

Convolutional neural network-based face recognition using non-subsampled shearlet transform and histogram of local feature descriptors

Yallamandaiah S, N. Purnachand

School of Electronics Engineering, VIT-AP University, Amaravati, India

Article Info

Article history:

Received Feb 6, 2021

Revised Sep 23, 2021

Accepted Oct 4, 2021

Keywords:

Convolutional neural network

Face recognition

Feature extraction

Histogram

Non-subsampled shearlet transform

ABSTRACT

Face recognition has been using in a variety of applications like preventing retail crime, unlocking phones, smart advertising, finding missing persons, and protecting law enforcement. However, the ability of face recognition techniques reduces substantially because of changes in pose, illumination, and expressions of the individual. In this paper, a novel face recognition approach based on a non-subsampled shearlet transform (NSST), histogram-based local feature descriptors, and a convolutional neural network (CNN) is proposed. Initially, the Viola-Jones algorithm is used for face detection and then the extracted face region is preprocessed by image resizing operation. Then, NSST decomposes the input image into a low and high-frequency component image. The local feature descriptors such as local phase quantization (LPQ), pyramid of histogram of oriented gradients (PHOG), and the proposed CNN are used for extracting features from the low-frequency component of the NSST decomposition. The extracted features are fused to generate the feature vector and classified using support vector machine (SVM). The efficiency of the suggested method is tested on face databases like Olivetti Research Laboratory (ORL), Yale, and Japanese female facial expression (JAFPE). The experimental outcomes reveal that the suggested face recognition method outperforms some of the state-of-the-art recognition approaches.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Yallamandaiah S

School of Electronics Engineering

VIT-AP University

Amaravati, Andhra Pradesh 522237, India

Email: yallamandaiah@gmail.com

1. INTRODUCTION

Face recognition has grabbed noticeable attention in several areas like surveillance, information security, and entertainment [1]-[3] due to its uniqueness, low-cost, and easy accessibility compared to other biometric approaches. Face recognition is a process of recognizing an individual from the available face database [4]. A general face recognition methodology comprises pre-processing, extracting features, and classification stages. The pre-processing step involves operations like image de-noising, scaling, image registration, face detection, and normalization. In the feature extraction phase, features are obtained for efficient image representation and visual description. Feature extraction plays a major role in computer vision applications like face recognition [5]-[7], texture analysis [8], [9], and sketch synthesis [10]-[12]. A precise image feature should be both robust and discriminative to distinct variations like noise and illumination changes. The last step of the face recognition system is the classification which incorporates robust classifiers

namely K-Nearest neighbors (KNN), support vector machine (SVM), and extreme learning machine (ELM). to recognize the input face.

Recently, several face recognition methods have been developed with a good recognition rate under certain constraints [13]. However, in practice, the face recognition process is affected by some external factors like illumination, conclusion, and imaging equipment, which leads to a reduction in the efficiency of the recognition system. Therefore, face recognition is still a challenging task [14].

2. RELATED WORK

From the last few decades, several methodologies were developed to identify the faces in an image. Among all these methods feature extraction plays a major role. Typically, feature extraction techniques are categorized into subspace learning [5]-[7] and local feature descriptors [8], [9], [15], [16] methods. Principle component analysis (PCA) and fisher linear discriminant analysis (FLDA) are conventional subspace learning approaches. Local linear embedding (LLE) [17], isometric feature mapping [18], Laplacian Eigenmap [19] are different manifold learning techniques to unwrap the intrinsic low-dimensional representation. Abusham *et al.* [20] demonstrated an approach to face recognition by integrating PCA and LLE. An and Ruan [21] propounded Enhanced fisher's linear discriminant (EFLD) method and it outperforms the earlier algorithms. PCA reduces the dimension and eliminates correlation, however, it is not appropriate for classification [22]. Zhou *et al.* [23] introduced a face recognition method depending upon PCA image reconstruction and linear discriminant analysis. But the above-mentioned methods are computationally expensive since they deal with the Eigen decomposition and also require a lot of memory.

Compared to subspace learning methods, local feature descriptors are more efficient and robust. Further, they can be classified into handcrafted and learning-based descriptors. Local binary patterns (LBP) and Gabor wavelets are two typical handcrafted features. Ahonen *et al.* [24] primarily used LBP in face recognition and they attained promising results due to its effectiveness and simplicity [25]. Owing to this idea, several LBP approaches have been evolved [26]-[28]. However, the handcrafted features are sensitive to illumination variations, and also lose some texture information under specific conditions. These problems are resolved by learning-based descriptors. Among them, local quantized patterns [29] and discriminant face descriptors [30] staging a good performance.

Dai *et al.* [31] manifested a decorrelated 2D-feed-forward neural network ensemble with random weights for face recognition. Chen *et al.* [32] addressed the problem of multi-pose classification using 2D-gabor features and the Deep Belief Nets. Muqeet *et al.* [33] utilized LBP and directional wavelet transform for face recognition. Tai *et al.* [34] proposed the orthogonal procrustes problem (OPR) as a framework to recognize pose varying faces. Li *et al.* [35] introduced a new method to estimate the low-rank representation for image classification. Khan *et al.* [36] proposed a system that can recognize faces with varying illumination and expressions by employing particle swarm optimization (PSO). Lin *et al.* [37] propounded a new dictionary learning approach for face recognition. In recent years convolutional neural network (CNN) methods have grabbed substantial attentiveness in face recognition. The CNNs considerably enhances the model generation ability by establishing effective regularization strategies such as dropout [38]. The research group at Facebook developed a deep learning facial recognition system named DeepFace [39]. Sun *et al.* [40] proposed a CNN-based face representation named deep hidden IDentity feature (DeepID), whose features are learned by training a group of small CNNs. Features extracted from all the CNNs are concatenated to form a powerful feature. Yin and Liu [41] proposed multi-task learning for face recognition with the illumination, expression, and pose estimation as the side tasks. Görgel and Simsek [42], deep stacked denoising and sparse autoencoders (DSDSA) were used for face recognition. In this paper, a new face recognition technology is introduced by utilizing non-subsampled shearlet transform (NSST), histogram-based local feature descriptors, and CNN.

The remainder of the paper is planned as: we discuss the proposed face recognition method in section 3. Experimental outcomes are demonstrated in section 4. Section 5 consists of the conclusion of the paper.

3. PROPOSED WORK

The proposed approach consists of five major phases, detecting a face from the input image, preprocessing, NSST decomposition, extracting features, and classification. Face detection removes unwanted parts like hands, neck, and surroundings from the images, and gives the region of interest. Here Viola-Jones [43] algorithm is utilized for face detection. After the detection of the face region, the image resizing preprocessing operation is performed. Later, NSST is applied to the preprocessed image, and features are extracted by using LPQ, pyramid of histogram of oriented gradients (PHOG), and the proposed

CNN. The extracted features are fused to obtain a hybrid feature vector. Finally, SVM is employed as a classifier to recognize the face images. The whole process is shown in Figure 1.

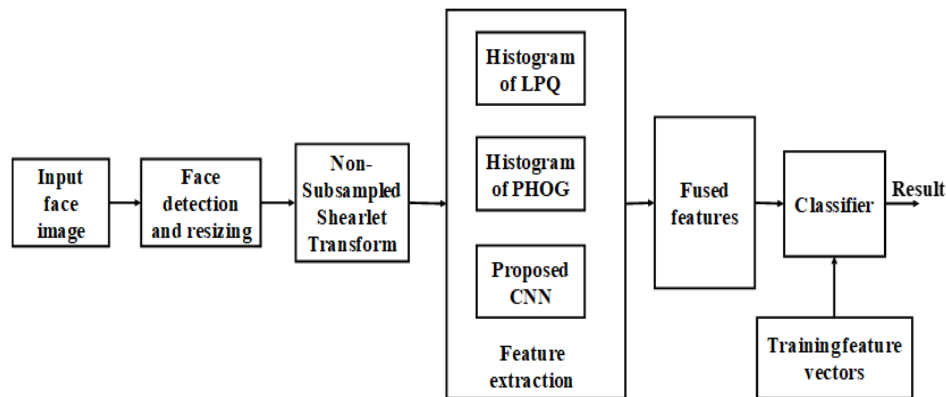


Figure 1. Block diagram of the proposed face recognition method

3.1. Non-subsampled shearlet transform

Traditional multiscale methods like wavelets, curvelets, and contourlet transforms are unable to capture the anisotropic features in multidimensional data. These problems are overcome by shearlets since they can efficiently represent the data in multidimensional phenomena [44]. Let dimension $n = 2$, the discrete shearlet transform can be given as (1),

$$\{\psi_{p,q,r}(x) = |\det M|^{p/2} \psi(S^q M^p x - r) : p, q \in \mathbb{Z}, r \in \mathbb{Z}^2\} \tag{1}$$

where ψ is a group of basis functions that satisfies $\psi \in L^2(\mathbb{R}^2)$, M indicates the anisotropy matrix, S is a shear matrix, p, q, r are scale, dimension, and shift parameters. Both M , and S are invertible matrices with size 2×2 and $|\det S| = 1$. For each $k > 0$ and $s \in \mathbb{R}$, the matrices M and S are given by (2),

$$M = \begin{pmatrix} k & 0 \\ 0 & \sqrt{k} \end{pmatrix}, S = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \tag{2}$$

the matrix M controls the scaling of shearlet and S controls the orientation of shearlet. For $k = 9, s = 1$, (2) becomes

$$M = \begin{pmatrix} 9 & 0 \\ 0 & 3 \end{pmatrix}, S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{3}$$

the basic function $\hat{\psi}^{(0)}$ for shearlet transform, for any $\beta = (\beta_1, \beta_2) \in \hat{\mathbb{R}}^2, \beta_1 \neq 0$ is given by (4),

$$\hat{\psi}^{(0)}(\beta) = \hat{\psi}_1(\beta_1) \hat{\psi}_2(\beta_2/\beta_1) \tag{4}$$

here $\hat{\psi}$ is the Fourier transform of ψ . $\hat{\psi}_1 \in C^\infty(\mathbb{R}), \hat{\psi}_2 \in C^\infty(\mathbb{R})$ are both wavelets.

The NSST decomposition consists of multi-scale and multi-directional factorization steps. To achieve multiscale factorization, the non-subsampled laplacian pyramid (NSLP) is utilized and it consists of a dual-channel non-subsampled filter bank to ensure multi-scale property, which separates the input image into low and high-frequency components. Implementation of successive NSLP decomposition is done to decompose the low-frequency component repeatedly and hence singularities in images are found. Similarly, to realize, multi-directional factorization improved shearing filters are used. In our proposed approach, initially, we detect the face region and then resize it to 64×64 and then NSST is applied to it. Figure 2(a) shows the input face image, Figure 2(b) gives the detected face region, and the low-frequency sub-band component from the NSST is shown in Figure 2(c).

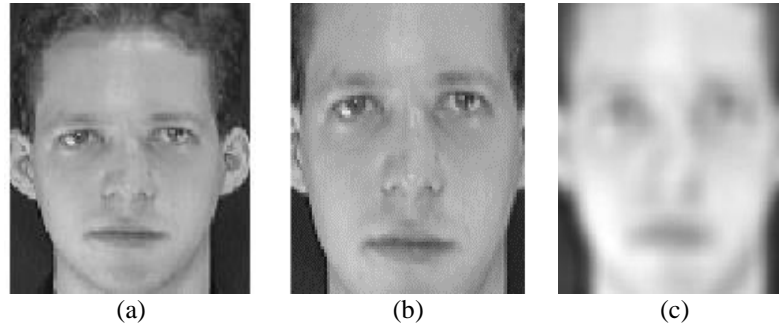


Figure 2. (a) input face image, (b) detected face region, (c) NSST low-frequency sub-band component

3.2. Local phase quantization

LPQ is a well-known local texture feature descriptor and used to extract the textual details, which are robust to blurring [45]. Initially, LPQ performs short time Fourier transform (STFT) to obtain the phase details for every pixel of the source image and then encrypts the corresponding phase information. Finally, estimates the distribution of the encrypted details to get the LPQ features. The mathematical description of LPQ is described as:

Let us assume that $p(m, n)$ be an original image. Then the spatial invariant blurring of the image $p(m, n)$ is obtained by a convolution operation (5):

$$q(m, n) = p(m, n) \otimes h(m, n) \quad (5)$$

where $q(m, n)$ is a blurred image, $h(m, n)$ is the point spread function (PSF) and \otimes represents the convolution [46].

The Fourier representation of (5) is given by (6),

$$Q(u, v) = P(u, v) \cdot H(u, v) \quad (6)$$

where $Q(u, v)$, $P(u, v)$, and $H(u, v)$ are the Fourier transforms of $q(m, n)$, $p(m, n)$, and $h(m, n)$ respectively. After that, the phase information of the blurred image is attained by the following expression,

$$\angle Q(u, v) = \angle P(u, v) + \angle H(u, v) \quad (7)$$

where $\angle Q(u, v)$, $\angle P(u, v)$, and $\angle H(u, v)$ are the phases of $q(m, n)$, $p(m, n)$, and $h(m, n)$ respectively. When the PSF, $h(m, n)$ is centrally symmetric, its phase has only two values and is represented by (8),

$$\angle H(u, v) = \begin{cases} 0, & \text{if } H(u, v) \geq 0 \\ \pi, & \text{otherwise} \end{cases} \quad (8)$$

thus, the phase invariance between $Q(u, v)$ and $P(u, v)$ is obtained as (9),

$$\angle Q(u, v) = \angle P(u, v), \quad \text{for all } H(u, v) \geq 0 \quad (9)$$

However, in LPQ the phase details are evaluated over the $M \times M$ neighborhood region of image $q(m, n)$. To obtain these local spectra features estimate the STFT by (10),

$$Q(u, v) = \sum_{m \in N_m} \sum_{n \in N_n} q(m, n) e^{-j2\pi(um+vn)/M} \quad (10)$$

where N_m and N_n indicate the neighborhood region. LPQ finds the phase detail at frequency points $z_1 = (a, 0)$, $z_2 = (0, a)$, $z_3 = (a, a)$, $z_4 = (a, -a)$ using STFT, where a is a small integer that obeys (9). The acquired results are arranged as (11)

$$V = [Q(z_1), Q(z_2), Q(z_3), Q(z_4)] \quad (11)$$

and

$$W = [Re\{V\}, Im\{V\}] \quad (12)$$

where $Re\{V\}$ represents the real part of V and $Im\{V\}$ denotes the imaginary part of V . The textural details can be obtained by encrypting the elements in W as (13),

$$c = \sum_{i=1}^8 k_i 2^{i-1} \quad (13)$$

where k_i is the quantization of the i^{th} element in W , given by (14)

$$k_i = \begin{cases} 1, & \text{if } W_i \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

Finally, the LPQ is obtained by detecting the distribution histogram of the encoded values c . In the proposed method, after applying NSST on the face detected image, the obtained low-frequency sub-band component is applied to LPQ to obtain the blur insensitive texture features. The detected face region from the input face image is given in Figure 3(a), and Figure 3(b) represents the NSST low-frequency sub-band component. Figure 3(c) and Figure 3(d) show the LPQ descriptor image and the corresponding histogram.

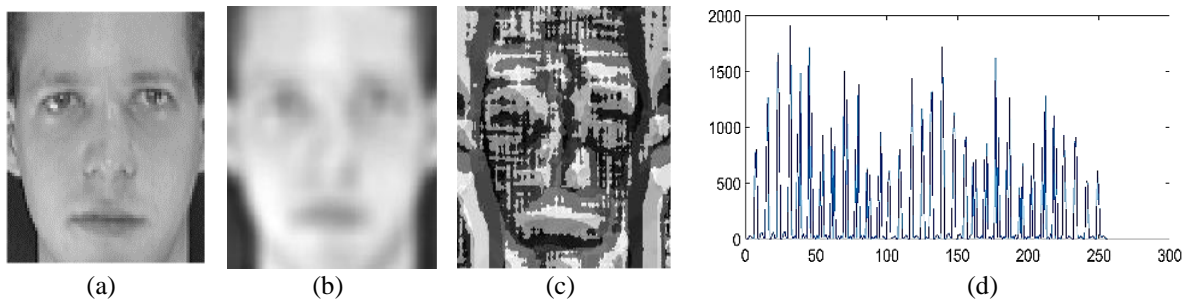


Figure 3. (a) detected face region, (b) NSST low-frequency sub-band component, (c) LPQ descriptor image, (d) histogram of (c)

3.3. Pyramid of histogram of oriented gradients

For effective face recognition, we require shape information along with texture details. To obtain such shape information we apply the PHOG descriptor which is built by utilizing the histogram of oriented gradients (HOG) features and pyramid representation of the images [47]. HOG descriptor is used to find the local shape of the objects in images and pyramid representation addresses spatial structure. The image is split into tiny regions (cells) and HOG features [48] are computed for every spatial region. The cells are split recurrently to maintain the local shape information completely. The extracted features from all the cells are integrated to form the final HOG features and they are concatenated with the pyramid structure to incorporate the details associated with the spatial design. Canny edge detection algorithm was utilized to identify the edges in the face image, and then the face image is split into cells by following the quad-tree concept. Let the M be the number of levels, and N be the number of bins for HOG features, then the dimension for PHOG descriptor is given by $N * \sum_{k=0}^M 4^k$. In this work, we choose 3 ($M = 0,1,2$) number of levels and the number of bins as 8, then the resultant feature vector has a size 168. Figure 4(a) shows the detected face image and Figure 4(b) represents the NSST low-frequency sub-band component. The PHOG descriptor image and the final histogram of the PHOG for the corresponding input face image are given in Figure 4(c) and Figure 4(d) respectively.

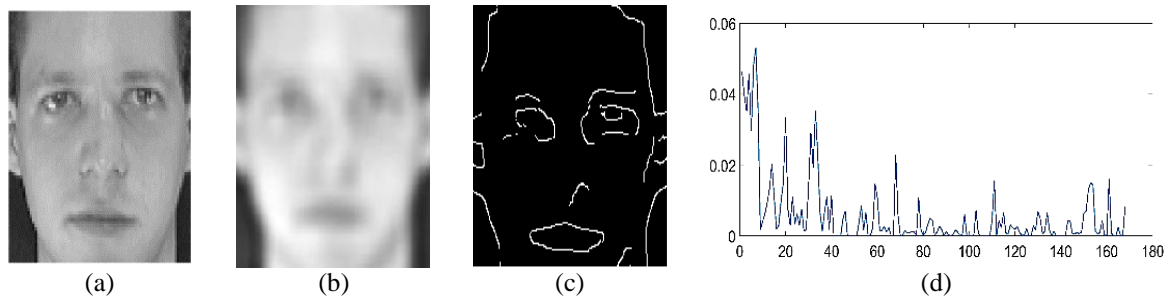


Figure 4. (a) detected face region, (b) NSST low-frequency sub-band component, (c) PHOG descriptor image (d) histogram of (c)

3.4. Proposed convolutional neural network

Convolutional neural networks have attained noticeable progress in image classification and they have been utilized in face recognition applications because they can extract robust facial features. The CNNs are generally made up of three types of layers namely, convolutional, pooling, and fully connected layers. A convolutional layer includes many convolutional kernels that are utilized to generate different feature maps. After each convolutional layer, a pooling layer is utilized that decreases the dimension of the feature maps and thus reduces the computational complexity of the CNN model. A fully connected layer considers all the neurons in the previous layer and associates them with every neuron of the current layer.

The architecture of the proposed CNN is shown in Figure 5. The proposed convolutional neural network contains three convolutional, three pooling, and two fully connected layers. The input to the proposed CNN is a $64 \times 64 \times 1$ grayscale image. The first convolutional layer has six 5×5 filters and the convolution stride is set to one pixel. Thus, the output of the first convolutional layer contains six feature maps with size 60×60 . Here, ReLU non-linear activation function is used in the convolutional layer. After each convolutional layer, max-pooling is accomplished over a 2×2 window, with stride two. Hence, the outcome of the maxpooling1 is feature maps with a 30×30 dimension. In each convolutional layer, the stride is considered as one whereas for max-pooling layers it is taken as two. The depth of the second and third convolutional layers is eight and ten, with output feature map dimensions $26 \times 26 \times 8$ and $9 \times 9 \times 10$ respectively. Maxpooling2 and maxpooling3 layers generate an output of $13 \times 13 \times 8$ and $5 \times 5 \times 10$ respectively. The last two layers are fully connected layers with 200 and 120 hidden units.

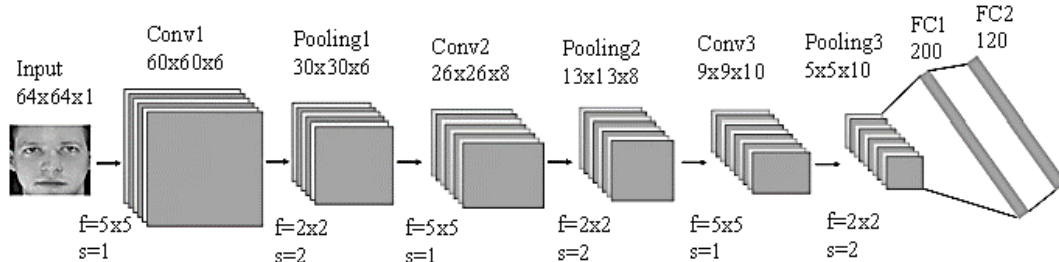


Figure 5. Architecture of the proposed CNN

4. EXPERIMENTAL RESULTS AND DISCUSSION

The capability of the suggested method, with different filters for the Laplacian pyramid decomposition [49], is tested using three face databases: i) ORL [50], ii) Yale [51], and iii) Japanese female facial expression (JAFFE) [52]. In every class of the database, 70% of images were utilized for training and the rest of the images were used for testing. While training the proposed CNN, stochastic gradient descent has been used for optimization with a base learning rate of 0.0001, and the maximum number of epochs as 20. Each experiment was done 10 times with the chosen datasets and the average recognition rate was given.

The ORL database comprises 40 different subjects. Each subject contains ten different images with distinct lighting environments, facial expressions, and attributes. In total ORL database includes 400 images each with 112×92 image resolution. The Yale face database includes a total of 165 face images of 15 subjects with 11 images per class. These images are considered under different configurations such as normal, happy,

sad, sleepy, surprised, no-glasses, wink, and center light. The JAFFE database includes face images of 10 Japanese females with seven facial expressions like the surprise, happiness, sadness, anger, fear, disgust, and neutral. It consists of 213 static grayscale images each with 256x256 image resolution. Figure 6(a)-(c) show the images belong to a single subject of ORL, Yale, and JAFFE databases respectively.

The performance of LPQ, PHOG, CNN, LPQ+PHOG, LPQ+CNN, PHOG+CNN, and LPQ+PHOG+CNN (proposed) is evaluated on the chosen databases with two different classifiers. The experiments were performed by utilizing k-fold cross-validation with 'k' value as 5. From every input image, the face region is extracted using the Viola-Jones algorithm, and then it is resized to 64x64 resolution. The features are extracted with LPQ, PHOG, and proposed CNN. These features are concatenated and fed to SVM. Another traditional classifier KNN is used for the classification of the extracted features.

To observe the effect of various NSLP filters [49], the recognition rate of the proposed approach for different filters is tabulated in Tables 1-4. The 'kos' filter produces a recognition rate of 97.61% with SVM, and 97.45% with the KNN classifier on the ORL database as given in Table 1. On the Yale database, the suggested method attains a recognition rate of 97.88% for SVM and 96.52% for KNN. Also, the recognition rate achieved is 98.75% with SVM and 97.48% with KNN on the JAFFE database. By using the 'pyr' filter face recognition rate achieved is 99.32% with SVM, 98.67% with KNN on the ORL database as tabulated in Table 2. On the Yale database, the face recognition rate is 98.72% with SVM and 97.85% with the KNN classifier. On the JAFFE database, the suggested method attains a recognition accuracy of 99.45% with SVM and 98.54% with KNN. The recognition rate of the proposed method for the 'pyrex' and 'maxflat' filters is tabulated in Tables 3 and 4 respectively.

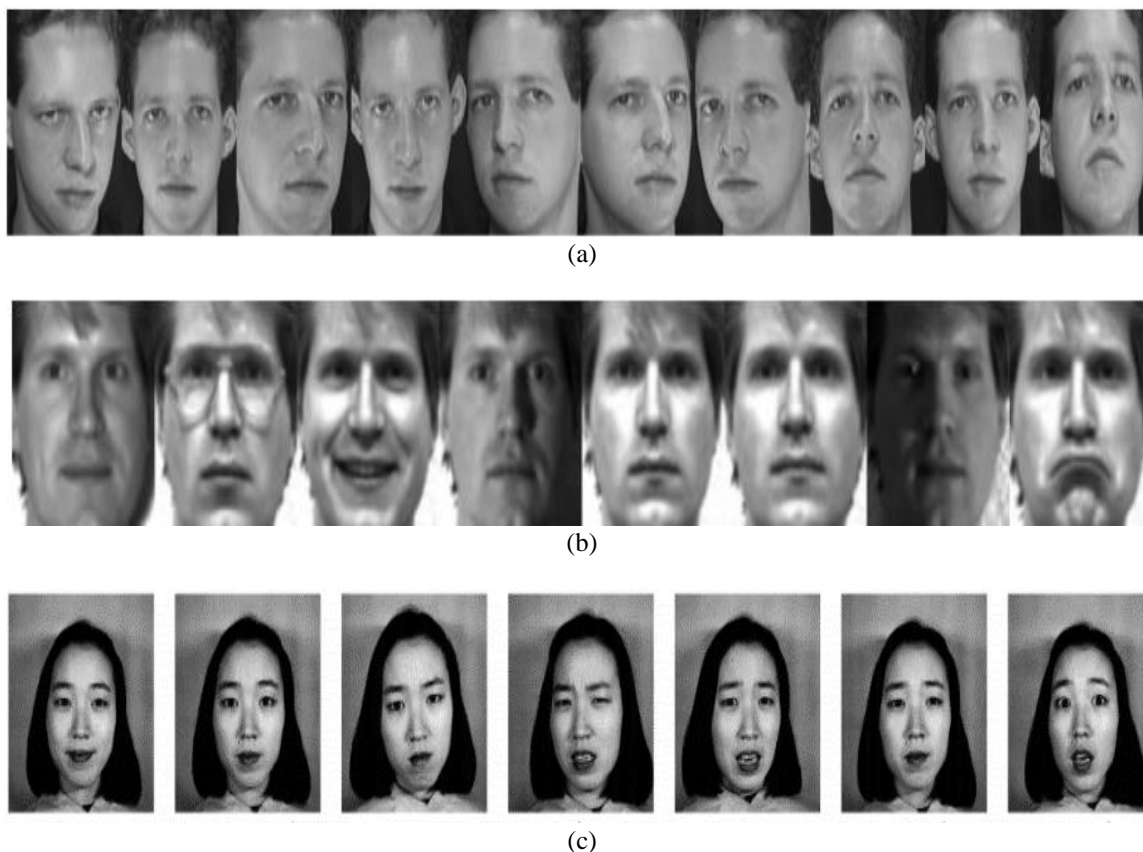


Figure 6. Images belong to a single subject of; (a) ORL, (b) Yale, and (c) JAFFE databases

From Tables 1-4, it is observed that, among the chosen NSLP filters, 'pyr' has given a good recognition rate for the proposed technique. The performance metrics for the proposed method with different classifiers on the three databases are given in Table 5. The ROC curve of the proposed face recognition system for the ORL database is shown in Figure 7(a)-(c) represent the ROC curves for Yale and JAFFE databases respectively. The required time to train the ORL, Yale, and JAFFE database to the proposed CNN

is 3.24 minutes, 1.56 minutes, and 2.35 minutes respectively. The execution time required to recognize the probe image of ORL, Yale, and JAFFE databases is 2 seconds, 1.2 seconds, and 1.6 seconds respectively.

Table 1. Recognition rate (%) for the proposed method with 'kos' filter

Database Classifier	ORL		Yale		JAFFE	
	KNN	SVM	KNN	SVM	KNN	SVM
LPQ	94.32	95.78	93.42	94.64	93.71	94.17
PHOG	92.72	93.52	92.33	93.76	92.41	92.32
CNN	94.54	96.31	95.56	96.37	95.28	96.84
LPQ+PHOG	95.46	95.92	94.37	95.27	93.36	94.58
LPQ+CNN	97.16	97.43	96.28	97.63	97.38	98.49
PHOG+CNN	96.46	96.87	96.15	97.26	96.27	97.92
Proposed	97.45	97.61	96.52	97.88	97.48	98.75

Table 2. Recognition rate (%) for the proposed method with 'pyr' filter

Database Classifier	ORL		Yale		JAFFE	
	KNN	SVM	KNN	SVM	KNN	SVM
LPQ	95.36	96.78	93.61	94.24	93.32	93.85
PHOG	93.53	94.52	92.43	93.66	92.42	92.25
CNN	96.43	97.91	96.21	96.55	96.72	97.41
LPQ+PHOG	96.28	96.39	94.86	95.29	94.78	94.84
LPQ+CNN	98.22	98.57	97.48	98.35	98.25	98.72
PHOG+CNN	97.24	98.82	96.47	97.38	98.14	98.26
Proposed	98.67	99.32	97.85	98.72	98.54	99.45

Table 3. Recognition rate (%) for the proposed method with 'pyrecx' filter

Database Classifier	ORL		Yale		JAFFE	
	KNN	SVM	KNN	SVM	KNN	SVM
LPQ	94.36	95.88	93.51	94.78	93.81	93.71
PHOG	93.42	94.53	92.33	93.56	92.43	92.25
CNN	96.25	96.58	95.24	96.34	96.29	95.62
LPQ+PHOG	95.38	96.51	94.38	95.54	94.47	94.48
LPQ+ CNN	97.65	97.16	96.58	97.43	97.58	98.51
PHOG+CNN	97.29	96.97	96.14	97.23	97.36	97.46
Proposed	97.78	97.94	97.28	97.87	97.67	98.87

Table 4. Recognition rate (%) for the proposed method with 'maxflat' filter

Database Classifier	ORL		Yale		JAFFE	
	KNN	SVM	KNN	SVM	KNN	SVM
LPQ	94.66	95.31	93.71	94.93	93.65	93.21
PHOG	93.52	94.82	92.63	93.65	92.47	92.14
CNN	96.23	97.23	95.84	96.34	96.14	96.95
LPQ+PHOG	95.45	96.81	94.75	95.96	94.25	94.28
LPQ+CNN	96.67	97.42	96.82	97.31	97.34	97.64
PHOG+CNN	96.45	97.24	96.26	96.93	97.19	97.28
Proposed	97.13	97.83	97.24	97.92	98.45	98.95

From the values of Tables 1-4, it is inferred that the proposed technique achieves a maximum face recognition rate of 99.32%, 98.72%, and 99.45% on ORL, Yale, and JAFFE databases respectively with 'pyr' filter. The comparison of the recognition rate for the proposed face recognition system on the ORL, Yale, and JAFFE databases with some of the existing methods shown in Table 6 appendix, to show its effectiveness.

Table 5. Performance metrics for the proposed method with 'pyr' filter

Database Classifier	ORL		Yale		JAFFE	
	KNN	SVM	KNN	SVM	KNN	SVM
Precision	0.9833	0.9916	0.9812	0.9833	0.9925	0.9937
Recall	0.9809	0.9904	0.9793	0.9777	0.9912	0.9916
Specificity	0.8678	0.9334	0.9632	0.9761	0.9873	0.9752
F1-Score	0.9821	0.9909	0.9758	0.9804	0.9895	0.9926

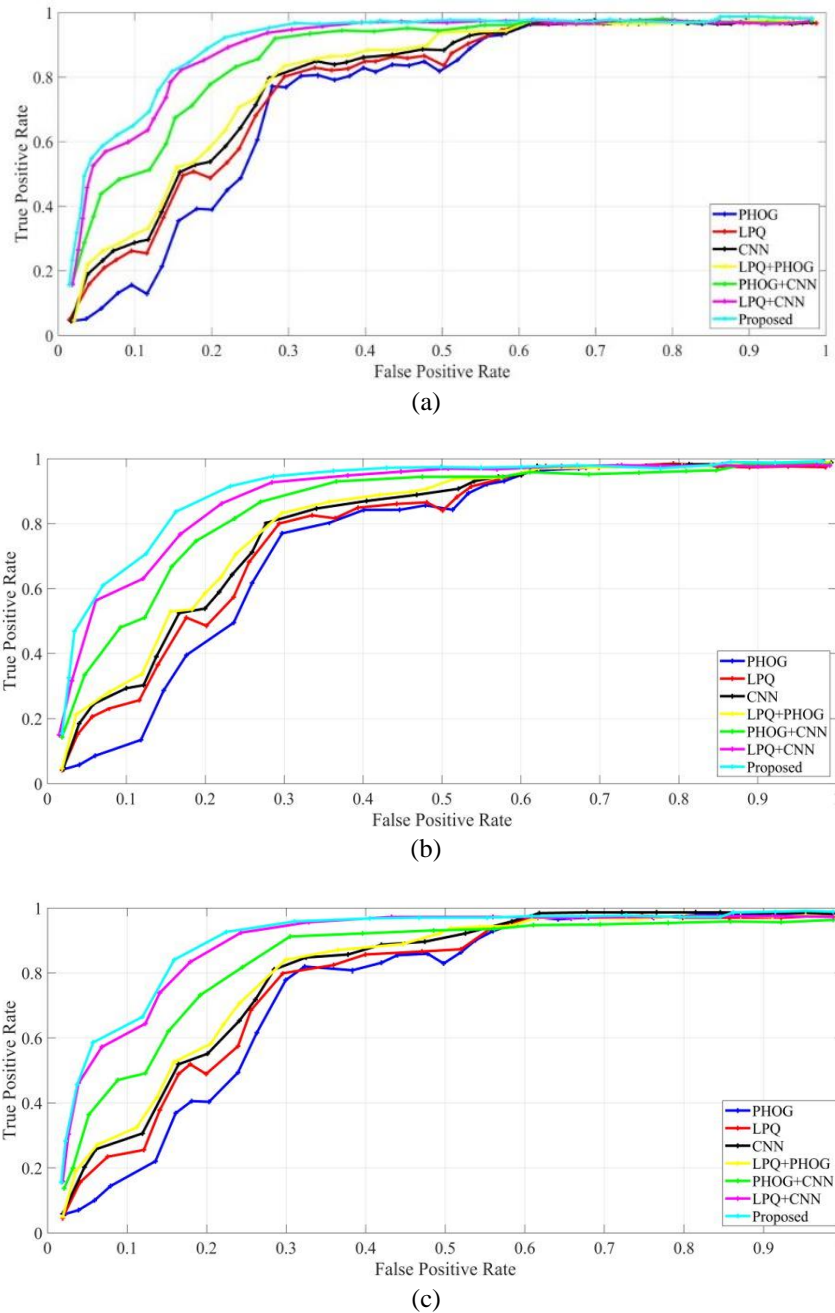


Figure 7. ROC curves for the proposed method on; (a) ORL, (b) Yale, and (c) JAFFE databases

5. CONCLUSION

A reliable and effective face recognition system using the NSST, the histogram of local feature descriptors, and CNN is proposed. The significant contribution of this work is presenting a novel method using histogram-based local feature descriptors, and CNN features on a transformed image for robust face recognition. NSST decomposes the input face image, into low and high-frequency sub-band components using the Non-Subsampled Laplacian Pyramid. Histograms of the local feature descriptors namely LPQ, PHOG, and the deep features from CNN are obtained from the low-frequency sub-band component and concatenated to form the feature space. In our proposed method compared to KNN classifier SVM produces better results on the chosen face databases. The experimental results reveal that the suggested method effectively recognizes the faces with different illuminations, poses, and expressions. Compared to some of the existing approaches, the proposed method achieves a better recognition rate.

APPENDIX

Table 6. Comparison of recognition rate (%) of the proposed method with some of the existing methods

Method	ORL	Yale	JAFFE
PCA [21]	89.50		
EFLDA [21]	93.00		
CLDA [21]	94.06		
PCA image reconstruction+LDA+SVM [23]	97.48		
GFDBN [32]	94.98		
DIWTLBP [33]	97.00		
DSDSA [42]	98.00		
Proposed	99.32		
OPR [34]		94.15	
PLR [35]		96.23	
Yin [41]		95.02	
RDCDL [38]		97.22	
DSDSA [42]		98.16	
Proposed		98.72	
FLLEPCA [20]			94.98
Single 2D-NNRW [31]			97.00
PSO [36]			98.80
Proposed			99.45

REFERENCES

- [1] Z. Huang, R. Wang, S. Shan, and X. Chen, "Hybrid Euclidean-and-Riemannian metric learning for image set classification," *Asian Conference on Computer Vision*, 2015, vol. 9005, pp. 562-577, doi: 10.1007/978-3-319-16811-1_37.
- [2] S. Z. Li and A. K. Jain, *Handbook of face recognition*, New York: Springer, vol. 1, 2011, doi: 10.1007/978-0-85729-932-1.
- [3] W. Zhao, R. Chellappa, P. J. Philips, and A. Rosenfield, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399-458, 2003, doi: 10.1145/954339.954342.
- [4] M. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, 2002.
- [5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991, doi: 10.1162/jocn.1991.3.1.71.
- [6] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997, doi: 10.1109/34.598228.
- [7] Xiaofei He, Shuicheng Yan, Yuxiao Hu, P. Niyogi and Hong-Jiang Zhang, "Face recognition using Laplacianfaces," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, March 2005, doi: 10.1109/TPAMI.2005.55.
- [8] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," in *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467-476, April 2002, doi: 10.1109/TIP.2002.999679.
- [9] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002, doi: 10.1109/TPAMI.2002.1017623.
- [10] N. Wang, X. Gao, L. Sun and J. Li, "Bayesian face sketch synthesis," in *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1264-1274, March 2017, doi: 10.1109/TIP.2017.2651375.
- [11] N. Wang, X. Gao, and J. Li, "Random sampling for fast face sketch synthesis," *Pattern Recognition*, vol. 76, pp. 215-227, 2018, doi: 10.1016/j.patcog.2017.11.008.
- [12] N. Wang, X. Gao, L. Sun and J. Li, "Anchored neighborhood index for face sketch synthesis," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2154-2163, Sept. 2018, doi: 10.1109/TCSVT.2017.2709465.
- [13] J. Wang, H. Zhou, and Z. Ji, "Research on face recognition technology in complex background," *Computer Engineering*, vol. 8, pp. 196-199, 2013.
- [14] J. Galbally, S. Marcel and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," in *IEEE Access*, vol. 2, pp. 1530-1552, 2014, doi: 10.1109/ACCESS.2014.2381273.
- [15] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [16] N. Dalal 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 doi: 10.1109/CVPR.2005.177.
- [17] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000, doi: 10.1126/science.290.5500.2323.

- [18] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, 2000, doi: 10.1126/science.290.5500.2319
- [19] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," in *Neural Computation*, vol. 15, no. 6, pp. 1373-1396, 1 June 2003, doi: 10.1162/089976603321780317.
- [20] E. E. Abusham, D. Ngo, and Andrew Teoh, "Fusion of locally linear embedding and principal component analysis for face recognition (FLLEPCA)," *International Conference on Pattern Recognition and Image Analysis*, 2005, pp. 326-333, vol. 3678.
- [21] G. An and Q. Ruan, "Novel mathematical model for enhanced fisher's linear discriminant and its application to face recognition," *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 524-527, doi: 10.1109/ICPR.2006.873.
- [22] G. Lu, J. Zou, and Y. Wang, "Incremental complete LDA for face recognition," *Pattern Recognition*, vol. 45, pp. 2510-2521, 2012, doi: 10.1016/j.patcog.2012.01.018.
- [23] C. Zhou, L. Wang, Q. Zhang, and X. Wei, "Face recognition based on PCA image reconstruction and LDA," *Optik*, vol. 124, no. 22, pp. 5599-5603, 2013, doi: 10.1016/j.ijleo.2013.04.108.
- [24] T. Ahonen, A. Hadid, and M. Pietikäinen. "Face recognition with local binary patterns," *European conference on computer vision*, Springer, Berlin, Heidelberg, 2004.
- [25] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen. "Local binary patterns and its application to facial image analysis: A survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 765-781, Nov. 2011, doi: 10.1109/TSMCC.2011.2118750.
- [26] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635-1650, June 2010, doi: 10.1109/TIP.2010.2042645.
- [27] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern recognition*, vol. 43, pp. 706-719, 2010, doi: 10.1016/j.patcog.2009.08.017.
- [28] Z. Guo, L. Zhang and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," in *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657-1663, June 2010, doi: 10.1109/TIP.2010.2044957.
- [29] S. U. Hussain and B. Triggs, "Visual recognition using local quantized patterns," *European Conference on Computer Vision*, 2012, pp. 716-729, vol. 7573, doi: 10.1007/978-3-642-33709-3_51.
- [30] Z. Lei, M. Pietikäinen and S. Z. Li, "Learning discriminant face descriptor," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289-302, Feb. 2014, doi: 10.1109/TPAMI.2013.112.
- [31] K. Dai, J. Zhao, and F. Cao, "A novel decorrelated neural network ensemble algorithm for face recognition," *Knowledge-Based Systems*, vol. 89, pp. 541-552, 2015, doi: 10.1016/j.knosys.2015.09.002.
- [32] Y. Cen, T. Huang, H. Liu, and D. Zhan, "Multi-pose face ensemble classification aided by Gabor features and deep belief nets," *Optik*, vol. 127, no. 2, pp. 946-954, 2016, doi: 10.1016/j.ijleo.2015.10.179.
- [33] M. A. Muqheet and R. S. Holambe. "Local binary patterns based on directional wavelet transform for expression and pose-invariant face recognition," *Applied Computing and Informatics*, vol. 15, no. 2, pp. 163-171, 2019, doi: 10.1016/j.aci.2017.11.002 .
- [34] Y. Tai, J. Yang, Y. Zhang, L. Luo, J. Qian and Y. Chen, "Face Recognition with Pose Variations and Misalignment via Orthogonal Procrustes Regression," in *IEEE Transactions on Image Processing*, vol. 25, no. 6, pp. 2673-2683, June 2016, doi: 10.1109/TIP.2016.2551362.
- [35] J. Li, Y. Kong, H. Zhao, J. Yang and Y. Fu, "Learning fast low-rank projection for image classification," in *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4803-4814, Oct. 2016, doi: 10.1109/TIP.2016.2598654.
- [36] S. A. Khan, M. Ishtiaq, M. Nazir, and M. Shaheen, "Face recognition under varying expressions and illumination using particle swarm optimization," *Journal of Computational Science*, vol. 28, pp. 94-100, 2018, doi: 10.1016/j.jocs.2018.08.005.
- [37] G. Lin, M. Yang, J. Yang, L. Shen, and W. Xie, "Robust, discriminative and comprehensive dictionary learning for face recognition," *Pattern Recognition*, vol. 81, pp. 341-356, 2018, doi: 10.1016/j.patcog.2018.03.021.
- [38] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*.
- [39] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.
- [40] Y. Sun, X. Wang and X. Tang, "Deep learning face representation from predicting 10,000 classes," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891-1898, doi: 10.1109/CVPR.2014.244.
- [41] X. Yin and X. Liu, "Multi-task convolutional neural network for pose-invariant face recognition," in *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 964-975, Feb. 2018, doi: 10.1109/TIP.2017.2765830.
- [42] P. Görgel and A. Simsek, "Face recognition via deep stacked denoising sparse autoencoders (DSDSA)," *Applied Mathematics and Computation*, vol. 355, pp. 325-342, 2019, doi: 10.1016/j.amc.2019.02.071.
- [43] P. Viola and M. J. Jones. "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004, doi: 10.1023/B:VISI.0000013087.49260.fb.
- [44] M. Mardanpou and M. A. Z. Chahooki, "Robust transparent image watermarking with Shearlet transform and bidiagonal singular value decomposition," *AEU-International Journal of Electronics and Communications*, vol. 70, pp. 790-798, 2016, doi: 10.1016/j.aeue.2016.03.004.
- [45] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," *International Conference on Image and Signal Processing*, 2008, pp. 236-243, vol. 5099, doi: 10.1007/978-3-540-69905-7_27.

-
- [46] Y. Xiao, Z. Cao, L. Wang, and T. Li, "Local phase quantization plus: A principled method for embedding local phase quantization into fisher vector for blurred image recognition," *Information Sciences*, vol. 420, pp. 77-95, 2017, doi: 10.1016/j.ins.2017.08.059.
- [47] N. Gour and P. Khanna, "Automated glaucoma detection using GIST and pyramid histogram of oriented gradients (PHOG) descriptors," *Pattern Recognition Letters*, vol. 137, pp. 3-11, 2019, doi: 10.1016/j.patrec.2019.04.004.
- [48] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.
- [49] A. Vishwakarma, M. K. Bhuyan, and Y. Iwahori, "An optimized non-subsampled shearlet transform-based image fusion using Hessian features and unsharp masking," *Journal of Visual Communication and Image Representation*, vol. 57, pp. 48-60, 2018, doi: 10.1016/j.jvcir.2018.10.005.
- [50] O. Olivetti, Oracle research laboratory face database of faces.
- [51] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997, doi: 10.1109/34.598228.
- [52] M. J. Lyons, M. Kamachi, and J. Gyoba, "The Japanese female facial expression (JAFFE) dataset," *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 14-16.