Optimizer**	**Epochs**	**Training Accuracy**	**Validation Accuracy**	**Testing Accuracy**
MobileNet	Adam	10	92.73	88.34	88.26
ResNet50	Adam	10	98.73	95.09	95.27

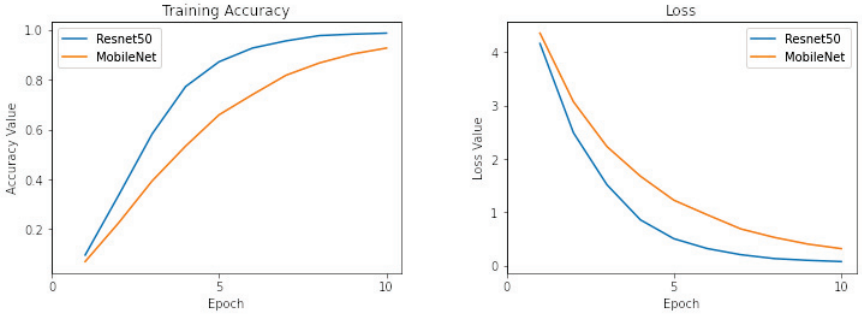
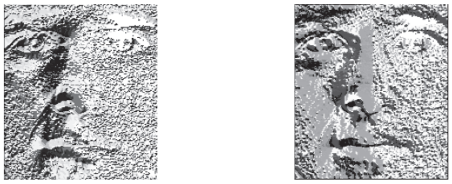


Fig. 6: Training accuracy and loss plot of CNN models



(a)

(b)

Fig. 7: Feature maps of LBP and LMeP

comparison, the performance of the ensemble of ResNet50 trained on original patch and LBP-CNN is also evaluated.

The system performance is evaluated using accuracy, precision, recall, and F-Measure. Expression for different performance measures is given in equations (5)-(8).

$$Accuracy = \frac{(T_{pv} + T_{nv})}{(T_{pv} + T_{nv} + F_{pv} + F_{nv})} \tag{5}$$

$$Precision = \frac{T_{pv}}{(T_{pv} + F_{pv})} \tag{6}$$

$$Recall = \frac{(T_{pv}}{(T_{pv} + F_{nv})} \tag{7}$$

$$F - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \tag{8}$$

where,

T_{pv}=	True	positive
T_{nv}=	True	negative
F_{pv}=	False	positive
F_{nv}=	False	negative

Table II shows the performance of LBP-CNN, LMeP-CNN, and ensemble models on the patches. The ROC plot of the two ensemble models is shown in Fig. 8. From the ROC plot, it is seen that the ensemble of LBP model and LMeP model shows AUC values of 0.98 and 0.99, respectively. The results show that the ensemble of CNN with LMeP-CNN model outperforms other models in the patch-based face recognition process. The proposed system gives high values for accuracy, precision, recall, and F-measure.

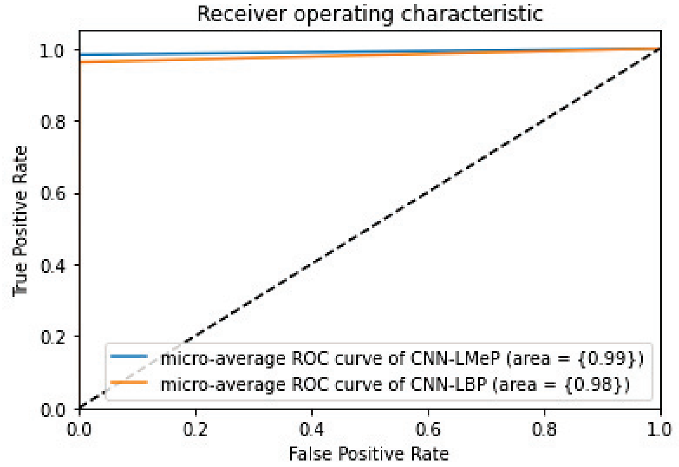


Fig. 8: ROC plot of ensemble models

IV. Conclusion

Face recognition is one of the important techniques among different biometric identifications. This paper presents a face recognition system by an ensemble of ResNet50 applied with original and local texture feature images. The face images are split into overlapping patches. Each patch is then converted into a textural feature map by applying local mesh descriptor on each neighbourhood. The feature map images of patches and the original patch images are applied as input image set to CNN networks.

TABLE II: Comparison of the proposed system with other methods

Model	Accuracy	Precision	Recall	F-measure
LBP-CNN	84.01	85.00	84.00	84.00
LMeP-CNN	90.21	91.00	90.00	90.00
Ensemble of CNN + LBP-CNN	95.87	96.00	96.00	96.00
Ensemble of CNN + LMeP-CNN	98.05	98.00	98.00	98.00

ResNet50 pre-trained on ImageNet is used in the study. The last fully connected layers of the networks are modified. The weights of the network are adjusted by using the cross-entropy error function with adam optimizer. The network is trained for 10 epochs with a batch size of 8. The ensemble of the trained models shows an accuracy of 98.05% in the classification process. The system performance is also evaluated in terms of precision, recall, and F-measure.

References

[1] M. Ghorbani, A. T. Targhi, and M. M. Dehshibi, "Hog and lbp: Towards a robust face recognition system," in *2015 Tenth International Conference on Digital Information Management (ICDIM)*. IEEE, 2015, pp. 138–141.

[2] F. Z. Chelali, N. Cherabit, and A. Djeradi, "Face recognition system using skin detection in rgb and ycbcr color space," in *2015 2nd World Symposium on Web Applications and Networking (WSWAN)*. IEEE, 2015, pp. 1–7.

[3] T. Abhishree, J. Latha, K. Manikantan, and S. Ramachandran, "Face recognition using gabor filter based feature extraction with anisotropic diffusion as a pre-processing technique," *Procedia Computer Science*, vol. 45, pp. 312–321, 2015.

[4] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015.

[5] A. Rikhtegar, M. Pooyan, and M. T. Manzuri-Shalmani, "Genetic algorithm-optimised structure of convolutional neural network for face recognition applications," *IET Computer Vision*, vol. 10, no. 6, pp. 559–566, 2016.

[6] M. Coşkun, A. Uçar, Ö. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *2017 International Conference on Modern Electrical and Energy Systems (MEES)*. IEEE, 2017, pp. 376–379.

[7] S. Guo, S. Chen, and Y. Li, "Face recognition based on convolutional neural network and support vector machine," in *2016 IEEE International Conference on Information and Automation (ICIA)*. IEEE, 2016, pp. 1787–1792.

[8] X. Yang, M. Li, and S. Zhao, "Facial expression recognition algorithm based on cnn and lbp feature fusion," in *Proceedings of the 2017 international conference on robotics and artificial intelligence*, 2017, pp. 33–38.

[9] P. Ke, M. Cai, H. Wang, and J. Chen, "A novel face recognition algorithm based on the combination of lbp and cnn," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*. IEEE, 2018, pp. 539–543.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[11] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[12] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[13] S. Murala and Q. J. Wu, "Local mesh patterns versus local binary patterns: biomedical image indexing and retrieval," *IEEE journal of biomedical and health informatics*, vol. 18, no. 3, pp. 929–938, 2013.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[16] X. Zhou, S. Wang, H. Chen, T. Hara, R. Yokoyama, M. Kanematsu, and H. Fujita, "Automatic localization of solid organs on 3d ct images by a collaborative majority voting decision based on ensemble learning," *Computerized Medical Imaging and Graphics*, vol. 36, no. 4, pp. 304–313, 2012.

[17] B. M. Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, "Using the jupyter notebook as a tool for open science: An empirical study," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 2017, pp. 1–2.