Advances in Consumer Research
ISSN (Print): 0098-9258
ISSN(Online): 3079-1766

Vol. 2, Issue 4 (2025) *https://acr-journal.com/*

Face Recognition using Local Feature Descriptors and Convolutional Neural Networks

Sneha D P[1] and Dr. Vasudev T[2]

[1]Research Scholar, Department of Computer Science, University of Mysore, Mysuru, Karnataka, India
Email-sneha.p012@gmail.com

[2]Professor, Department of MCA, Maharaja Institute of Technology Mysore, Karnataka, India
Email-vasu@mitmysore.in

Cite this paper as: Sneha D P, Dr. Vasudev T, (2025) Face Recognition using Local Feature Descriptors and Convolutional Neural Networks. *Advances in Consumer Research*, 2 (4), 3382-3392

KEYWORDS

Face Recognition, Histogram of Gradients, Local Binary Pattern, Convolution Neural Network.

ABSTRACT

Face recognition is normally used in automated surveillance, individual identification, and database searches for specific faces. Face detection, representation, and matching are the different stages of the face recognition process. The face detection starts from the query image, and then features are retrieved using a face recognition algorithm in the next stage. Matching the query face with the database is the final stage. However, face recognition algorithms perform low in unrestricted environments such as those with variations in an individual's lighting, posture, and facial expressions. This paper proposes a face recognition system designed to address these challenges using Convolutional Neural Networks (CNNs), Local Binary Pattern (LBP) histograms, and Histogram of Oriented Gradients (HOG). Initially, face detection from the input image is accomplished using the Viola-Jones technique. The feature space is created through fusing the features that were extracted using CNN, HOG, and the LBP histogram. SVM and KNN classifiers are used to assess the classification ability for various HOG cell sizes.

1. INTRODUCTION

The hardware and software technologies have advanced significantly in the past two decades. Due to this advancement, information technologies like Artificial Intelligence and Machine Learning have grown rapidly. These technologies employ current gadgets to build more effective and comfortable methods of human-computer interaction. Computer vision technology focuses on replicating or simulating visual perception, which is one area of machine intelligence. Applications for computer vision systems have included automated industrial quality control and assembly line inspections. As the cost of computer systems and video image-collecting technology has dropped, computer vision technologies have grown to more sophisticated vision applications including facial recognition and facial tracking techniques. Due to scaling and illumination issues, face recognition in the computer vision domain has remained a challenge to date. The other challenges with face detection are facial expression recognition and face authentication. Problems have been solved traditionally using the segmentation approach, facial feature detection, and face verification in complex contexts.

The difficulties of face detection are exacerbated by variations in size, location, orientation, posture, facial expression, occlusion, and illumination. A sub problem of the broader field of visual object tracking under computer vision studies is face tracking. A good number of research has been noticed on object tracking in the context of computer vision including autonomous robots based on this applications.

A real-time captured image sequence generally exhibits minimal variation from one frame to the next. Consequently, the object information present across frames within a defined time interval tends to be significantly redundant. This redundancy can be leveraged to monitor particular objects and differentiate between various visual elements. Since the human visual system struggles to differentiate between a face and a complex background, identifying redundancy in a series of images continues to pose one of the most significant challenges in the field of recognition.

2. RELATED WORK

Many references address works on face tracking through face recognition, we first discuss the papers on face detection and later continue with face tracking methods. Chi et al.

[2] studied the current schemes in the field of visual surveillance. The recent development in computer vision-based applications has sparked much interest in Face detection and Recognition. Hao et al. [3] explained that face detectors on Convolutional Neural Networks (CNNs) are ineffective when dealing with faces of various sizes. It relies on multi-scale testing or applying a single sizeable model that represents faces on a large-scale spectrum. Further, Zhang et al. [4] discussed

[1] In this study, the words “app” and “platform” are used interchangeably.

that face detection with high performance remains a complex problem, mainly when there are several small faces. They presented Refine Face, a single-shot specialization face detector with high efficiency. Lenz et al. [5] presented the first purely event- based approach for face detection which uses an event-based camera's high temporal resolution properties to track the movement of an object in a shot. Guo et al. [6] stated that convolutional neural network-based face and object recognition methods (such as OverFeat, R-CNN and DenseNet) precisely extract multi-scale features based on an image. Zhang et al. [7] concluded that anchor-based deep face recognition techniques had shown promising outcomes but deep learning-based methods have difficulty in identifying stern faces that are small, fuzzy, or partially obscured. Further, Liang et al. [8] explained that face recognition from low-light exposures is complex due to the small number of photos available and the unavoidable noise, which is also spatially unevenly distributed, making the task even more difficult. According to Zhou et al. [9] in the field of security, the image taken by an outside surveillance camera, normally has distorted faces occluded in a variety of poses and tiny which is influenced by external factors such as camera pose and distance as well as weather conditions. Chen et al. [10] also explained that face spoofing puts the security of face recognition systems in jeopardy. Previous anti-spoofing research has focused on supervised methods with binary or auxiliary supervision being the most common. Xu et al. [11] introduced interface, a one-stage approach that predicts the location of the facial box and landmark in real time with more accuracy. Mahmoud et al. [12] suggested a robust method for detecting hidden faces in various camera angles and lighting conditions. A hybrid non-linear transform model that blends the RGB color space model and the YCbCr color model identifies human skin patches. Li et al. [13] suggested a face identification technique in the wild that uses a multi-task discriminative learning framework to integrate a ConvNet with a 3D mean face model. Zhang et al. [14] presented a unique cascaded Convolutional Neural Network called the Supervised Transformer Network to predict face regions and related facial landmarks. Tao et al. [15] used the kernel combination (LS- KC-SVM) approach to construct a locality-sensitive support vector machine to solve the problems. Lian et al. [16] presented using multiple objects tracking algorithms to create a real- time face tracking device. Ren et al [17], presented a tool for detecting and monitoring the human face in real-time using Convolution Neural Networks and Kalman Filters.

Further, the review on Face-tracking algorithms proposed by various authors are discussed subsequently. Lin et al. [18] added face tracking with region-based CNN, or FT- RCNN, is an effective face tracker based on the Faster-RCNN platform. In addition, Zheng et al. [19] suggested a deep learning-based face detection and tracking system that includes a Regression Network-based Face Tracking (RNFT) model to precisely monitor human faces in video sequences. The Squeeze and Excitation Network (SEN) and the Residual Neural Network (RNN) are combined in the SENResNet model (ResNet). Li et al. [20] demonstrated a multi- target face real-time detection, monitoring and recognition algorithm including three stages fast- tracking, detection and rapid recognition methods. A new GOTURN-based network is used in this work for quick face tracking. Chakravorty et al. [21] addressed visual face monitoring in real- world situations, covering various challenges in face matching. The authors introduced the FaceTrack method, which uses multiple appearance models as well as long- and short-term memory to provide effective face monitoring. Su et al. [22], proposed a quick Face Tracking-by-Detection (FFTD) that works independently for tasks like tracking, facial detection and discrimination. Li et al. [23] explained face tracking could be used to monitor faces reliably in various situations, including variation in lighting, background clutter, rapid

motion and partial occlusion. Soldie et al. [24] presented a powerful real-time face tracking device with numerous novel capabilities. Short and Long-Term memories (STM and LTM) are built into the framework and are used to monitor re-initialization throughout the online learning process. Pham et al. [25] presented a comprehensive hybrid 3D face tracking framework based on RGBD (Red Green Blue-Depth) video streams that tracks head pose and facial gestures without the need for re-calibration or user involvement. Ranganatha et al. [26] proposed an innovative face tracking method that combines the corner measured algorithm and the KLT (Kanade-Lucas-Tomasi) tracker. In the first frame of the video sequence, the Viola-Jones approaches first and detects the face and then extracts the detected portion of the face and applies to the Harris corner measured algorithm. Male et al. [27] developed a new reference architecture based on four paradigms. The suggested framework that allows deep learning ideas, a traditional approach to addressing the domain problem, cognitive agents with social concerns, and nature-inspired computing concepts to be integrated. According to Yuan et al. [28], traditional face tracking algorithms have obtained good results in some confined contexts. On the other hand, these methods necessitate the creation of manual facial features based on the researcher’s experience. Wu et al. [29] developed a unique framework for maintaining identification that simultaneously clusters and connects the faces of different persons in extended video sequences. Congcong et al. [30] proposed Dual-Cycle Deep Reinforcement Learning (DCDRL) to learn a robust face-tracking policy using just weakly-labeled annotations sparsely acquired from raw video data.

The approaches proposed in literature were unable to reach human-level performance in identifying faces. Further, accuracy and datasets look poor in proposed methods of face detection and tracking. Although progress in facial recognition was encouraging, the task has also turned out to be a difficult endeavor. To achieve better results, the work propose a face recognition system using CNN, LBP histogram, and HOG to perform face identification under difficult conditions. Initially, face detection from the input image was accomplished using the Viola-Jones technique [19]. The feature space was created by fusing later features that were extracted using CNN, HOG, and the LBP histogram. SVM and KNN classifiers were used to assess the proposed method's classification ability for various HOG cell sizes.

3. METHODOLOGY

Initially, the exact face area from the input image is extracted using the Viola-Jones algorithm, and then the retrieved face

region is resized to 64 × 64 dimensions for ensuring

accurate recognition and computational efficiency in processing of the images. Subsequently, by combining HOG, LBP, and CNN features, a comprehensive feature vector is created that leverages the strengths of each method, i.e., robust features are obtained with the assistance of CNN, while HOG acquires the local shape information from the input face image and LBP extracts texture features. The CNN, LBP histogram, and HOG features are combined to form a feature vector, which is then categorized using SVM. The entire process is depicted in Fig.1.

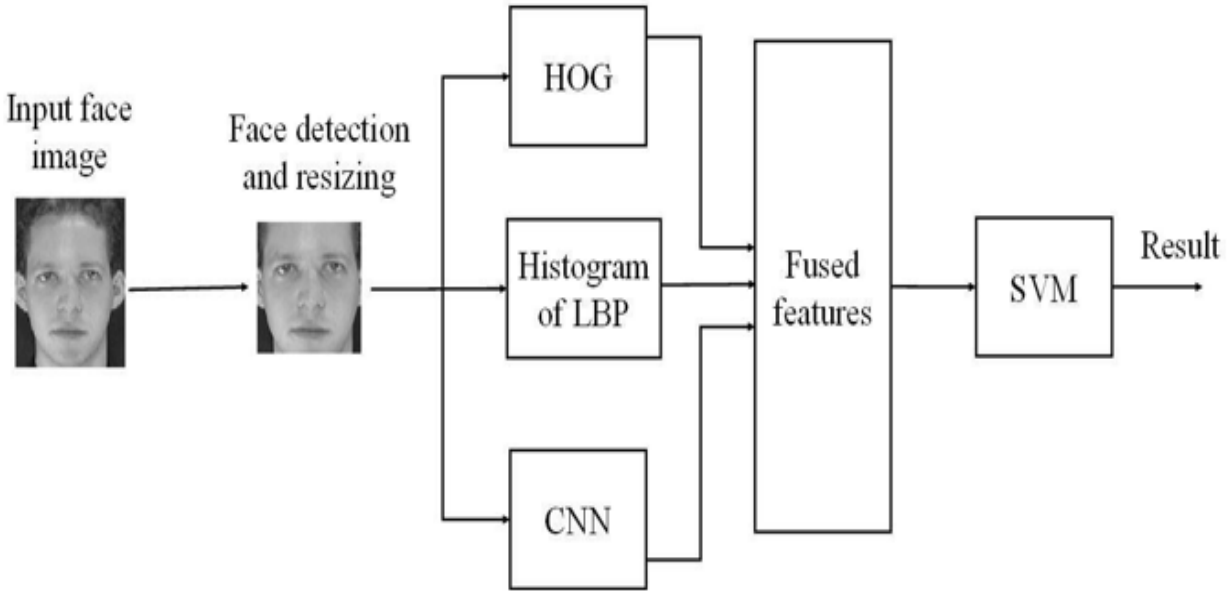


Fig. 1 The process of the proposed face recognition method

3.1 Histogram of Oriented Gradients (HOG)

The HOG is a feature-based descriptor used in image processing and computer vision for detecting the faces. The HOG feature extraction preserves the edges and also the directionality of the edge information. In this process, the entire image is divided in to cells. Each cell has a matrix of pixels. Each pixel casts a weight vote for an oriented based histogram channel. Histogram channels are evenly spread over 0 to 360 degrees. The HOG shape descriptor is used to find the shape of the local objects in computer vision (Dalal and Triggs [31]). HOG splits the image into tiny connected blocks, which are further segmented as cells. The HOG directions of each pixel in the cell is determined. Let P (.) is an intensity function denoting the grayscale values of the image. Each pixel's gradient in horizontal and vertical directions are determined as

The weights of Gradient magnitudes are combined to create

$$G_x = P(x+1, y) - P(x-1, y) \quad (1)$$

$$G_y = P(x, y+1) - P(x, y-1) \quad (2)$$

The magnitude of the gradient at a pixel is

$$G(x, y) = \sqrt{G_x^2 + G_y^2} \quad (3)$$

and the gradient orientation is given as

$$\theta(x, y) = \arctan\left(\frac{G_x}{G_y}\right) \quad (4)$$

a histogram vector for each cell. To enhance the robustness against edge intensity, shadows, and illumination, these histogram vectors are normalized. The final HOG representation consists of vectors from all normalized cells within each block. With a cell size of 4 × 4 for a 64 × 64 dimension image, this results in a feature vector of 1 × 8100. The input image and the respective HOG descriptor image is shown in Fig. 2.



Fig. 2 (a) Face detected image (b) HOG Descriptor

Local Binary Patterns

$$h(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Ojala et al. [32] presented LBP as a local texture descriptor. The grayscale value of the eight adjacent pixels in the 3×3 neighborhood is compared with the center pixel in the window. If the value is more than the central pixel then it is replaced with one, otherwise, zero is placed at that particular pixel location as given in Eqn. (5).

From Eqn. (6), eight bits are generated and then sum them up with a weight of 2n to obtain the value of LBP

$$LBP_{(P,R)} = \sum_{n=0}^{7} s\left(g_p - g_c\right) 2^n \quad \text{-------}$$

where gp (p = 0, 1, 2, ..., 7) represents eight pixels around the center pixel and gc

denotes the center pixel grayscale value, (xc, yc) is the location of the central pixel. (P, R) represents the P neighboring points with a radius of R. The way of generating LBP is given in Fig. 3. For the resized face obtained the histogram of the LBP feature vector of 1 × 59 dimension. The resized face image, LBP image, and corresponding histogram are shown in Fig 4.

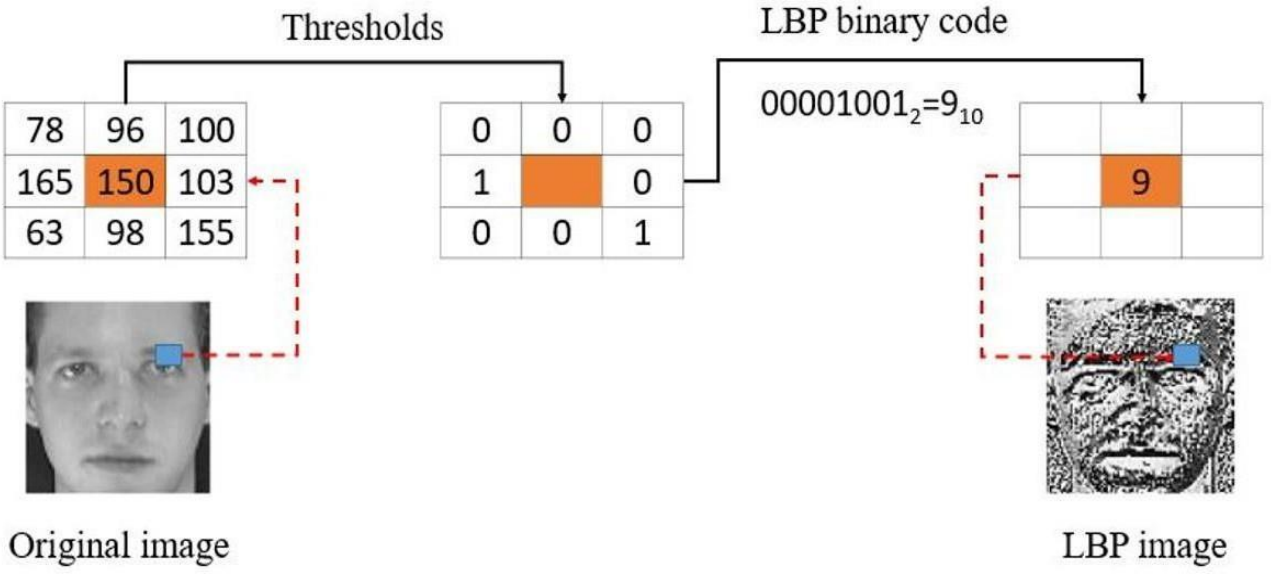


Fig. 3 The LBP with R=1 and P=8.

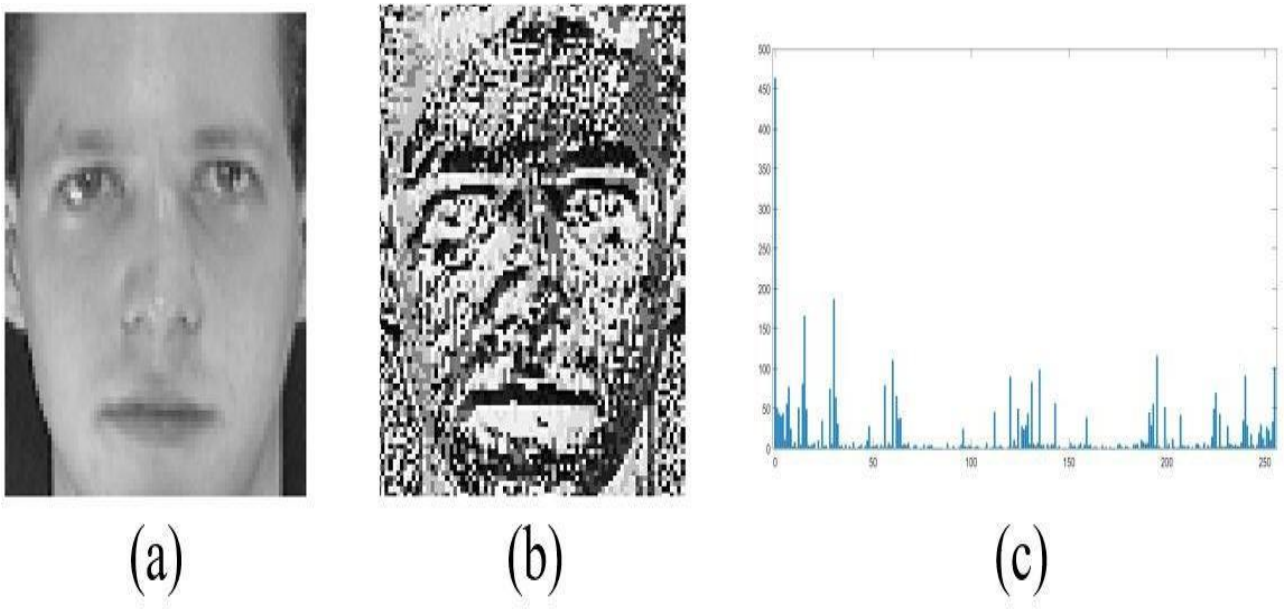


Fig. 4: Sample of (a) Resized image (b) LBP image (c) Histogram of (b)

4. PROPOSED CNN

The framework of the CNN is depicted in Fig. 5. It contains three convolutional layers with 8, 16, and 32 filters. In each convolutional layer, the ReLU is utilized as an activation function. The input to the first convolutional layer is an image with a 64x64x1 dimension. The first convolutional layer comprises 3x3 kernels with eight filters and the stride is set to one. Thus, the output of the Convolution 1 is eight feature maps with a 62x62 dimension. In the proposed CNN, every convolutional layer is succeeded by a max-pooling layer with a kernel size of 2 × 2 and stride two. Maxpooling1 produces an output with a dimension of 31 × 31 ×

8. The Conv2 and Conv3 produce output feature maps with dimensions 29 × 29 × 16, and 12

× 12 × 32 respectively. Maxpooling2 and Maxpooling3 generate an output with sizes 14×14×

16 and 6 × 6 × 32 respectively. The size of the fully connected layers is 250 and 120 that follow the Maxpooling3 layer. The number of learnable parameters for the proposed CNN are tabulated in Table 1. While training the data to the proposed CNN, Stochastic Gradient Descent is employed with a batch size of four. In every class of the face database, 70% of images were utilized for training and 30% for testing.

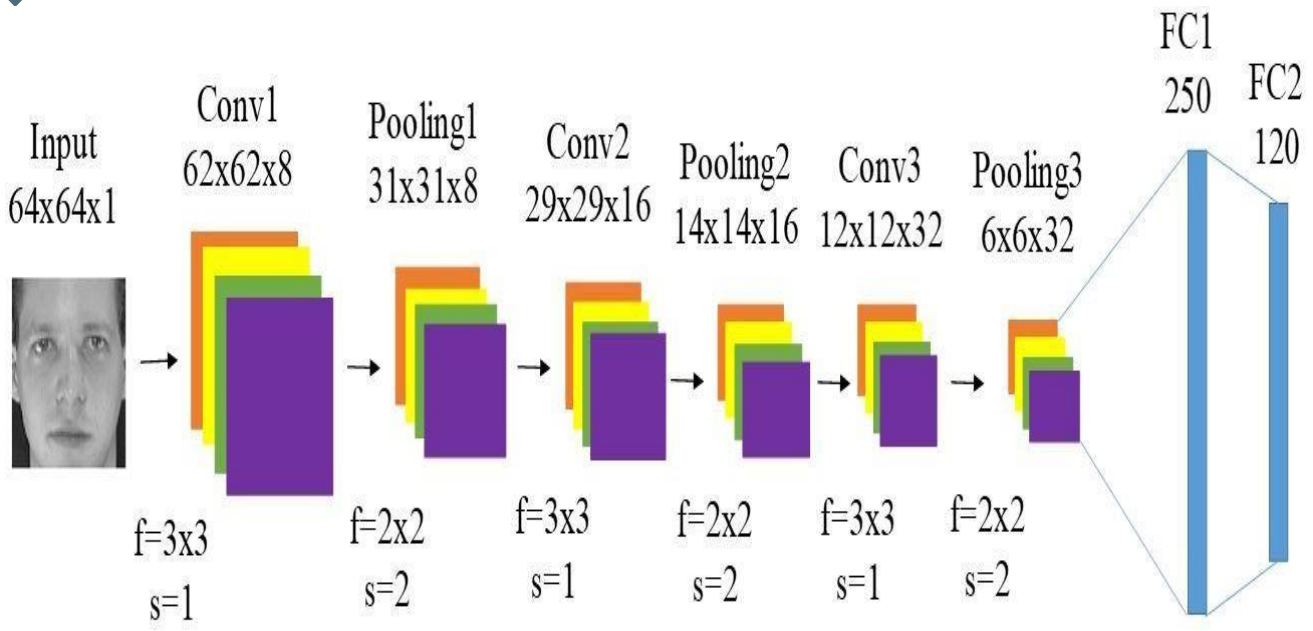


Fig. 5 The Architecture of the proposed CNN

Table 1 Number of learnable parameters of the proposed CNN

Layer	Activation shape	Number of learnable parameters
Conv1	(62, 62, 8)	80
Conv2	(29, 29, 16)	1168
Conv3	(12, 12, 32)	4640
FC1	(250, 1)	288250
FC2	(200, 1)	30120
Total number of learnable parameters		324258

Experiments and Results

We conducted experimentation on the ORL (Jin et al. [33]), Extended Yale B (Georghi- ades et al. [34], and CMU-PIE (Gross et al. [35]) face datasets. The ORL includes 400 images of 40 subjects with 10 different images for each person. Each subject contains images with various lighting, poses, illuminations, and facial details. The Extended Yale B comprises 16,128 face images of 28 persons with nine different poses and 64 lighting environments. The CMU-PIE includes 41,368 images of 68 classes. The images were captured from all subjects under 13 distinct poses, 43 dissimilar lighting environments, and four distinct variations. The few images of the aforesaid datasets are given in Fig. 6. The recognition rate for HOG, histogram of LBP, and CNN features is shown in Table 2 individually and the combination of these three methods on chosen face databases. For comparison purposes, the recognition rate with the KNN classifier is also given. From the values of Table 2, it is

observed that, compared to HOG and histogram of LBP, CNN has given a good recognition rate across all the chosen databases. Among all the combinations, the proposed method (HOG + histogram of LBP + CNN) has given a good recognition rate. Table 3 consists of the recognition rate values for HOG with a cell size of 8x8 on different databases. For 4x4 cell size, the recognition rate for the suggested approach on ORL, Extended Yale B, and CMU-PIE is 98.48%, 97.33%, and 97.28% respectively, whereas for the 8x8 cell size the recognition rate is 98.12%, 96.95%, and 96.74% respectively. From Tables 2 and 3, it is noticed that the HOG with a cell size of 4x4 produced good results compared to the HOG with an 8x8 cell size for the suggested method. To estimate the capability of the proposed methodology, the following performance metrics were utilized: precision, recall, specificity, and F1-score. The performance metrics on the aforementioned databases are specified in Tables 4, 5, and 6.

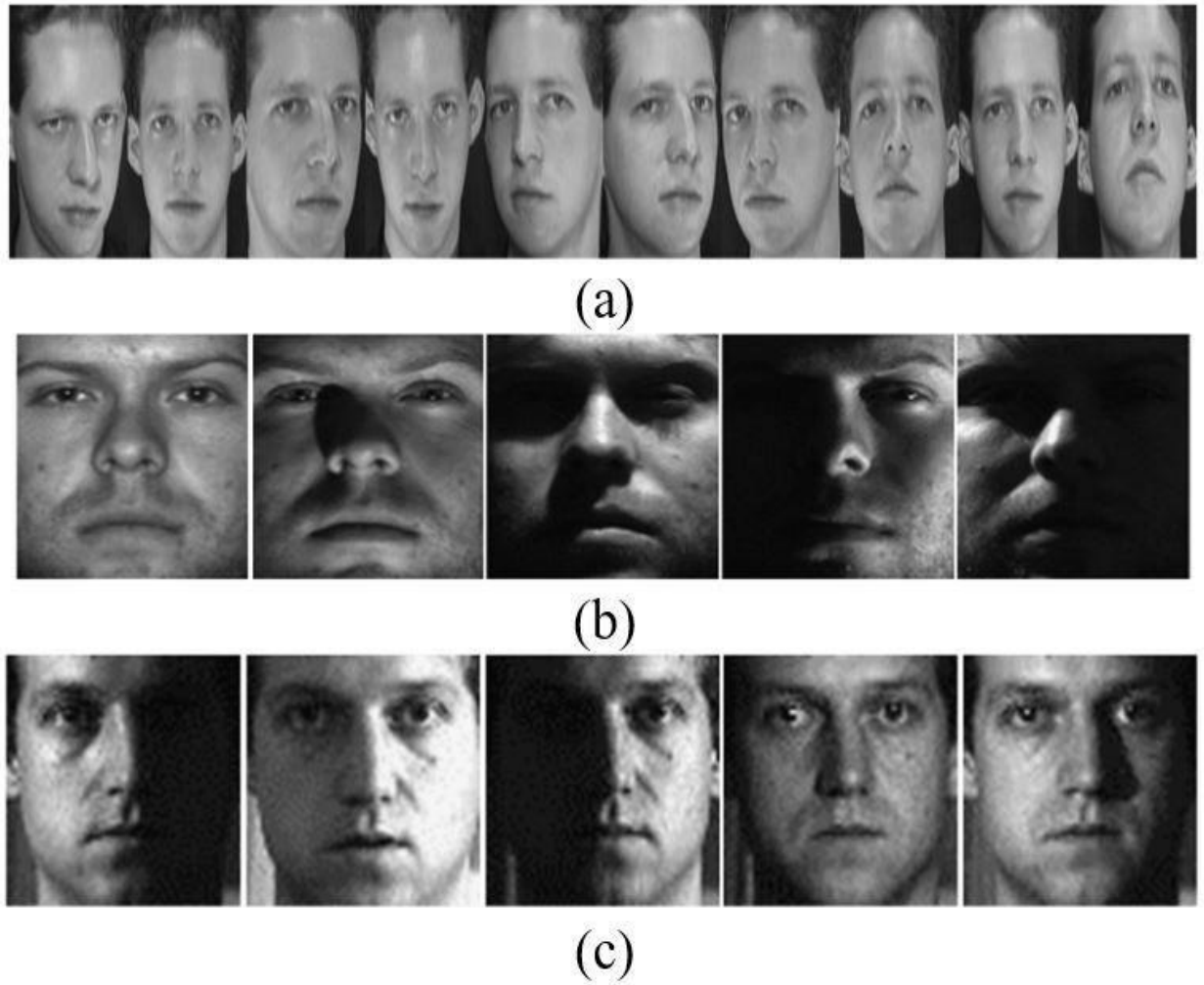


Fig. 6: Database images of (a)ORL (b)Extended YALE B (c)CMUPIE databases

Table 2 Recognition rate (%) using KNN and SVM classifier with HOG Cell size =4x4

Method	ORL		Extended Yale B		CMU-PIE	
	KNN	SVM	KNN	SVM	KNN	SVM
LBP	95.27	97.34	93.25	94.43	94.35	95.52
HOG	95.69	97.73	93.47	94.57	94.74	95.84
CNN	96.82	97.91	94.84	95.93	95.46	96.73
LBP+HOG	96.61	97.83	94.21	95.42	95.31	95.86
LBP+ CNN	97.21	98.17	95.46	96.61	95.82	96.88
HOG+CNN	97.46	98.23	95.83	96.94	96.66	97.11
LBP+HOG+CNN	97.83	98.48	96.57	97.33	96.83	97.28

Table 3 Recognition rate (%) using KNN and SVM classifier with HOG Cell size =8x8

Method	ORL		Extended Yale B		CMU-PIE	
	KNN	SVM	KNN	SVM	KNN	SVM
LBP	93.37	94.76	93.74	94.24	93.43	95.58
HOG	93.64	94.83	93.48	94.63	93.68	95.42
CNN	94.68	95.46	94.82	95.47	94.94	95.86
LBP+HOG	94.42	95.84	94.62	94.95	94.25	95.67
LBP+ CNN	95.62	96.23	94.64	95.24	95.24	96.32
HOG+CNN	95.86	96.78	95.22	96.37	95.83	96.68

LBP+HOG+CNN	96.15	98.12	95.64	96.95	95.96	96.74

Table 4 Performance Metrics on the ORL database

Classifier	Precision	Recall	Specificity	F1-Score
KNN	0.9925	0.9846	0.9763	0.9785
SVM	0.9887	0.9972	0.9742	0.9836

Table 5 Performance Metrics on the Extended Yale B database

Classifier	Precision	Recall	Specificity	F1-Score
KNN	0.9763	0.9749	0.8868	0.9741
SVM	0.9936	0.9884	0.9364	0.9779

Table 6 Performance Metrics on the CMU-PIE database

Classifier	Precision	Recall	Specificity	F1-Score
KNN	0.9742	0.9723	0.9682	0.9768
SVM	0.9983	0.9767	0.9863	0.9756

Table 7 Comparison of the suggested method with other techniques on the ORL.

Method	Recognition accuracy (%)
PCA (Cavalcanti et al. (2013))	95.85
LDA (Lu et al. (2012))	91.45
DLPV (Wen, Zhang, von Deneen and He (2016))	96.65
LBP (Ojala et al. (2002))	95.35
LOOP (Chakraborti et al. (2018))	97.31
GA-CNN (Rikhtegar et al. (2016))	94.61
SIAMESE (Wang, Yang, Xiao, Li and Zhou (2014))	92.10
DCT+LBP (Khan et al. (2015))	95.10
SSV (Zaaraoui et al. (2021))	96.75
Proposed method	98.48

Table 8 Comparison of the proposed method with other methods on the Extended Yale B.

Method	Recognition accuracy (%)
PCA (Cavalcanti et al. (2013))	83.47
LDA (Lu et al. (2012))	85.41
DLPV (Wen, Zhang, von Deneen and He (2016))	89.90
LBP (Ojala et al. (2002))	89.32
LOOP (Chakraborti et al. (2018))	95.36
GA-CNN (Rikhtegar et al. (2016))	93.84
SIAMESE (Wang, Yang, Xiao, Li and Zhou (2014))	92.52
DCT+LBP (Khan et al. (2015))	94.36
SSV (Zaaraoui et al. (2021))	94.42
Proposed method	97.33

Comparison of the proposed method with other techniques

To demonstrate the efficiency, the suggested approach is compared with the existing approaches. The holistic feature extraction methods like PCA (Cavalcanti et al. [36], LDA (Lu et al. [37]), Discriminative Locality Preserving Vectors (DLPV) (Wen, Zhang, von Deneen and He [38]), 2 Dimensional Random projection (2DRP) (Leng et al. [39]), and local feature descriptors namely LBP (Ojala et al. [40], Full Ranking (FR) (Chan et al. [41]), Local Optimal Oriented Pattern

(LOOP) (Chakraborti et al. [42]), Local Quadruple Pattern (LQP) (Chakraborty et al. [43]), and Strings of Successive Values (SSV) (Zaaraoui et al. [44]) were used for comparison. Moreover, the deep learning techniques like the Genetic Algorithm optimized structure of CNN (GA-CNN) (Rikhtegar et al. [45]), and SIAMESE network (Wang, Yang, Xiao, Li and Zhou [46]), additionally the approaches depending on a fusion technique like DCT+LBP (Khan et al. [47]) are also used. The comparison of the recognition accuracy for the suggested approach with other methods on chosen databases is given in Fig. 7.

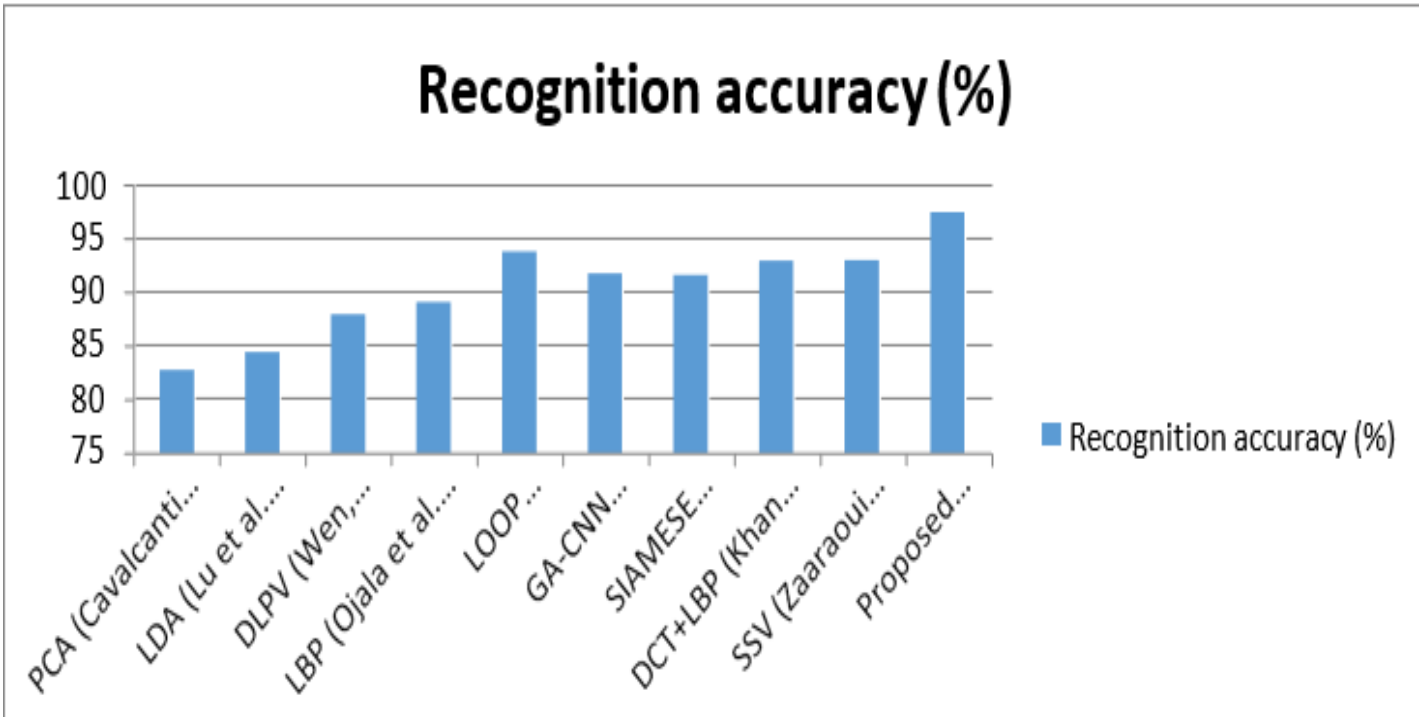


Fig. 7: Comparison of the proposed method with other techniques on the CMU-PIE

5. Summary

In this work, a convolutional neural network-based novel face recognition method is proposed. Initially, the Viola-Jones algorithm was used for face detection from the input image. Later features were extracted using HOG, histogram of LBP, and proposed CNN and are fused to create the feature space. The classification capacity of the suggested approach was tested with SVM and KNN classifiers for different cell sizes of HOG. Among these two classifiers, SVM has given a good recognition rate. The ORL, Extended Yale B, and CMU- PIE databases are used for experimental work and attained a recognition rate of 98.48%, 97.33%, and 97.28% respectively. Our experimental work reveals that the proposed approach remarkably improved the face recognition rate compared to some of the existing techniques. In future, we extend the proposed system to track the faces in videos of real surveillance system.

6. REFERENCES

[1] Dang, K., and S. Sharma. "Review and Comparison of Face Detection Algorithms." 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, Jan. 2017, pp. 629–633. IEEE.

[2] Chi, C., et al. "Selective Refinement Network for High Performance Face Detection." Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 1, July 2019, pp. 8231–8238.

[3] Hao, Z., et al. "Scale-Aware Face Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6186–6195.

[4] Zhang, S., et al. "RefineFace: Refinement Neural Network for High Performance Face Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[5] Lenz, G., S. H. Ieng, and R. Benosman. "Event-Based Face Detection and Tracking Using the Dynamics of Eye Blinks." Frontiers in Neuroscience, vol. 14, 2020, p. 587.

[6] Guo, G., et al. "A Fast Face Detection Method via Convolutional Neural Network." Neurocomputing, vol. 395, 2020, pp. 128–137.

[7] Zhang, Z., et al. "Robust Face Detection via Learning Small Faces on Hard Images." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 1361–1370.

[8] Liang, J., et al. "Recurrent Exposure Generation for Low-Light Face Detection." arXiv preprint arXiv:2007.10963,

2020.

[9] Zhou, Z., et al. “Context Prior-Based with Residual Learning for Face Detection: A Deep Convolutional Encoder–Decoder Network.” Signal Processing: Image Communication, vol. 88, 2020, p. 115948.

[10] Chen, C., et al. “Spoof Face Detection via Semi-Supervised Adversarial Training.” arXiv preprint arXiv:2005.10999, 2020.

[11] Li, Y., et al. “Face Detection with End-to-End Integration of a ConvNet and a 3D Model.” European Conference on Computer Vision, Oct. 2016, pp. 420–436. Springer, Cham.

[12] Zhang, K., et al. “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks.” IEEE Signal Processing Letters, vol. 23, no. 10, 2016, pp. 1499–1503.

[13] Tao, Q. Q., et al. “Robust Face Detection Using Local CNN and SVM Based on Kernel Combination.” Neurocomputing, vol. 211, 2016, pp. 98–105.

[14] Pham, H. X., et al. “Robust Real-Time Performance-Driven 3D Face Tracking.” 2016 23rd International Conference on Pattern Recognition (ICPR), Dec. 2016, pp. 1851–1856. IEEE.

[15] Ranganatha, S., and Y. P. Gowramma. “A Novel Fused Algorithm for Human Face Tracking in Video Sequences.” 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Oct. 2016, pp. 1–6. IEEE.

[16] Lian, Z., Shao, S., and C. Huang. “A Real Time Face Tracking System Based on Multiple Information Fusion.” Multimedia Tools and Applications, vol. 79, 2020, pp. 16751–16769.

[17] Ren, Z., et al. “A Face Tracking Framework Based on Convolutional Neural Networks and Kalman Filter.” 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Nov. 2017, pp. 410–413. IEEE.

[18] Ren, Z., et al. “A Face Tracking Framework Based on Convolutional Neural Networks and Kalman Filter.” 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Nov. 2017, pp. 410–413. IEEE.

[19] Zheng, G., and Y. Xu. “Efficient Face Detection and Tracking in Video Sequences Based on Deep Learning.” Information Sciences, 2021.

[20] Li, J., et al. “Real-Time Detection Tracking and Recognition Algorithm Based on Multi-Target Faces.” Multimedia Tools and Applications, 2020, pp. 1–16.

[21] Chakravorty, T., Bilodeau, G. A., and É. Granger. “Robust Face Tracking Using Multiple Appearance Models and Graph Relational Learning.” Machine Vision and Applications, vol. 31, no. 4, 2020, pp. 1–17.

[22] Chakravorty, T., Bilodeau, G. A., and É. Granger. “Robust Face Tracking Using Multiple Appearance Models and Graph Relational Learning.” Machine Vision and Applications, vol. 31, no. 4, 2020, pp. 1–17.

[23] Li, T., Zhou, P., and H. Liu. “Multiple Features Fusion Based Video Face Tracking.” Multimedia Tools and Applications, vol. 78, no. 15, 2019, pp. 21963–21980.

[24] Soldić, M., et al. “Real-Time Face Tracking under Long-Term Full Occlusions.” Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, Sept. 2017, pp. 147–152. IEEE.

[25] Maleš, L., Marčetić, D., and S. Ribarić. “A Multi-Agent Dynamic System for Robust Multi-Face Tracking.” Expert Systems with Applications, vol. 126, 2019, pp. 246–264.

[26] Yuan, S., Yu, X., and A. Majid. “Robust Face Tracking Using Siamese VGG with Pre-Training and Fine-Tuning.” 2019 4th International Conference on Control and Robotics Engineering (ICCRE), Apr. 2019, pp. 170–174. IEEE.

[27] Wu, B., Hu, B. G., and Q. Ji. “A Coupled Hidden Markov Random Field Model for Simultaneous Face Clustering and Tracking in Videos.” Pattern Recognition, vol. 64, 2017, pp. 361–373.

[28] Congcong, Z., et al. “Dual-Cycle Deep Reinforcement Learning for Stabilizing Face Tracking.” 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), July 2019, pp. 543–548. IEEE.

[29] Ding, C., and D. Tao. “Robust Face Recognition via Multimodal Deep Face Representation.” IEEE Transactions on Multimedia, vol. 17, no. 11, 2015, pp. 2049–2058.

[30] Sun, Y., et al. “DeepID3: Face Recognition with Very Deep Neural Networks.” arXiv preprint arXiv:1502.00873, 2015.

[31] Dalal, N., and B. Triggs. “Histograms of Oriented Gradients for Human Detection.” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1, 2005, pp. 886–893. IEEE.

[32] Ojala, T., Pietikäinen, M., and T. Mäenpää. “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, 2002, pp. 971–987.

[33] Jin, X., and X. Tan. “Face Alignment In-the-Wild: A Survey.” Computer Vision and Image Understanding, vol.

162, 2017, pp. 1–22.

[34] Georghiades, A. S., Belhumeur, P. N., and D. J. Kriegman. “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, 2001, pp. 643–660.

[35] Gross, R., et al. “Multi-PIE.” Image and Vision Computing, vol. 28, no. 5, 2010, pp. 807–813.

[36] Cavalcanti, G. D., Ren, T. I., and J. F. Pereira. “Weighted Modular Image Principal Component Analysis for Face Recognition.” Expert Systems with Applications, vol. 40, no. 12, 2013, pp. 4971–4977.

[37] Lu, G.-F., Zou, J., and Y. Wang. “Incremental Complete LDA for Face Recognition.” Pattern Recognition, vol. 45, no. 7, 2012, pp. 2510–2521.

[38] Wen, Y., et al. “Face Recognition Using Discriminative Locality Preserving Vectors.” Digital Signal Processing, vol. 50, 2016, pp. 103–113.

[39] Leng, L., et al. “Two-Directional Two-Dimensional Random Projection and Its Variations for Face and Palmprint Recognition.” International Conference on Computational Science and Its Applications, Springer, 2011, pp. 458–470.

[40] Ojala, T., Pietikäinen, M., and T. Mäenpää. “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns.” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, 2002, pp. 971–987.

[41] Chan, C. H., et al. “Full Ranking as Local Descriptor for Visual Recognition: A Comparison of Distance Metrics on SN.” Pattern Recognition, vol. 48, no. 4, 2015, pp. 1328–1336.

[42] Chakraborti, T., et al. “LOOP Descriptor: Local Optimal-Oriented Pattern.” IEEE Signal Processing Letters, vol. 25, no. 5, 2018, pp. 635–639.

[43] Chakraborty, S., Singh, S. K., and P. Chakraborty. “Local Quadruple Pattern: A Novel Descriptor for Facial Image Recognition and Retrieval.” Computers and Electrical Engineering, vol. 62, 2017, pp. 92–104.

[44] Zaaraoui, H., et al. “Face Recognition with a New Local Descriptor Based on Strings of Successive Values.” Multimedia Tools and Applications, vol. 80, no. 18, 2021, pp. 27017–27044.

[45] Rikhtegar, A., Pooyan, M., and M. T. Manzuri-Shalmani. “Genetic Algorithm Optimized Structure of Convolutional Neural Network for Face Recognition Applications.” IET Computer Vision, vol. 10, no. 6, 2016, pp. 559–566.

[46] Sun, Y., Wang, X., and X. Tang. “Deep Learning Face Representation from Predicting 10,000 Classes.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.

[47] Khan, S. A., Usman, M., and N. Riaz. “Face Recognition via Optimized Features Fusion.” Journal of Intelligent and Fuzzy Systems, vol. 28, no. 4, 2015, pp. 1819–1828.