IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 16, 2021, accepted April 19, 2021, date of publication April 28, 2021, date of current version May 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3076359

DeepWTPCA-L1: A New Deep Face Recognition Model Based on WTPCA-L1 Norm Features

AYYAD MAAFIRI[1], OMAR ELHARROUSS[2], (Member, IEEE), SAAD RFIFI[3], SOMAYA AL-MAADEED[2], (Senior Member, IEEE), AND KHALID CHOUGDALI[1]

[1]Engineering Science Laboratory, National School of Applied Sciences (ENSA), Ibn Tofail University, Kenitra 14000, Morocco
[2]Department of Computer Science and Engineering, Qatar University, Doha, Qatar
[3]Advanced Science and Technology Laboratory, Larache Polydisciplinary Faculty, Abdelmalek Essaadi University, Larache 92004, Morocco

Corresponding author: Ayyad Maafiri (maafiri.ayyad@gmail.com)

ABSTRACT In this paper, we propose a robust face recognition model called DeepWTPCA-L_1 using WTPCA-L_1 features and a CNN-LSTM architecture. First, WTPCA-L_1 algorithm, composed of Three-level decomposition of discrete wavelet transform followed by PCA-L_1 algorithm, is exploited to extract face features. Then, the extracted features are used as inputs of the proposed CNN-LSTM architecture. To evaluate the robustness of the proposed approach, several face recognition datasets have been used. In addition, the proposed method is trained on noisy images using Gaussian, and Salt & Pepper noise added to the facial images of each dataset. The results of the experiment indicate that the proposed model achieves high recognition performance on three well-known standard face databases. When compared to state-of-the-art methods, the proposed model achieves a better face recognition rate.

INDEX TERMS Face recognition, WTPCA-L_1 algorithm, CNN-LSTM architecture.

I. INTRODUCTION

Data analysis is becoming a challenging task due to the advancement in the concerned domains such as real-time processing or those associated with large communication, computation, storage, and transmission of data [1]. Furthermore, the advancement of video surveillance systems that protect our lives necessitates the continuous processing of captured data, especially when people's security or objects' safety is jeopardized [2]. The acquisition using multiple views with multiple cameras can also complicate the analysis of each source of data, which makes the use of certain representative and unique features, like faces, a suitable solution. Nowadays, video technologies face a range of challenges and difficulties, primarily related to the retrieval of information in real-time from a vast number of recordings, as well as video management as collection and storage techniques are rapidly evolving. However, massive storage is required for all these videos which not only consumes space but also time. A requirement for storage of meaningful and interesting information is an essential task. Furthermore, the extracted data can be used to identify and detect multitude of events that can aid in analyses such as abnormal events and human specific behaviour, as well as predict a variety of events that commonly occur in scenes, especially scenes with large crowds such as football fans in stadiums. Eventually, a number of researchers focused on their studies on finding some techniques to extract the pertinent information, according to the purpose and the analyzed situations, from the captured data [3]. This research field is essential for the improvement of computer vision applications that require complex data analysis.

The face is an important feature for many computer vision applications, due to the unicity of its characteristics for each person, Which makes the face a better feature for identifying a human being. Due to the fact that humans are the prime source of insecurity in society, facial recognition can be useful for detecting and recognizing anomalistic behaviour [4]. In addition, the detection and the recognition of the existing faces in the monitored scenes can assist in managing and monitoring uncontrolled scenes like events. This has lead to face detection and recognition being a hot research topic during the last decades. It has attracted researchers to find better techniques with high performance. Many methods have been proposed using different image processing algorithms and techniques [5]. However, these approaches are limited by many problems including lack of datasets for training, illuminations changes, and pose variations of the faces which affect

The associate editor coordinating the review of this manuscript and approving it for publication was Yu-Da Lin.

the performance of such method [6]. With the introduction of deep learning techniques, the performance of face recognition has increased and helped overcome some problem like the selection of the suitable features that can be used during the recognition process. However, many problems still exist even using deep learning methods like pose variations, environmental changes, and image resolution.

In this work, a face recognition model is proposed called DeepWTPCA-L1. The proposed model starts by extracting face features using WTPCA-L1, while each face is presented with 40 features for ORL and GTFD datasets, and 120 features for FERET dataset. Due to the effectiveness of deep learning models in different computer vision tasks on image data, we exploit these techniques with other features like WTPCA-L1. After the extraction of WTPCA-L1 features of each face, a proposed deep learning architecture based on CNN and LSTM networks is used for face identification. The results obtained demonstrate the robustness of the proposed method for recognizing faces on different datasets. Also, the proposed approach can recognize faces even on noisy images.

The rest of this paper is structured as follows. In Section II, we review briefly some related face recognition algorithms. The PCA and PCA-L1 algorithms are explained in section III. The proposed approach is explained in Section IV. Experimental results are presented and discussed in Section V. Finally in Section VI, we give a conclusion relevant to the paper.

II. RELATED WORKS

Face recognition (FR) has attracted great interest from the research community due to its wide use in authentication applications. Also, the face structure makes the recognition of a face a very challenging task. The probability of appearance of the same face in a dataset is affected by different lighting conditions, poses, expressions, and occlusions, which increases the problem of intra-class variability. The key idea to overcome this problem is to find discriminant and optimal facial representations. In literature, the four most used techniques for extracting discriminant face features are Principal Component Analysis (PCA) [7], [8], Independent Component Analysis (ICA) [9], Linear Discriminant Analysis (LDA) [10], and Nonnegative Matrix Factorization (NMF) [11] like illustrated in Figure 1. PCA is adapted to extract the optimal features from an image. While LDA seeks a discriminant projection matrix which used for inter-class distance maximization and intra-class distance minimization. ICA is also looking for projection bases such that the projected data are uncorrelated. NMF finds an approximate decomposition of the original data matrix into two non-negative matrices. In addition, the global geometric structure of the data space is preserved while using these methods but ignores its local geometric structure. In order to overcome this problem, many data dimension reduction methods have been proposed. In [12] the authors proposed a learning method based on neighborhood preserving

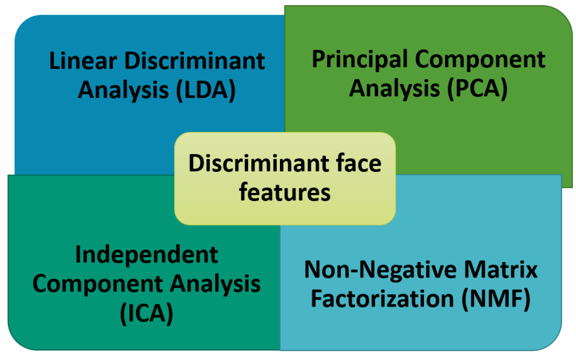


FIGURE 1. Discriminant facial representations techniques.

embedding (NPE). For the same purpose, the authors in [13] used locality preserving projection (LPP) for data dimension reduction. NPE and LPP are unsupervised-learning techniques based on approximations of Laplace embedding (LE) [14], [15] and local linear embedding (LLE) [16], respectively.

Motivated by NPR and LPP, many algorithms have been proposed for image recognition. but in subspace learning methods, a transformation from 2D images to 1D vectors should be done first. however, the spatial characteristics of the images are an important feature for recognition tasks but it's not taken into consideration when using these techniques. To solve this limitation, the researchers proposed many feature extraction methods, such as two-dimensional NMF (2DNMF) [17], two-dimensional PCA (2DPCA) [18], [19], and two-dimensional LDA (2DLDA) [20]. The advantage of these approaches is that the relevant features are formed from the original image. In addition, the approaches depends mainly on the L2-norm is suitable as a distance metric to compute the data similarity. However, the L2-norm can be affected by noises and outliers. This makes the approaches declared above not very effective.

In order to reduce the impact of outliers and noises, some researchers replaced L2-norm by L1-norm for distance metric computation. Several techniques based on the L1-norm have been widely used in image recognition. L1-PCA [21] and PCA-L1 [22] are the optimal and representative algorithms based on L1-norm. L1-PCA exploits the L1-norm in the objective function to minimize reconstruction error. PCA-L1 uses the projection bases which maximizes the objective function in sense of L1-norm. Inspired by the L1-norm-based algorithms, several dimensionality reduction methods have been exploited. For instance, the most representative are LDA-L1 [23], 2DLDA-L1 [24], 2DPCA-L1 [25], and 2DCRP-L1 [26].

During the last recent years, face recognition using deep learning techniques has given remarkable results and becomes one of the most active lines of research. Many works proposed for the same purposes such as Supervised convolutional neural network (CNN) which is adapted in DeepID [27] and DeepFace [28] for multi-class classification limitation

Algorithm 1 PCA-L_1 ($\tilde{d} = 1$)

Input: Data variance $F = \{f_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$
Output: Projection basis $\phi \in \mathbb{R}^{d \times 1}$
1: Initialization: Take any $\phi(t = 0)$. Set $\phi(0) \longleftarrow \frac{\phi(0)}{\|\phi(0)\|_2}$ and $t = 0$
2: Singularity check:
for $i = 1, 2, \ldots, n$ do
if $\phi^T f_i < 0$ then $p_i(t) = -1$
else then $p_i(t) = 1$
end for
3: Maximization:
$t \longleftarrow t + 1$
$\phi(t) = \sum_{i=1}^n p_i(t+1) f_i$
$\phi(t) \longleftarrow \frac{\phi(t)}{\|\phi(t)\|_2}$.
4: Stability test:
if $\phi(t) \neq \phi(t-1)$, go to Step 2.
Else if there exists i s.t $\phi^T f_i = 0$,
$\phi(t) \longleftarrow \frac{(\phi(t)+\Delta\phi)}{\|\phi(t)+\Delta\phi\|_2}$ and go to Step 2.
Otherwise, Set $\phi^* = \phi(t)$ and Stop.

and learning-features on large databases. Motivated by triples-loss, unified face-embedding [29] exploited Euclidean space embedding, triplet-loss, and 200 million facial images to have the best accuracy. To improve the recognition performance using feature embedding, a combination of softmax loss and contrastive loss called DeepID2 [27] aims to oversee CNN training for verification and identification. In [30], the authors have fully explained the center for CNNs by LDA and obtained a higher recognition rate. DeepID2+ [31] and DeepID3 [32] used an advanced networks to extract representative face features. Recently, effective CNNs in FR are based on either contrastive loss or triplet loss. Many techniques give considerable performances by improving on FR deep CNN's [33], [34]. Whereas in [35] the authors have proposed an efficient method, called robust discriminative non-negative dictionary learning for occluded FR. Also, the Local-circular quaternary pattern is discussed by Roy and Bhattacharjee [36].

The proposed methods in the literature, that used Deep learning techniques on images, have achieved good performance. But, it's still sensitive to any changes in the face images like noises, pose variation and illuminations.

III. PCA AND PCA-L1

Principal Component Analysis (PCA) is one of the most widely used and adapted unsupervised learning approaches in several application areas. Such as pattern recognition and computer vision. Its main purpose is to reduce the number of variables of the input data $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$ into a representation of small dimension $Y = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^{\tilde{d} \times n}$ with $\tilde{d} \ll d$. This is done mathematically by the search for a projection bases $\phi \in \mathbb{R}^{d \times \tilde{d}}$ which maximizes the data variance $F = \{f_i\}_{i=1}^n$, where $f_i = x_i - m$ and $m = \frac{1}{n}\sum_{i=1}^N x_i$ is the centroid of training-data.

$$\phi = arg \max_{\phi^T\phi = I_{\tilde{d}}} \sum_{i=1}^n \|\phi^T f_i\|_2^2 = arg \max_{\phi^T\phi = I_{\tilde{d}}} tr(\phi^T C_r \phi) \quad (1)$$

We assume that $C_r = \frac{1}{N} F F^T$ is a covariance matrix and $tr(.)$ is the trace operator. The solution of the Eq. (1) is composed of the orthonormal eigenvectors of C_r that correspond to the first $\tilde{d}$ largest eigenvalues. However, the advantage of conventional PCA in face recognition is its high recognition accuracy in ideal facial-databases. Also, it is not robust to outliers and noises for its inherent properties of L_2-norm. To handle this problem, Kwak [22] proposed a robust variant called PCA-L_1. This variant is based on L_1-norm instead of using L_2-norm. Expressed mathematically by the following equation :

$$\phi = arg \max_{\phi^T\phi = I_{\tilde{d}}} \|\phi^T F\|_1 = arg \max_{\phi^T\phi = I_{\tilde{d}}} \sum_{i=1}^n \|\phi^T f_i\|_1 \quad (2)$$

where $\|.\|_1$ denotes the L_1-norm of a vector. However, the authors in [22] reformulates the equation into an ordinary sum expressed by the following equation:

$$\phi = arg \max_{\phi^T\phi = I_k} \sum_{i=1}^n \|\phi^T f_i\|_1 = arg \max_{\phi^T\phi = I_k} \sum_{i=1}^n p_i \phi^T f_i \quad (3)$$

where p_i represents a polarity function equal to 1 if $\phi^T f_i \geq 0$; or -1 if $\phi^T f_i < 0$. For more technical details on PCA-L_1 algorithm can be referred to [22]. The steps to form the projection bases of PCA-L_1 are summarized in Algorithm 2. The projection vector ϕ converges to ϕ^*, which is a local maximum point of $\sum_{i=1}^n \|\phi^T f_i\|_1$. Nevertheless, its possible that this obtained solution it may not be a global solution. The first principal base of L_2-PCA is used as an initial vector $\phi(0)$ and presents a greedy search algorithm for $\tilde{d} > 1$.

Algorithm 2 PCA-L_1 ($\tilde{d} > 1$)

Input: Data variance $F = \{f_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$
Output: Projection bases $\phi = \{\phi_i\}_{i=1}^{\tilde{d}}$
1: $\phi(0) = 0$ and $F_0 = F$
2: for $i = 1, 2, \ldots, \tilde{d}$ do
3: $f_i = (I_{\tilde{d}} - \phi_{i-1}\phi_{i-1}^T) f_{i-1}$
4: Apply PCA-L_1 on f_i
5: end for
6: **return** ϕ

The PCA-L_1 algorithm is very robust to severe outliers and noise. This implies that PCA-L1 is better than the classical version in terms of recognition accuracy. However, PCA-L_1 suffers of two major critical points which are:

1) The method does not exhibit-invariance of rotation, because $\sum_{i=1}^n \|\omega\phi^T f_i\|_1 \neq \sum_{i=1}^n \|\phi^T f_i\|_1$ $(\omega^T\omega = I)$, where ω is an arbitrary rotation matrix.
2) The heuristic solution of the method consumes a lot of training time to form projection bases ϕ.

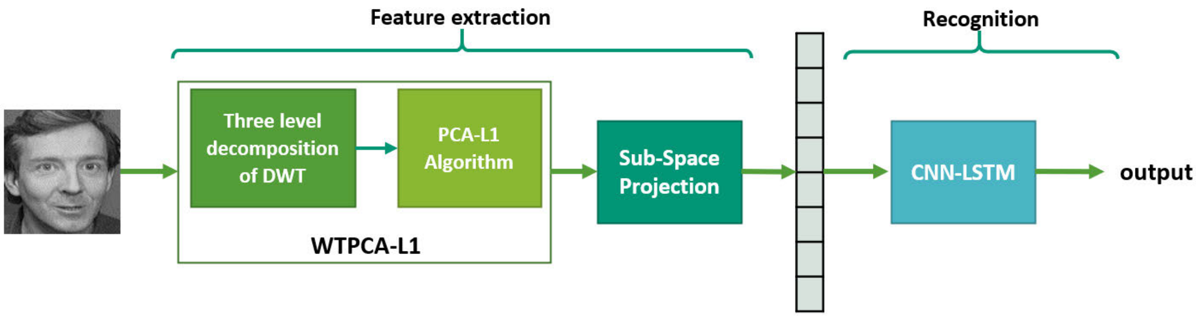


FIGURE 2. Flowchart of the proposed DeepWTPCA-L1 model.

Several researchers have worked to overcome this limitation. For example, the researchers in [37] proposed an algorithmic solution which is based on an initialization procedure of gram-schmidt orthogonalization. In this paper, we propose a new face recognition system which constitutes two major steps: 1) We exploit the WTPCA-L1 method [38] to better represent the raw images in a low dimensional mathematical space. 2) We propose a new CNN-LSTM architecture. In the next section, the proposed model will be described in details.

IV. PROPOSED APPROACH

Face recognition is a hot topic and an important task in computer vision applications. To achieve a high recognition efficiency, many researchers employ a variety of techniques. Deep learning methods have become the most prominent method for this task. Compared with traditional techniques, Convolutional Neural Networks (CNNs) on images, has achieved high performance in recognition. Unlike the existing method, this paper proposes a new method for face recognition using WTPCA-L1 as a features extraction module followed by a recognition module using a proposed CNN-LSTM architecture.

The proposed contribution consists of two main steps: the first step consists of extracting features from the face images using the WTPCA-L1 technique. The extracted features are used as inputs for the proposed CNN-LSTM architecture. This architecture is composed of a sequential convolutional and pooling layers that provides features extraction from the input data, followed by an LSTM network and then dense Layers. Figure 2 illustrates the flowchart of the proposed DeepWTPCA-L1 model. The following sections describe our proposed approach in detail.

A. FEATURE EXTRACTION USING WTPCA-L1

From WTPCA-L1 feature extraction algorithm stated in the previous section which aims to accelerate the process of computing PCA-L1 algorithm with the use of Three-level Wavelet decomposition Transformation (Three-level DWT) based on Daubechies [39] of the original images. Instead of extracting the projection bases of the raw images from the database, we only exploited the LL_3 band of the Three-level decomposition of DWT as illustrated in Figure 2.

The features extraction process start by selecting the training space, which defined mathematically by $X \in \mathbb{R}^{d\times n}$. WTPCA-L1 technique is exploited to compute the sub-space projection $\phi \in \mathbb{R}^{d\times \tilde{d}}$ using Algorithm 2. Then, the data X is projected using the following equation:

$$Y = \phi^T X \tag{4}$$

where $Y \in \mathbb{R}^{\tilde{d}\times n}$ contains the feature vectors of the training set. Using WTPCA-L1 on each image a vector of values is extracted. The obtained vectors are the features used as inputs of the proposed CNN-LSTM architecture that will be presented in the following section.

B. CNN-LSTM ARCHITECTURE

Convolutional-based models can extract useful information for effective learning from time-series or images data. Unlike traditional RNN model's LSTM networks are capable to identify "long-term and short-term dependencies". The combination of CNN and LSTM networks can improve the system performance, while the use of features extracted from face images can also help for a better learning for face recognition. By the following, a detailed description of the proposed data representation and the deep learning architecture will be presented.

C. CONVOLUTIONAL AND POOLING LAYERS

A network combining a sequence of convolutional and pooling layers dedicated to filtering the input data to extract desired information. The output of these layers is exploited by fully connected layers which are the final step in a network before the prediction of the desired output.

To obtain the output feature values of a convolutional layer, a convolution operation is applied to raw input data and convolution kernels. The structure of the input data must be in matrix form, due to the fact that the convolution operation was designed for image-based features extraction. Usually, each model is a set of blocks of convolutional layer followed by an activation (linear) and then a pooling layer. Pooling used to reduce the matrix dimension by extracting values from the convolutional layer output. Here, WTPCA-L1 data is a vector of values, in order to use convolutional-pooling layers, an Embedding layer is used first.

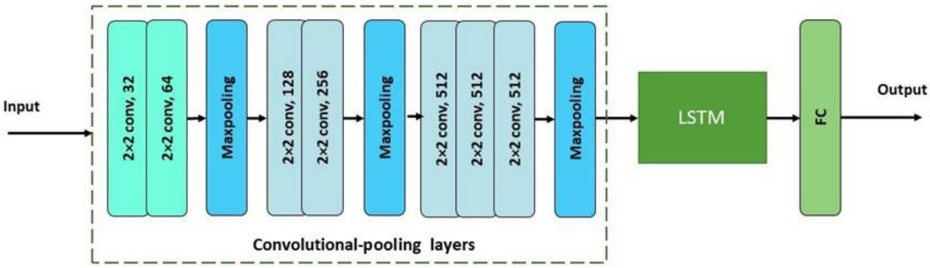


FIGURE 3. Proposed CNN-LSTM architecture.

TABLE 1. Training hyper parameters (General classifier).

Optimizer	LR	Epsilon	Beta_1	Beta_2	Decay	Epochs	Batch size
Adam	0.001	1e-08	0.91	0.999	0	100	32

D. LSTM

Learning long-term dependencies using feedback connections is the main process of an LSTM, which is a special type of Recurrent Neural Networks (RNNs). The traditional RNNs models tried to handle the feedforward problem in neural networks named as "lack memory" which affects the performance of the model on sequences and time series. In order to capture useful features form inputs data (time series of sequences) as well as to gain short-term memory, cyclic connections are used in these models. To avoid the famous RNNs problem vanishing or exploding the gradient when backpropagating through time, LSTM overcame this problem by saving the necessary information on memory cells and then vanishing the useless information, for reaching better performance. A basic LSTM unit consists of a short or long-term memory cell, an activation function as well as three gates including input i_t, forget f_t, and output o_t. The input i_t permits the incoming signal to modify the memory cell state or to block it. While f_t is the forget gate of control of the information that should be forgotten and the one that should be remembered. The last gate is o_t, which gives the state of the memory cell the possibility of infecting some neurons or to prevent some others. All of these allow LSTM to capture extremely complex and long-term temporal dynamics and to overcome the vanishing gradients problems. At time step, and an input x_t, LSTM calculates a hidden state named h_t, and memory cell state c_t, which is an encoding of everything observed by cell until time t:

$$i_t = \sigma(U_i x_i + W_i h_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma(U_g x_t + W_g h_{t-1} + b_g) \tag{6}$$

$$c_t^* = \tanh(U_c x_t + W_c h_{t-1} + b_c) \tag{7}$$

$$c_t = g_t \odot c_{t-1} + i_t \odot c_t^* \tag{8}$$

$$o_t = \sigma(U_0 x_t + W_0 h_{t-1} + b_0) \tag{9}$$

where x_t denotes the input, W_* and U_* are weight matrices, b_* are the vectors of bias term, σ is the sigmoid function, and the operator $\odot$ denotes component-wise multiplication. Finally, the hidden state h_t which constitutes the output of the memory cell is calculated by:

$$h_t = o_t \odot \tanh(c_t) \tag{10}$$

The LSTM layers are usually consecutive and the c_t and h_t which represent the memory and hidden state of each LSTM are used as next LSTM layer inputs.

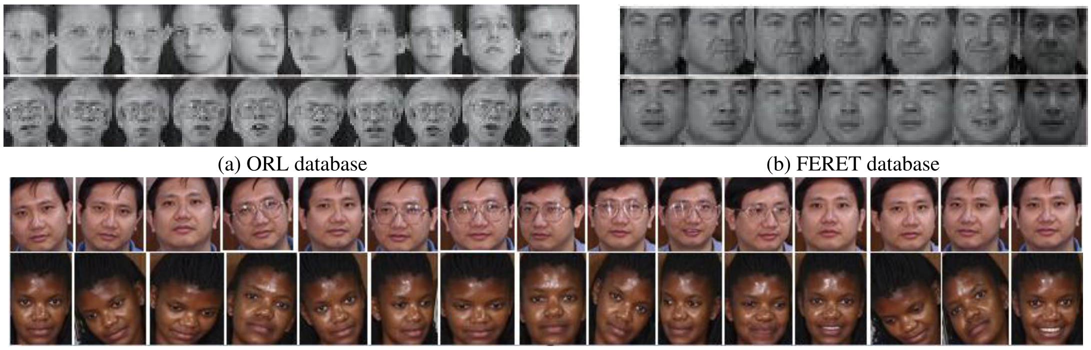
E. DeepWTPCA-L1

The implementation of the CNN-LSTM proposed architecture in Figure 3, consist of two blocks of convolutional and pooling layers before an embedding layer which allows the conversion of vector data to be usable with convolutional layer. The first block contains two convolutional layers of 32 and 64 filters of size 2, followed by a pooling layer. Where the second composed of another two convolutional layers of 128 and 256 filters of size 2 respectively, followed by a pooling layer. The third block contain three convolutional layers of 512 filters of size 2, followed by a pooling layer. The convolutional-pooling block is followed by an LSTM layer of 100 units then a dense layer of 128 neurons followed by an output layer of one neuron. The proposed model is represented in Figure 3.

The input of the CNN-LSTM architecture is a set of vectors that represent the features extracted using WTPCA-L1 for all images in the dataset. The deepWTPCA-L1 method used the CNN-LSTM model for face identification learning from the output of WTPCA-L1. The CNN-LSTM architecture is trained using *CrossEntrpy* loss function with a batch size of 32 examples, a learning rate of 0.001 as described in Table 1.

V. EXPERIMENTAL RESULTS

This section demonstrates the relevance of the proposed model by providing the experimental results. The evaluation has been performed on the ORL [40], GTFD [41], and FERET [42] datasets. The results obtained are compared with a set of state-of-the-art methods which are recent effective



(a) ORL database

(b) FERET database

(c) GT Face Database

FIGURE 4. Some face images from each dataset.

facial recognition techniques including PCA, 2DPCA, PCA-L_1, LDA, 2DLDA, WTPCA-L_1, $L_{2,p}$-norm PCA, and Discriminative PCA, which are trained on the same datasets. We have implemented all these methods to compare them with our proposed model. All the experiments are implemented using a machine with a 2.00 GHz i7 processor, 8 GB of RAM. The features extraction has been WTPCA-L1 has been implemented in MATLAB "R2018a" as the development environment. While the deep learning model has been implemented and trained using Python.

A. FACE RECOGNITION DATABASES

The ORL face database comprises 40 individuals. Each individual is represented by 10 different images. So the base contains 400 grayscale faces with a fixed of 92 × 112 pixel size. All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The Figure 4.(a) below shows some facial image of this database. We resize the facial-image size to 56 × 46, with quantization to 256 gray-level.

Facial Recognition Technology (FERET) is a facial database containing over 10,000 facial images. These images are taken in different situations with a size of 80 × 80 pixels. There are other versions of this database with colorful and gray facial images. We selected 50 individuals and 7 images per individual of gray-version FERET database. All images are tiff format and partial of them are displayed in Figure. The Figure 4.(b) below shows some facial image of this dataset.

GTFD dataset contains the images of 50 people, each person represented in the dataset by 15 facial images. These images are captured with different facial expressions, pose variations, rotations, and several lighting conditions as shown in Figure 4. (c). The image resolutions in the GTFD dataset is fixed to 50 × 50. In this paper, we used just 10 images from 15 images for each person for training and testing the proposed method.

TABLE 2. Recognition rate of the proposed approach compared with PCA [7], PCA-L1 [22], and WTPCA-L1 [38].

Dataset	PCA	PCA-L_1	WTPCA-L_1	DeepWTPCA-L_1
ORL	93.75%	95.00%	96.70%	**99.85%**
GTFD	79.00%	80.40%	85.00%	**93.89%**
FERET	65.86%	67.31%	72.98%	**80.56%**

B. RESULTS AND DISCUSSION

In the first experiment, we exploited the three data set using the following test protocol: we randomly select 5 facial images for each individual for ORL, 10 facial images for each individual for GTFD, 4 images for the FERET facial data set. This selection is intended to form the training set and we took the remaining images as a test set. Then, each test protocol is run ten times to calculate the average recognition accuracy. Table 2 displays the Average performance accuracies of four methods including Traditional-PCA [7], PCA-L1 [22], WTPCA-L1 [38], and the proposed method DeepWTPCA-L1. Based on these results, it can be clearly seen that our approach gives higher recognition accuracy for all three datasets. While, the performance accuracies are 99.85% for ORL, 93.89% for GTFD, and 80.56% for the FERET dataset. Compared with the other state-of-the-art methods, the proposed method is more accurate than these SOTA methods by a difference of 3% to 6% for the ORL dataset, 7% to 14% for GTFD, and 8% to 15% for the FERET dataset. These results shows the impact of the deep learning model, which is used for face identification on the new presentation of features, on the recognition rate.

In the second experiment, we modify the test protocol as follows: we select the first five facial images for ORL, the first ten facial images for GTFD, and the first four facial images for FERET. Then we take the remaining images to form the test set. The main goal of this experiment is to test the robustness degree of our FR learning model to the two forms of noise namely Gaussian noise and Salt and Pepper noise. The variation of the noise densities values are 10^{-4}, 5×10^{-4}, 10^{-3}, 5×10^{-3}, 10^{-2}, 5×10^{-2}, and 10^{-1}.

TABLE 3. Performance of the proposed method and the other methods on noisy images (Salt and Pepper) from on ORL, GTFD, and FERET databases. The bold and underline fonts respectively represent the first and second place.

Method	Dataset	Noise-free	Salt and Pepper noise			
			10^{-4}	10^{-3}	10^{-2}	10^{-1}
PCA [43]	ORL	88.50%	88.50%	88.50%	88.30%	78.25%
	GTFD	70.20%	77.20%	77.12%	76.16%	62.60%
	FERET	61.33%	60.46%	60.08%	58.40%	28.26%
2DPCA [18]	ORL	90.50%	90.50%	90.55%	90.30%	88.95%
	GTFD	77.24%	77.24%	77.12%	76.80%	72.32%
	FERET	66.00%	66.00%	66.00%	65.66%	59.63%
PCA-L_1 [22]	ORL	89.75%	89.75%	89.05%	88.40%	80.50%
	GTFD	79.91%	79.92%	79.72%	79.04%	70.48%
	FERET	63.33%	62.60%	62.06%	61.53%	32.73%
WTPCA-L_1 [38]	ORL	92.10%	92.10%	92.05%	92.15%	88.60%
	GTFD	84.22%	84.20%	83.72%	83.96%	74.88%
	FERET	64.66%	63.80%	63.53%	52.86%	59.33%
LDA [10]	ORL	90.25%	90.25%	90.30%	88.45%	74.20%
	GTFD	78.89%	78.56%	78.24%	76.56%	59.32%
	FERET	64.00%	63.40%	62.80%	57.40%	21.86%
2DLDA [20]	ORL	93.15%	92.95%	92.70%	92.15%	89.85%
	GTFD	80.71%	79.32%	78.92%	78.60%	76.08%
	FERET	65.33%	64.00%	64.06%	62.53%	61.93%
$L_{2,p}$- norm PCA (p=0.5) [44]	ORL	92.25%	92.25%	92.03%	91.06%	87.64%
	GTFD	83.96%	83.96%	84.36%	83.80%	73.29%
	FERET	63.12%	63.12%	62.69%	62.07%	57.64%
Discriminative PCA [45]	ORL	91.45%	90.23%	90.03%	87.96%	87.63%
	GTFD	80.76%	80.32%	77.92%	76.59%	71.35%
	FERET	62.78%	61.30%	61.65%	60.23%	30.89%
DeepWTPCA-L_1 (proposed)	ORL	**98.54%**	**98.54%**	**98.31%**	**97.49%**	**97.08%**
	GTFD	**93.28%**	**93.28%**	**93.21%**	**92.37%**	**90.98%**
	FERET	**79.97%**	**79.97%**	**78.88%**	**78.24%**	**78.09%**

FIGURE 5. Some facial-images of the FERET database with and without noises. The second row and the third row are noisy with Salt & Pepper noises and Gaussian noises respectively.

TABLE 4. The performance of each method on GTFD face database. The bold and underline fonts respectively represent the first and second place.

Method	Recognition Rate
SVD based VR (2018) [46]	64.40%
INNC (2018) [46]	63.60%
Naive CR (2020) [47]	64.00%
Method based on CR (2020) [47]	74.40%
RNLRLSR (2020) [48]	72.00%
CLSR (2020) [48]	65.60%
DWT(SVD/LR+RWLDA/QR)+MIN-MAX (2019) [49]	89.24%
DWT(SVD/LR+RWLDA/QR)+Z-score (2019) [49]	88.20%
CMBZZBP (2020) [50]	91.20%
DeepWTPCA-L_1	**93.89%**

Figure 5 face images after adding noise to images from the FERET database. The obtained results are compared with the existing algorithms that used same procedure including PCA [43], 2DPCA [18], PCA-L1 [22], WTPCA-L1 [38], LDA [10], and 2DLDA [20], L_{2_p}- norm PCA [44], and Discriminative PCA [45]. In addition, the effect of Salt & Pepper and Gaussian noises on each image performance on the three facial databases is shown in Table 3 and Figure 6.

From Table 3 and Figure 6, we can easily notice that the recognition accuracies decreasing with the increasing of noise density. Even with noisy images, the proposed system has achieved high recognition accuracies for different noise density degradations. The accuracy reached 97.08%, 90.98%, and 78.09% on ORL, GTFD, and FERET respectively. Also, from figure 7, we can clearly observe the effect of the added Gaussian noises on the recognition accuracies of our approach applied to the FERET database. From these results, we can find that the recognition accuracy of PCA, 2DPCA, PCA-L1, WTPCA-L1, LDA, and 2DLDA, L_{2_p}- norm PCA, Discriminative PCA decreases with increasing noise variance. However, the recognition accuracy of the proposed system remains the topmost one, which indicates the efficiency and robustness of our proposed approach. Also, comparing with the other methods we can find that the performance deceases rapidly when we add the noise density, while the proposed method accuracies are stable with almost all noise densities or decrease slowly.

C. EVALUATION ON GTFD DATASET

We have also compared the performances of the proposed system with 9 state-of-the-art approaches on GTFD database and adopted the same experimental protocol. Table 4 shows the comparison of the recognitionp rate

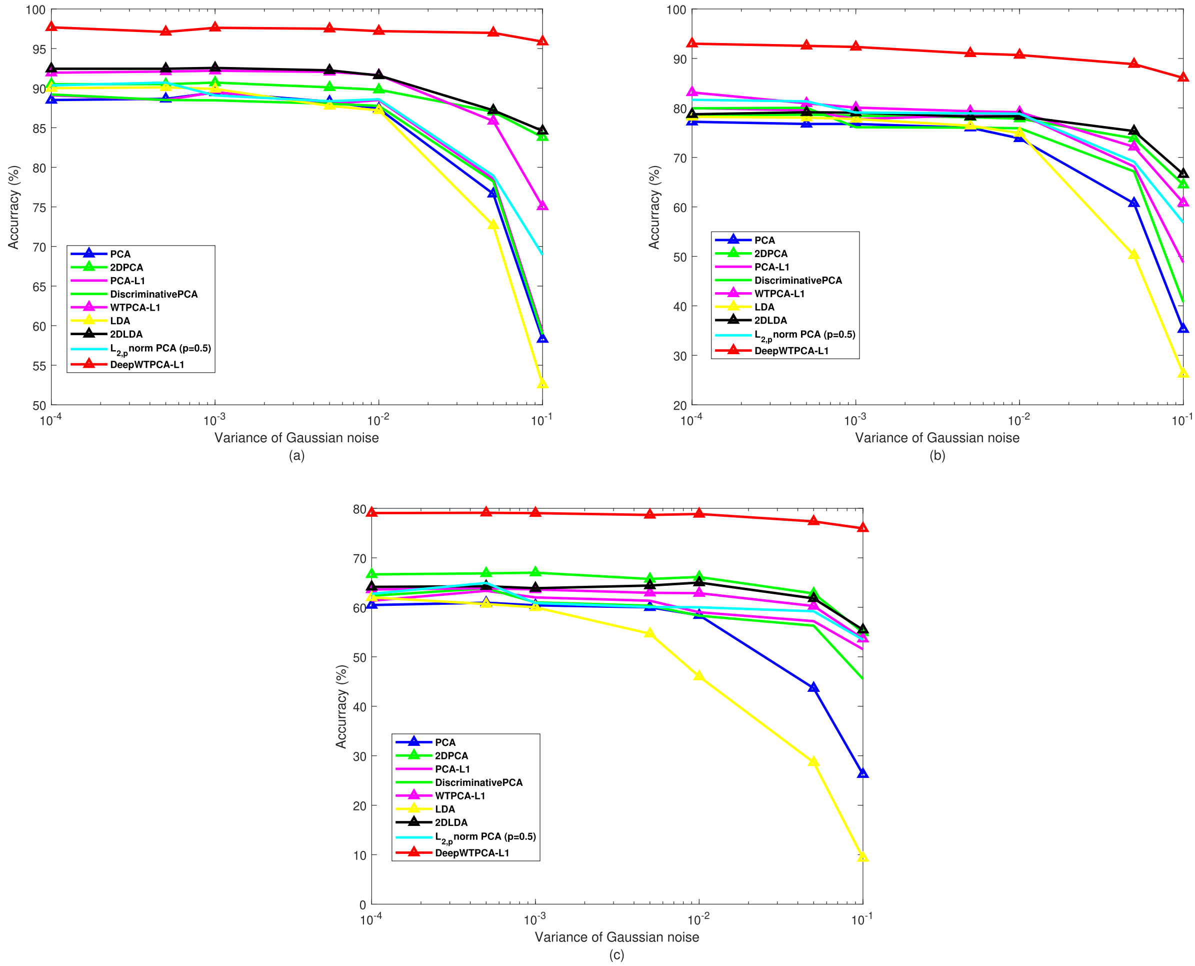


FIGURE 6. Impact of Gaussian-noise on recognition accuracy for three databases: (a) ORL, (b) GTFD, and (c) FERET.

between our system proposed and these methods including SVD based VR [46], INNC [46], Naive CR [47], Method based on CR [47], RNLRLSR [48], CLSR [48], DWT(SVD/LR + RWLDA/QR) using MIN-MAX method [49], DWT(SVD/LR+RWLDA/QR) using Z-score method [49], and CMBZZBP [50]. We can clearly observe that the recognition rate of the proposed system is much better than the other approaches. Our proposed method produces the highest recognition rate of 93.89%, which is better with 2% to the second best results using CMBZZBP and better by more than 10% for the others.

VI. CONCLUSION

In this paper, a new face recognition model has been proposed. By exploiting the WTPCA-L_1 norm as a method for features extraction, while these features are used as input of the proposed deep neural network architecture. The proposed deep learning model is used for face identification and classification. We have used the WTPCA-L_1 algorithm instead of using PCA or PCA-L_1 in order to obtain a better data representation in a low-dimensional space. Our approach not only makes use of the strong robustness of the L_1-norm optimization method to outliers and noises but also utilizes an effective combination of CNN-LSTM networks to improve face recognition performance. Many experiments on ORL, GTFD, and FERET datasets prove that the proposed approach is more robust than several advanced face recognition algorithms in terms of recognition accuracy under noisy and noise-free conditions.

ACKNOWLEDGMENT

The authors are very grateful to the Editor and the anonymous reviewers for their valuable and useful comments, which helped improve the quality of this work.

REFERENCES

[1] W. Xu, Y. Shen, N. Bergmann, and W. Hu, "Sensor-assisted multi-view face recognition system on smart glass," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 197–210, Jan. 2018.

[2] N. Almaadeed, O. Elharrouss, S. Al-Maadeed, A. Bouridane, and A. Beghdadi, "A novel approach for robust multi human action recognition and summarization based on 3D convolutional neural networks," 2019, *arXiv:1907.11272*. [Online]. Available: http://arxiv.org/abs/1907.11272

[3] M. K. Bhowmik, P. Saha, A. Singha, D. Bhattacharjee, and P. Dutta, "Enhancement of robustness of face recognition system through reduced gaussianity in log-ICA," *Expert Syst. Appl.*, vol. 116, pp. 96–107, Feb. 2019.

[4] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, "LFR face dataset: Left-front-right dataset for pose-invariant face recognition in the wild," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, 2020, pp. 124–130.
[5] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2439–2450, May 2019.
[6] I. Masi, S. Rawls, G. Medioni, and P. Natarajan, "Pose-aware face recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4838–4846.
[7] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[8] G. Ghinea, R. Kannan, and S. Kannaiyan, "Gradient-orientation-based PCA subspace for novel face recognition," *IEEE Access*, vol. 2, pp. 914–920, 2014.
[9] A. D. Back and A. S. Weigend, "A first application of independent component analysis to extracting structure from stock returns," *Int. J. Neural Syst.*, vol. 8, no. 4, pp. 473–484, Aug. 1997.
[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[11] W. Zhu and Y. Yan, "Joint linear regression and nonnegative matrix factorization based on self-organized graph for image clustering and classification," *IEEE Access*, vol. 6, pp. 38820–38834, 2018.
[12] Y. Wang and Y. Wu, "Complete neighborhood preserving embedding for face recognition," *Pattern Recognit.*, vol. 43, no. 3, pp. 1008–1015, Mar. 2010.
[13] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006.
[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
[15] S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, and K. D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization," *Opt. Exp.*, vol. 15, no. 7, pp. 4066–4082, Apr. 2007.
[16] S. T. Roweis, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[17] L. Hu, G. Guo, and C. Ma, "Combined new nonnegative matrix factorization algorithms with two-dimensional nonnegative matrix factorization for image processing," *Multimedia Tools Appl.*, vol. 75, no. 18, pp. 11127–11155, Sep. 2016.
[18] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
[19] B. Bai, Y. Li, J. Fan, C. Price, and Q. Shen, "Object tracking based on incremental Bi-2DPCA learning with sparse structure," *Appl. Opt.*, vol. 54, no. 10, pp. 2897–2907, Apr. 2015.
[20] J. Yang, D. Zhang, X. Yong, and J.-Y. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern Recognit.*, vol. 38, no. 7, pp. 1125–1129, Jul. 2005.
[21] Q. Ke and T. Kanade, "Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 739–746.
[22] N. Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–1680, Sep. 2008.
[23] Q. Ye, H. Zhao, L. Fu, and S. Gao, "Underlying connections between algorithms for nongreedy LDA-L1," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2557–2559, May 2018.
[24] C.-N. Li, Y.-H. Shao, and N.-Y. Deng, "Robust L1-norm two-dimensional linear discriminant analysis," *Neural Netw.*, vol. 65, pp. 92–104, May 2015.
[25] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao, "Robust 2DPCA with non-greedy ℓ_1-norm maximization for image analysis," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1108–1112, Aug. 2014.
[26] L. He, J. Ye, and J. E, "Robust L1-norm two-dimensional collaborative representation-based projection for dimensionality reduction," *Signal Process., Image Commun.*, vol. 81, Feb. 2020, Art. no. 115684.
[27] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.
[28] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.
[29] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.
[30] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 499–515.
[31] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2892–2900.
[32] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face recognition with very deep neural networks," 2015, *arXiv:1502.00873*. [Online]. Available: http://arxiv.org/abs/1502.00873
[33] Z. Wang, K. He, Y. Fu, R. Feng, Y.-G. Jiang, and X. Xue, "Multi-task deep neural network for joint face recognition and facial attribute prediction," in *Proc. ACM Int. Conf. Multimedia Retr.*, Jun. 2017, pp. 365–374.
[34] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 142–150.
[35] W. Ou, X. Luan, J. Gou, Q. Zhou, W. Xiao, X. Xiong, and W. Zeng, "Robust discriminative nonnegative dictionary learning for occluded face recognition," *Pattern Recognit. Lett.*, vol. 107, pp. 41–49, May 2018.
[36] H. Roy and D. Bhattacharjee, "A novel quaternary pattern of local maximum quotient for heterogeneous face recognition," *Pattern Recognit. Lett.*, vol. 113, pp. 19–28, Oct. 2018.
[37] M. Hirokawa and Y. Kuroki, "A fast implementation of PCA-L1 using gram-Schmidt orthogonalization," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 3, pp. 559–561, 2013.
[38] A. Maafiri and K. Chougdali, "Face recognition using wavelets based feature extraction and PCA-L1 norm," in *Proc. Int. Conf. Vis. Towards Emerg. Trends Commun. Netw. (ViTECoN)*, Mar. 2019, pp. 1–4.
[39] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.
[40] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Dec. 1994, pp. 138–142.
[41] A. Nefian. *Georgia Tech Face Database 1999*. Accessed: 1999. [Online]. Available: http://www.anefian.com/research/face reco.htm
[42] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
[43] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[44] Q. Wang, Q. Gao, X. Gao, and F. Nie, "$\ell_{2,p}$-norm based pca for image recognition," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1336–1346, Mar. 2017.
[45] H. Qiao, "Discriminative principal component analysis: A reverse thinking," 2019, *arXiv:1903.04963*. [Online]. Available: http://arxiv.org/abs/1903.04963
[46] G. Zhang, W. Zou, X. Zhang, and Y. Zhao, "Singular value decomposition based virtual representation for face recognition," *Multimedia Tools Appl.*, vol. 77, no. 6, pp. 7171–7186, Mar. 2018.
[47] Y. Qin, L. Sun, and Y. Xu, "Exploring of alternative representations of facial images for face recognition," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 10, pp. 2289–2295, 2020.
[48] K. He, Y. Peng, S. Liu, and J. Li, "Regularized negative label relaxation least squares regression for face recognition," *Neural Process. Lett.*, vol. 51, no. 3, pp. 1–19, 2020.
[49] M. Ayyad and C. Khalid, "New fusion of SVD and relevance weighted LDA for face recognition," *Procedia Comput. Sci.*, vol. 148, pp. 380–388, 2019.
[50] S. Karanwal and M. Diwakar, "Two novel color local descriptors for face recognition," *Optik*, vol. 226, Jan. 2021, Art. no. 166007.

AYYAD MAAFIRI was born in Rabat, Morocco, in 1988. He received the M.Sc. degree in computer systems security from Mohamed V University, Rabat, in 2012. He is currently pursuing the Ph.D. degree with the Engineering Science Laboratory, CED-ST Center of Doctoral Studies in Sciences and Technologies, National School of Applied Sciences (ENSA), Ibn Tofail University, Kenitra, Morocco. His current research interests include biometrics, image analysis, and statistical pattern recognition.

OMAR ELHARROUSS (Member, IEEE) received the master's degree from the Faculty of Sciences, Dhar El Mehraz, Fez, Morocco, in 2013, and the Ph.D. degree from the LIIAN Laboratory, USMBA-Fez University, in 2017. His research interests include pattern recognition, image processing, and computer vision.

SAAD RFIFI received the Ph.D. degree in theoretical physics disciplinary from Mohammed V University, in 2015. Since 2017, he has been working as an Assistant Professor with the Department of Physics, Abdelmalek Essaafi University. He is doing academical research with the Theoretical Physics Laboratory, Faculty of Sciences in Rabat, and publishing several articles with that affiliation. The affiliation of academical research is currently changed to be the Advanced Technology and Sciences Laboratory. He is working in many scientific fields as quantum cryptography and the quantum information.

SOMAYA AL-MAADEED (Senior Member, IEEE) received the Ph.D. degree in computer science from Nottingham, U.K., in 2004. She is currently the Head of the Computer Science Department, Qatar University, where she is also the Coordinator of the Computer Vision and AI Research Group. She enjoys excellent collaboration with national and international institutions and industry. She is a Principal Investigator of several funded research projects generating approximately five million. She has published extensively pattern recognition and delivered workshops on teaching programming for undergraduate students. In 2015, she was elected as the IEEE Chair for the Qatar Section. She attended workshops related to higher education strategy, assessment methods, and interactive teaching.

KHALID CHOUGDALI received the Ph.D. degree in computer science from Mohamed V-Agdal University, in 2010. He is currently an Associate Professor of computer science with the National School of Applied Sciences, Ibn Tofail University, Kenitra. His main research interests include information security, pattern recognition, and biometrics.