

# Face Recognition Employing DMWT Followed by FastICA

Ahmed Aldhahab<sup>1,2</sup>  · Wasfy B. Mikhael<sup>1</sup>

Received: 22 October 2016 / Revised: 24 August 2017 / Accepted: 29 August 2017 /  
Published online: 21 September 2017  
© Springer Science+Business Media, LLC 2017

**Abstract** Face recognition becomes a challenging topic in several fields since images of faces are varied by changing illuminations, facial rotations, facial expressions, etc. In this paper, two dimensional discrete multiwavelet transform (2D DMWT) and fast independent component analysis (FastICA) are proposed for face recognition. Pre-processing, feature extraction, and classification are the main steps in the proposed system. In the preprocessing step, each pose in the database is divided into six parts to reduce the effect of unnecessary facial features and highlight the local features in each part. For feature extraction, the 2D DMWT is applied to each part for dimensionality reduction and features extraction. This results in two facial representations. Then FastICA followed by  $\ell_2$ -norm is applied to each representation, which produces six and three different techniques for the first and second representation, respectively. This results in features that are more discriminating, less dependent, and more compressed. In the recognition step, the resulted compressed features from the two representations are fed to a neural network-based classifier for training and testing. The proposed techniques are extensively evaluated using five databases, namely ORL, YALE, FERET, FEI, and LFW, which have different facial variations, such as illuminations, rotations, facial expressions, etc. The results are analyzed using  $K$ -fold cross-validation. Sample results and comparison with a large number of recently proposed approaches are pro-

---

✉ Ahmed Aldhahab  
aldhahab2012@knights.ucf.edu

Wasfy B. Mikhael  
Wasfy.Mikhael@ucf.edu

<sup>1</sup> Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816, USA

<sup>2</sup> Department of Electrical Engineering, University of Babylon, Hilla, Babylon, Iraq

vided. The proposed approach is shown to yield significant improvement compared with the other approaches.

**Keywords** Discrete wavelet and multiwavelet transform · Independent component analysis (ICA) · FastICA · Neural network · Face recognition

## 1 Introduction

Face recognition has received considerable attention in computer vision, pattern recognition, and image processing. Face recognition is a popular research topic due to its wide applications in access control, personal identification, and security in which, for example, the identified person can access a specific building, confidential documents, etc. Instead of authenticating people through physical and virtual domains, which are based on passwords, PIN, smart cards, etc., that are subjected to be forgotten, stolen, etc., the face recognition is promising to recognize the individual. Face recognition does not require a direct contact with the acquisition devices or a voluntary action compared with the other biometrics, such as iris detection/recognition, finger print, etc., that require a voluntary action by the users. Also, faces can be captured and detected from far using inexpensive fixed small camera, while other biometrics might require expensive equipment [30,43].

There are a large number of algorithms that have been proposed for face recognition. Designing an efficient facial recognition system is a challenging task due to the large dimensions of the data. Data compression/feature extraction is always performed. The most broadly used subspace method for face recognition is principal component analysis (PCA) [47]. Truk and Pentland [56] applied the eigenfaces method for face recognition. The eigenfaces, defining a feature vector space that dramatically reduces the dimensions of the original faces, corresponded to the eigenvectors associated with the dominant eigenvalues of the known face Covariance matrix. The performance of PCA suffers significant degradation when the training face images have different facial variations in pose, illuminations, and expressions. Fast  $\ell_1$ -PCA was reported in [34] to improve PCA performance under facial variations and was integrated with a greedy search algorithm in [31] to achieve better results.

Another subspace approach called linear discriminant analysis (LDA), fisherfaces, is a supervised method. LDA constructs a linear subspace in which data are optimally discriminated by maximizing between class scatters and minimizes within class scatters in the projective linear subspace. It is widely used for facial representation and recognition [29]. Locality preserving projection (LPP), an alternative approach to PCA, seeks to find the optimal linear approximation to the eigenfunctions on the face manifold. LPP is used for dimensionality reduction and feature selection; hence, it is employed for facial representation and classification [51]. Local features were extracted using LPP approach in [7] for facial recognition in which the effect of facial variations was reduced. An extended version of PCA is an independent component analysis (ICA). In contrast to PCA that only considers the second-order statistics, ICA employs the high-order statistics in the data. ICA is successfully used for feature selection in image processing applications, such as face recognition [8]. In such

application, the basis images extracted using ICA are as independent as possible compared with the basis images extracted using PCA that are uncorrelated. Both ICA and supported vector machine (SVM) were introduced in [60] for face recognition. Non-negative matrix factorization (NMF) [36] has been recently shown to be one of the most powerful decomposition tools for multivariate data, which is inherent into several applications in image processing, computer vision, etc. NMF requires that the projective vectors and associated weighted coefficients are non-negative. Sparse spatial facial information was extracted in [40] using NMF for face recognition.

Discrete wavelet transform (DWT), based on multiresolution analysis (MRA), is broadly used in different fields. DWT is used for dimensionality reduction, feature extraction, and noise mitigation. Based on favorable properties of wavelet filters, DWT integrated with the other approaches, such as PCA, ICA, etc., to achieve superior performance. In [44], 2D PCA was applied to the approximation coefficients of the third level of DWT decomposition for feature selection/data compression. The authors in [44] demonstrated that using the integrated tools PCA/DWT led to better results than those reported using PCA or DWT alone. ICA and 2D DWT were proposed in [41] for facial representations and classification. In [41], ICA was applied to different levels of DWT decomposition and the results obtained using the integrated tools were higher than the results reported using traditional ICA or DWT.

Although discrete cosine transform (DCT) is well known in image compression, DCT is widely used for facial representations/classification. Tools are integrated to overcome the weakness that found in a single transform and to obtain features less sensitive to the facial variations. A hybrid technique based on DCT and Gabor wavelet transform (GWT) was presented in [2] to obtain discriminant facial features. Facial images were efficiently represented using the integrated tools DCT/PCA for feature extraction and classified using NN [10]. Vector quantization (VQ) is successfully used in image compression since Linde, Buzo, and Gray (LBG) published their distinguishable work in 1980 [39]. Four different algorithms-based VQ were studied in [45] for facial representations. In [45], the algorithms were tested against DCT-based system, and the Kekre's fast codebook generation (KFCG) algorithm accomplished highest results among all the algorithms.

Multiwavelet transform (MWT), based on MRA, is used for signal representation and data compression. Filters used in MWT have more degree of freedom than those filters used in WT. 2D DMWT, 2D Radon Transform (2D RT) that was used to align the features around the origin, and 2D DWT were proposed in [3] to achieve better facial representations and less storage requirements. Infrared facial images were classified using 2D PCA/DMWT [62]. For further compaction in [62], 2D PCA was applied to different MWT sub-bands and thereafter each sub-band was differently weighted using Fisher discriminant criterion. In [5], Facial images were efficiently represented using eigenvalues and eigenvectors of ICA based on MRA features.

Artificial neural network (ANN) and deep learning (DL) are used to improve the performance of the face recognition system. DL-based multimodal face representation was proposed in [17] to improve the system performance that evaluated using faces that appeared in multimedia applications, such as social media. Lei et al. [37] proposed stacked image descriptor (SID), which employs deep structure, for face recognition. More complex facial information was extracted using deep structure. Also, features'

compactness and discriminant were improved. A better result was accomplished when sparse autoencoder SADL was applied to the facial images [61].

The traditional canonical correlation analysis (CCA) can address the relationship between two multivariate data sets, through which some of the important information can be missing. To overcome this problem, the authors in [12] proposed intra-facial-features CCA, which involves multiple multivariate data sets, for face recognition. Moments are widely used for feature extraction due to their superior discriminatory power and geometrical invariance. Zernike moments and Hermite kernels were proposed in [20] for face recognition. Rahman et al. [42] showed that the 2D Gaussian–Hermite moments (2D-GHMs) has an ability to capture higher-order hidden non-linear 2D structures within images. Thereafter, Rahman et al. [50] used orthogonal 2D Krawtchouk moments (2D KCMs) for face recognition since KCMs have the ability to capture region-based higher-order hidden non-linear structures from discrete coordinates of finitely supported images.

To overcome the limitations of some algorithms, such as PCA, LDA, and LPP when applied to the databases under facial variations, local-based features can achieve promising results [38]. Extracting local features from facial images can be done automatically using suitable tools or by partitioning faces into several parts. The most common methods used for extracting local features are LBP [1] and GWT [57]. Duan and Tan in [19] segmented the facial images to  $n$  overlapping subregions. Then, PCA applied to each subregion for further feature compression. Although the complexity in [19] was increased, the highest rates achieved were 92.4% based on FERET database.

Several efficient methods for representing image data have been exploited for features extraction. Recently, 2D DMWT and FastICA were used for processing image signal. In this contribution, a new face recognition system is presented that employs successive application of 2D DMWT followed by FastICA on partitioned facial images. The contribution of the paper is as follows:

1. The first key is dividing the input facial image into 6 overlapping parts, each contains some of the local facial features, which are eyes, mouth, and nose. The motivation behind this is to identify people using local facial features. Also, this is performed to reduce the effect of unnecessary information, which is in common among all poses/persons, such as background. Also, partitioning improves the system performance since each part is participated with the other two parts, which increase the matching probability.
2. The second key is achieving less storage requirements as well as high recognition rates. In this paper, one level of DMWT decompositions is applied to each part. Each part is divided into four frequency sub-bands, which are low–low (LL), low–high (LH), HL, and HH pass filters. It is shown later that most of the information is localized in the LL sub-bands; hence, the LL sub-band is retained while all other sub-bands are eliminated. Therefore, the dimensionality reduction (less storage requirements) is accomplished. Furthermore, each sub-band is further divided into four sub-images. For each pose, there are two representations resulted from DMWT step, as shown in Fig. 7. The first representation is using all sub-images in the LL sub-band, while the second one is using only one sub-image.

3. The FastICA is applied to the two representations. This step is performed to produce discriminated and independent features since ICA is robust to the facial variations as reported in [8, 18, 59]. Then the FastICA is followed by  $\ell_2$ -norm for further dimensionality reduction and data compression. As a result, six different techniques are established in the first representation, and three different techniques are formed in the second representation.
4. Finally, the compressed/discriminated local features are fed NN-based classifier for the recognition task.

The proposed techniques of two representations are extensively evaluated using five databases, namely ORL, YALE, FERET, FEI, and LFW that have different facial variations, such as facial expressions, facial rotations, illuminations, makeups. Then, the experimental results are analyzed using  $K$ -fold cross-validation (CV) [6]. The proposed techniques are shown to improve the recognition accuracy in comparison with a large number of recently proposed methods. The performance improvement has been consistently obtained.

The rest of the paper is organized as follows. In Sect. 2, background about 2D DMWT and FastICA is discussed. The proposed techniques are presented and broadly explained in Sect. 3. In Sect. 4, the experimental results are illustrated. Discussion of the experimental results is provided in Sect. 5. The neural network performance of the proposed techniques is presented in Sect. 6. Finally, we give some concluding remarks in Sect. 7.

## 2 Background: Wavelet, Multiwavelet and FastICA

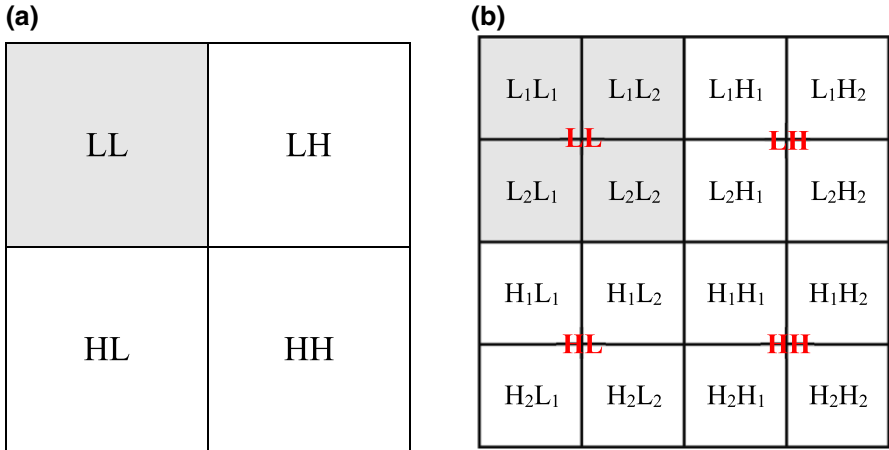
### 2.1 Discrete Wavelet and Multiwavelet Transform

The DWT is an extended version of the discrete Fourier transform (DFT). DWT, based on MRA, has one scaling function  $\phi(t)$  and one wavelet function  $\psi(t)$ . Both functions, ( $\phi(t)$  and  $\psi(t)$ ), are associated with low pass and high pass filters, respectively, [9]. The 2D DWT for 2D image  $f(x, y)$  is expressed as:

$$W_\phi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \phi_{j_0, m, n}(x, y) \quad (1)$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x, y) \quad (2)$$

where  $i = \{H, V, D\}$ , directional wavelets,  $j_0$  is an arbitrary start scale.  $\phi_{j_0, m, n}(x, y)$  and  $\psi_{j, m, n}^i(x, y)$  are the scaled and translated basis functions, respectively.  $W_\phi(j_0, m, n)$  represents the approximation of the image  $f(x, y)$  at scale  $j_0$ . The  $W_\psi^i(j, m, n)$  denotes to the horizontal ( $H$ ), vertical ( $V$ ), and diagonal ( $D$ ) coefficients at scale  $j \geq j_0$ . Usually,  $j_0 = 0$  and  $N = M = 2^J$  are selected such that  $j = 0, 1, \dots, J - 1$  and  $m = n = 0, 1, \dots, 2^j - 1$  [23]. Each level of DWT decompositions is dividing the image into four sub-bands, namely LL, LH, HL, and HH as



**Fig. 1** a, b 1-Level of decomposition for DWT and DMWT, respectively

shown in Fig. 1a, where L and H are the low and high pass filters, respectively. The LL sub-band is the approximation coefficients of the image. The LH, HL, and HH are the detail coefficients, which are the vertical, horizontal, and diagonal directions of the original image, respectively. Most of the image information is localized in LL sub-band.

The MRA for multiwavelet case is the same as the standard MRA except that MWT uses  $Q$  scaling functions,  $\phi_1(t-k), \phi_2(t-k), \dots, \phi_Q(t-k)$ . The  $Q$  scaling functions expressed in the vector notation as  $\Phi(t) = [\phi_1(t), \dots, \phi_Q(t)]^T \in L^2(R)^Q$ , which is called multiscaling functions. Analogue to the scalar case, this can satisfy the matrix dilations as:

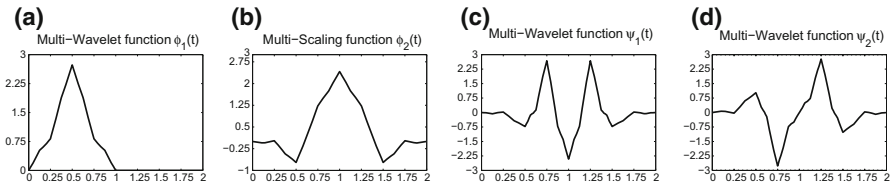
$$\Phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} H_k \cdot \Phi(2t - k) \tag{3}$$

where  $H_k$  is  $Q \times Q$  matrix dimensions for each integer  $k$  and it is associated with the low pass filter. The  $Q$  scaling functions are associated with  $Q$  wavelet functions that can satisfy the vector notation as  $\psi(t) = [\psi_1(t), \dots, \psi_Q(t)]^T \in L^2(R)^Q$ , which is called multiwavelet functions. The  $Q$  multiwavelet functions are satisfying wavelet matrix equation as:

$$\Psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} G_k \cdot \Phi(2t - k) \tag{4}$$

where  $G_k$  is  $Q \times Q$  matrix dimensions for each integer  $k$  and  $G_k$  is associated with high pass filter [53,54].

One of the famous multiwavelets system was constructed by Geronimo, Hardian, and Massopust (GHM) [21]. Their system uses multiplicity  $Q = 2$ . In such case, there are two scaling functions  $\Phi(t) = [\phi_1(t) \ \phi_2(t)]^T$  as shown in Fig. 2a, b, and two



**Fig. 2** Multiscaling and multiwavelet functions for GHM filter

wavelet functions  $\Psi(t) = [\psi_1(t) \ \psi_2(t)]^T$  are shown in Fig. 2c, d [53]. The matrix elements in the GHM filter provide extra degree of freedom compared with the scalar case. The GHM filter are combining the orthogonality, high order of approximation, and symmetry. Also, both multiscaling and multiwavelet functions of GHM filter are quite smooth [53]. For GHM case, both Eqs. (3) and (4) are written as [53]:

$$\Phi = \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix} = \sqrt{2} \sum_k H_k \cdot \begin{bmatrix} \phi_1(2t - k) \\ \phi_2(2t - k) \end{bmatrix} \tag{5}$$

$$\Psi = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \end{bmatrix} = \sqrt{2} \sum_k G_k \cdot \begin{bmatrix} \phi_1(2t - k) \\ \phi_2(2t - k) \end{bmatrix} \tag{6}$$

Hence, the dilation Eq. (3) and wavelet Eq. (4) can be written as:

$$\Phi = \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix} = H_0\Phi(2t) + H_1\Phi(2t - 1) + H_2\Phi(2t - 2) + H_3\Phi(2t - 3), \tag{7}$$

$$\Psi = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \end{bmatrix} = G_0\Phi(2t) + G_1\Phi(2t - 1) + G_2\Phi(2t - 2) + G_3\Phi(2t - 3), \tag{8}$$

and  $H_k$  and  $G_k$ , where  $k = 0, 1, 2, 3$ , are expressed as [53,54]:

$$H_0 = \begin{bmatrix} \frac{3}{5\sqrt{2}} & \frac{4}{5} \\ \frac{-1}{20} & \frac{-3}{10\sqrt{2}} \end{bmatrix}, \quad H_1 = \begin{bmatrix} \frac{3}{5\sqrt{2}} & 0 \\ \frac{1}{20} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad H_2 = \begin{bmatrix} 0 & 0 \\ \frac{9}{20} & \frac{-3}{10\sqrt{2}} \end{bmatrix}, \quad H_3 = \begin{bmatrix} 0 & 0 \\ \frac{-1}{20} & 0 \end{bmatrix}$$

$$G_0 = \begin{bmatrix} \frac{-1}{20} & \frac{-3}{10\sqrt{2}} \\ \frac{1}{10\sqrt{2}} & \frac{3}{10} \end{bmatrix}, \quad G_1 = \begin{bmatrix} \frac{9}{20} & \frac{-1}{\sqrt{2}} \\ \frac{-9}{10\sqrt{2}} & 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} \frac{9}{20} & \frac{-3}{10\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{-3}{10} \end{bmatrix}, \quad G_3 = \begin{bmatrix} \frac{-1}{20} & 0 \\ \frac{-1}{10\sqrt{2}} & 0 \end{bmatrix}$$

There are four favorable properties for GHM scaling functions [53,54]:

1. The short support for both scaling functions is [0, 1] and [0, 2].
2. The approximation of the GHM system is based on the second order.
3. Translated version of the scaling and the wavelet functions are orthogonal.
4. Both scaling and wavelet function are symmetric.

The transformation matrix  $T$ , GHM filter matrix, of DMWT is constructed using the above scaling and wavelet functions and can be written as [53,54]:

$$T = \begin{bmatrix} H_0 & H_1 & H_2 & H_3 & 0 & 0 & \dots \\ G_0 & G_1 & G_2 & G_3 & 0 & 0 & \dots \\ 0 & 0 & H_0 & H_1 & H_2 & H_3 & \dots \\ 0 & 0 & G_0 & G_1 & G_2 & G_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix} \tag{9}$$

As we mentioned that both  $H_k$  and  $G_k$  are associated with low pass and high pass filters, respectively. They are  $2 \times 2$  matrix dimensions for each integer  $k$  [53,54]. Hence, the transformation matrix  $T$  is written as [52–54]:

$$T = \begin{bmatrix} H_{0,0} & H_{0,1} & H_{1,0} & H_{1,1} & H_{2,0} & H_{2,1} & H_{3,0} & H_{3,1} & 0 & 0 & 0 & 0 & \dots \\ H_{0,1} & H_{0,1} & H_{1,1} & H_{1,1} & H_{2,1} & H_{2,1} & H_{3,1} & H_{3,1} & 0 & 0 & 0 & 0 & \dots \\ G_{0,0} & G_{0,1} & G_{1,0} & G_{1,1} & G_{2,0} & G_{2,1} & G_{3,0} & G_{3,1} & 0 & 0 & 0 & 0 & \dots \\ G_{0,1} & G_{0,1} & G_{1,1} & G_{1,1} & G_{2,1} & G_{2,1} & G_{3,1} & G_{3,1} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & H_{0,0} & H_{0,1} & H_{1,0} & H_{1,1} & H_{2,0} & H_{2,1} & H_{3,0} & H_{3,1} & \dots \\ 0 & 0 & 0 & 0 & H_{0,1} & H_{0,1} & H_{1,1} & H_{1,1} & H_{2,1} & H_{2,1} & H_{3,1} & H_{3,1} & \dots \\ 0 & 0 & 0 & 0 & G_{0,0} & G_{0,1} & G_{1,0} & G_{1,1} & G_{2,0} & G_{2,1} & G_{3,0} & G_{3,1} & \dots \\ 0 & 0 & 0 & 0 & G_{0,1} & G_{0,1} & G_{1,1} & G_{1,1} & G_{2,1} & G_{2,1} & G_{3,1} & G_{3,1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix} \tag{10}$$

Figure 1b shows the resultant of applying one level of DMWT decomposition to 2-D signal. It is obvious there that the input data is divided into four main sub-bands and each sub-band is further divided into four sub-sub-bands. For example, the sub-sub-band  $L_2H_1$  corresponds to the data from the first channel (high pass filter) in the horizontal direction and second low pass filter in the vertical direction. The total number of sub-bands related to the number of decompositions can be expressed as  $4 + 12 \times L^*$ , where  $L^*$  is the number of multiwavelet decomposition levels.

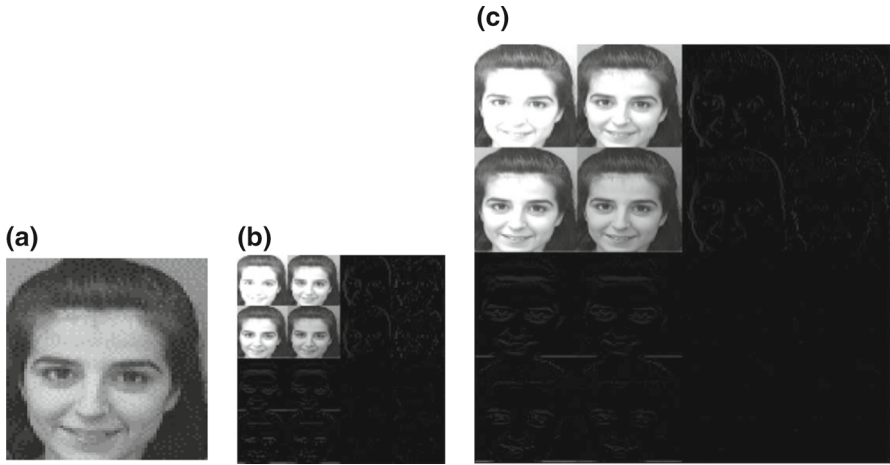
In contrast to the DWT, the DMWT has several favorable features through achieving a good reconstruction (orthogonality), better performance (linear phase symmetry), high order of approximation (vanishing moments), and compact support. These desirable features cannot be achieved at the same time in the scalar wavelet, while in MWT provide more degree of freedom and give good performance in signal and image applications [53].

2.1.1 Preprocessing for Multiwavelet

In multiwavelet, both low pass filter  $H_k$  and high pass filter  $G_k$  are  $Q \times Q$  matrix filter. DMWT is required a vector valued input signal compared to DWT that uses a scalar input signal during the convolution step. In MWT, the signal is passed through the so-called preprocessing step. In the preprocessing step, a sequence of an input vectors is required to be extracted from the scalar input to achieve better performance [11,53,54].

Let’s assume that the dimensions of the input image after resizing to power of two are  $Z \times Z$ . The preprocessing is called oversampling scheme (repeated row preprocessing) if it produces  $Z$ -length 2 vector. The preprocessing is called critical sampling scheme (approximation-based preprocessing) if the preprocessing produces





**Fig. 3** Shows different preprocessing algorithms. **a** Original image. **b** Application of DMWT using critically sampled preprocessing scheme. **c** Illustrates the result of DMWT using oversampled scheme preprocessing

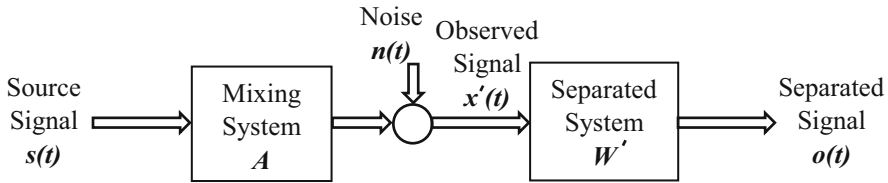
$Z/2$ -length 2 vector. In this paper, we used critically sampled preprocessing scheme (approximation-based preprocessing) [11, 53, 54]. Figure 3 shows an example of applying DMWT to facial image using different preprocessing techniques to the facial image. For sake of brevity, we have omitted the mathematical details of computing 1-D and 2-D MWT preprocessing algorithms. Reader refers to [11, 53, 54] for further mathematical details. The dimensions of the transformation matrix  $T$  using critically sampled preprocessing scheme has the same input dimensions.

### 2.2 Independent Component Analysis (ICA)

Blind source separation (BSS) is a method of restoring the original source signal from a set of mixed signals without any prior knowledge about the source signals and mixing signals. Several algorithms and approaches have been proposed to solve the BBS problems, such as PCA [32], Projection Pursuit [22], etc. ICA is developed to solve the BSS problems, and it is a powerful tool in signal processing and data analysis [8, 28].

The principle of the ICA is to express a set of random variables as a linear combination of statistically independent components (source signals), where both source signal and mixing matrix are unknown. Figure 4 shows the block diagram of ICA. Let a vector  $S = [s_1, s_2, \dots, s_r]^T$  denote as  $r$ -dimensional unknown source signals with zero mean and  $X' = [x'_1, x'_2, \dots, x'_r]^T$  denote as  $r$ -dimensional observed signals. Let assume a matrix  $A$  is a mixing matrix, then the mixture module can be expressed as:

$$X' = A.S = \sum_{i=1}^r a_i s_i, \quad i = 1, 2, 3, \dots, r \tag{11}$$



**Fig. 4** Block diagram for independent component analysis (ICA)

where  $A = (a_1, a_2, \dots, a_r) \in R^{r \times r}$  is a full rank matrix and  $a_i$  is the basis vector of the mixing matrix.  $s_i$  is the independent components [58]. Based in ICA assumptions explained in detail in [58], the noise  $n(t)$  introduced in Fig. 4 is very small compared with the observed signal  $X'$ ; hence, it is neglected.

Each observed  $x'_j$  is a linear combination of weighted independent components  $s_i$ . Based on several assumptions and constrains corresponding to ICA problems, which are explained in detail in [58], the ICA algorithm attempts to find the transformation matrix  $W'$  of the mixing signals such that

$$O = W'.X' = W'.A.S \quad (12)$$

is an estimation of the independent source signal [18,58]. ICA, which is an extension and a generalization of PCA that decorrelates the second-order statistic, is well known to decorrelate the high-order statistic of the input data. Therefore, only if the source signal has a Gaussian distribution, the PCA can be performed to achieve the separation. This is due to the fact that the uncorrelation is equivalent to the statistical independence for normal distribution random process. Hence, the signal can be described by only the second-order statistic. Data representation using ICA is more meaningful than data representation based on using PCA [8,58]. In the field of the face recognition, important information may be localized in the high-order relationships between pixels. Also, ICA can capture the local characteristics of the input face images. Due to these favorable properties of ICA, the ICA bases vectors are extremely robust to the natural face variations such as facial expressions, rotations, and illuminations [8,18,59]. As discussed in [8,27,58], there is another way of representing the input facial images by using unknown mixing matrix  $A$  that is shown in Eq. 12.

Both centering and whitening are preprocessing steps for ICA. Applying linear transformation to the  $r$ -dimensional signal  $X'$  such that the output  $O = W'.X'$  is a white signal which is the so-called whitening process. Before employing ICA, a transformation matrix  $W'$  is applied to the observed vector  $X'$  such that the covariance matrix of the output  $O = W'.X'$  is an identity matrix  $I$ . Let  $X'$  be  $r$ -dimensional random variables with zero mean and  $C_{X'}$  is

$$C_{X'} = Cov(X') = E[X'.X'^T] = ED'E^T \quad (13)$$

a covariance matrix of  $X'$  and it is positive definite.  $D'$  is the diagonal matrix of its eigenvalues and  $E$  is an orthogonal matrix of its eigenvectors such that  $EE^T = I$ .

Therefore, the linear whitening transform is  $W' = D'^{-\frac{1}{2}}E^T$ . Applying whitening process to the observed signal leads to the transformation matrix  $W'$  be an orthogonal matrix; hence, the output vector  $O$  is a white. ICA uses whitening preprocessing to reduce the complexity of the problem [8, 18, 58].

### 2.2.1 Fast Independent Component Analysis (FastICA)

FastICA, which is first proposed and developed by [27, 28], is a faster algorithm for ICA. There are several advantages of using FastICA compared to ICA algorithms, such as FastICA has faster convergence, FastICA does not required to select a step size compared with ICA based on gradient-based algorithm. FastICA finds the independent non-Gaussian signal using any arbitrary non-linear function  $\theta$ , while other algorithms required to evaluate the PDF of the selected non-linear function [27, 58].

The method of FastICA tries to find a local extreme of the Kurtosis of the observed mixing signals. The target of the fixed point is to find the maximum of  $o_i = w_i^T x'$  that has maximum non-Gaussianity. FastICA uses negentropy as the objective function [58]. [26] introduced a simple way to approximate the negentropy function:

$$J'(o_i) \approx [E \{G'(o_i)\} - E \{G'(v)\}]^2 \quad (14)$$

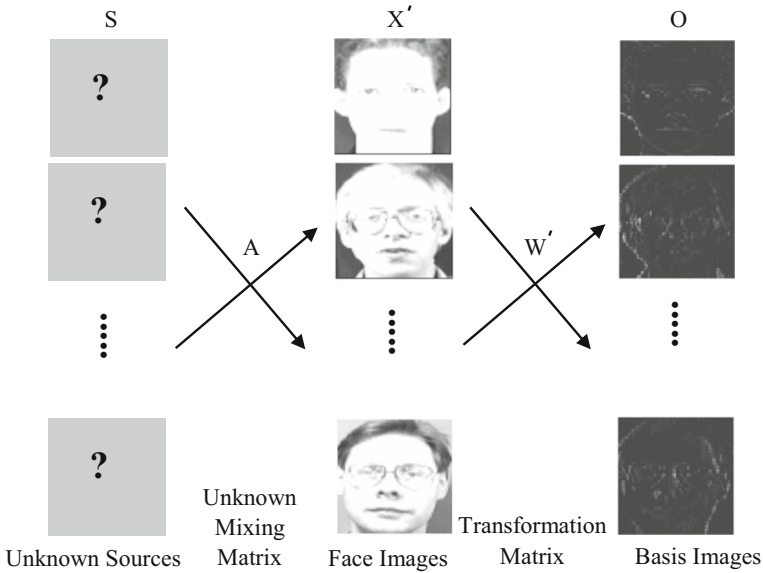
where  $o_i$  is the output random variable with zero mean and unit variance,  $G'$  is non-quadratic function, and  $v$  is a Gaussian random variable with zero mean and unit variance. The approximation given in Eq. (14) leads to use new objective function for estimating ICA. Hence, maximizing the function  $J'_{G'}$

$$J'_{G'(w')} = [E \{G'(w'^T x')\} - E \{G'(v)\}]^2 \quad (15)$$

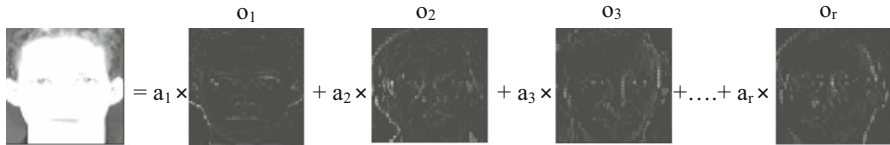
gives one independent component and  $w'$  is the  $r$ -dimensional weighted vector. The algorithm for FastICA based on maximum negentropy objective function is explained in details in [26–28, 58].

### 2.2.2 ICA Architectures

There are two different ICA architectures employed for the face recognition. In architecture-I, which is used in this paper as shown in Fig. 5, the goal is to find a statistically independent basis images [8, 18]. In this architecture, the input facial images are organized in the rows and pixels are in the columns. The input facial images ( $X' = A.S$ ) are considered as a linear combination of unknown statistically independent components  $S$  by unknown mixing matrix  $A$ . Then the basis components (images) are found by projected input facial images  $X'$  onto transformation matrix  $W'$ . As mention earlier, the goal of ICA is to find the transformation matrix  $W'$  such that the rows of  $O = W'.X'$  are as statistically independent as possible. Thereafter, the statistically independent source images estimated from the rows of  $O$  are considered as basis images that are used to construct the input facial images [8, 18].



**Fig. 5** ICA model for architecture-I



**Fig. 6** Reconstruct the input face image using the independent basis images

For input facial image reconstruction as shown in Fig. 6, the rows of  $O$  are used with the linear coefficients  $a_i$ . These coefficients are contained in the inverse of transformation matrix  $W'^{-1} \triangleq A$  [8, 18].

In architecture-II, the input facial images in  $X'$  are arranged so that the images are in the columns and pixels are in the rows. Therefore, the ICA basis are arranged in the columns of  $O = W'.X'$ . The independent basis coefficients contained in the columns of  $O$  are used with the coefficients contained in  $W'^{-1}$  to reconstruct the input facial images in  $X'$  [8, 18, 58].

### 3 Proposed System

The proposed system consists of three main phases as shown in Fig. 7. In the first phase, all poses in the databases are divided into six partitions to reduce the effect of unnecessary information on the system performance and to highlight the local features. The second phase aims to find the most efficient and discriminating features from the set of different databases. The determined discriminated features enhance the recognition rates. Finally, NN based on back propagation training algorithm (BPTA)

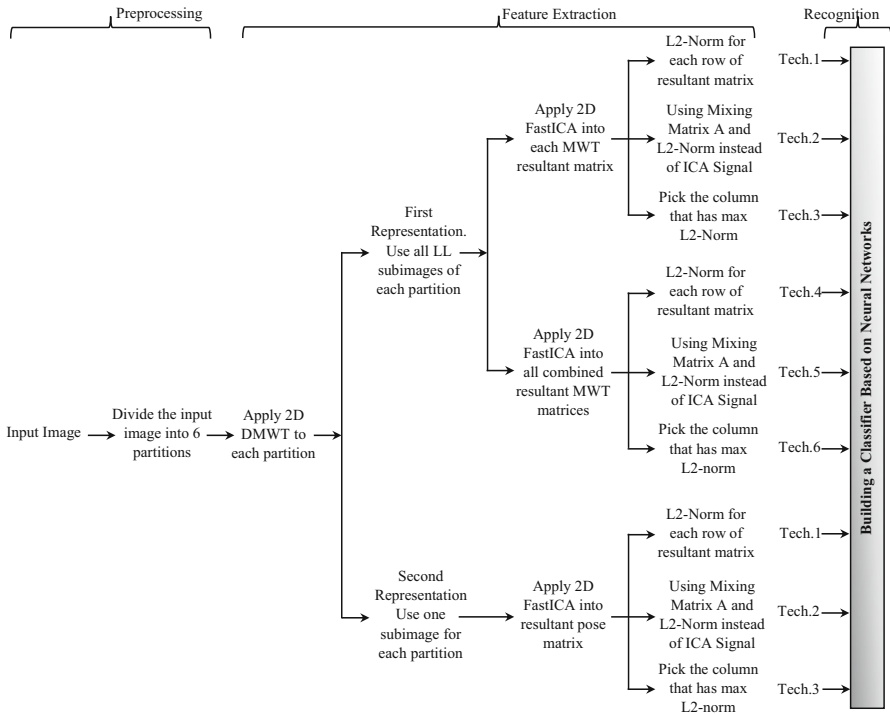


Fig. 7 The proposed system

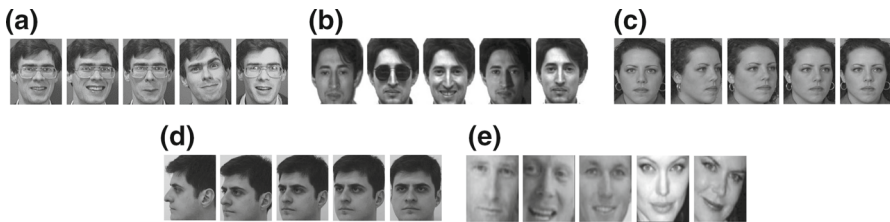


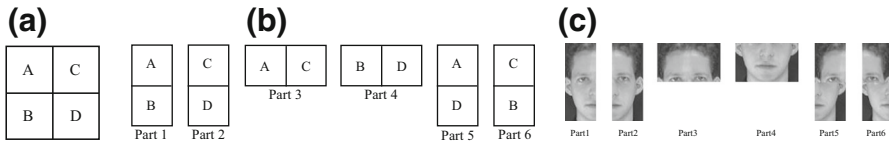
Fig. 8 Shows samples of databases after cropping, a ORL, b YALE, c FERET, d FEI, e LFW

is applied to the extracted features to build a classifier system. Five databases are used to evaluate the proposed techniques, namely ORL, YALE, FERET, FEI, and LFW that have different facial configurations. Samples of each database are shown in Fig. 8.

### 3.1 Preprocessing

The preprocessing phase consists of several steps:

1. Apply cropping technique to the databases used as shown in Fig. 8. This step aims to remove most of the background information (unnecessary information) found in each pose.



**Fig. 9** How each pose of each person is divided into 6 parts, **a** image, **b** parts, **c** example of ORL database

**Table 1** The dimensions of the databases

Databases	Actual size	Dimensions after cropping	Dimensions of each part	Proposed dimension of each part
ORL	112 × 92	112 × 92	112 × 46 OR 56 × 92	64 × 64
YALE	243 × 320	180 × 150	180 × 75 OR 90 × 150	128 × 128
FERET	384 × 256	220 × 170	220 × 85 OR 110 × 170	128 × 128
FEI	480 × 640	320 × 290	320 × 145 OR 160 × 290	256 × 256
LFW	250 × 250	64 × 64	64 × 32 OR 32 × 64	32 × 32

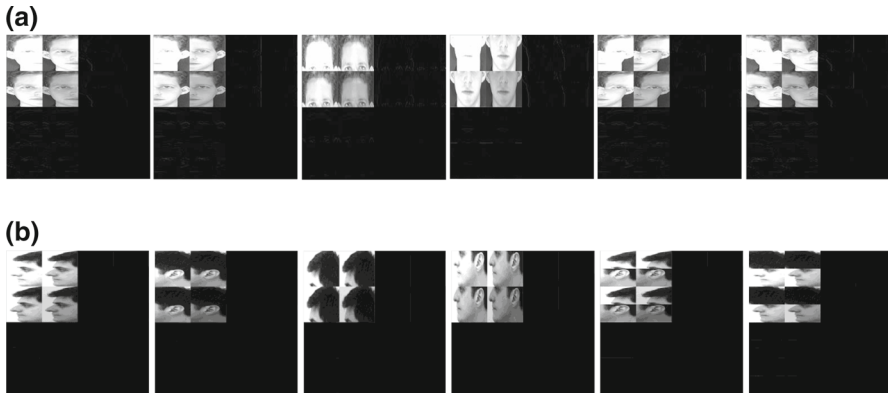
2. Divide all poses of all persons in the databases to 6 partitions, see Fig. 9. The motivation from partitioning is to differentiate the local features in each part. This reduces the effect of the common features, which are mutual in some partitions between persons, on the overall system performance.
3. Convert dimensions of each part of each pose into adequate dimensions as shown in Table 1. The proposed dimensions were chosen since the algorithm used in this paper required dimensions that are power of two.
4. Apply a critically sampled preprocessing scheme (approximation-based preprocessing) to the input facial images [11,53,54].

### 3.2 Feature Extraction

Variations in the faces of the subjects in terms of facial expressions, illumination, partial obstacles, rotations, makeups, etc., lead facial recognition to be a challenging topic in several fields although there are many samples for each subject. Since the human faces are not solid bodies, there is a lot of unnecessary information that is not essential for recognition. Feature extraction, in which the data compression is employed, is considered to be an important step in every recognition system. Since the input facial image has large dimensions, which have lot of useless information, the recognition rate decreases and the computational complexity increases.

Therefore, the goal of this step is to find the most pertinent features from the original images to best represent the faces and achieve high accuracy while fulfilling the low dimensionality and less storage requirements. Hence the following efficient tools are applied to get better facial representation:

1. The 2D DMWT based on MRA is used for:
  - (a) Dimensionality reduction, Less storage requirement.
  - (b) Noise alleviation.



**Fig. 10** An example of applying 2-D DMWT to ORL and FEI databases, **a** ORL, **b** FEI

(c) Feature selection by localizing most of the energy in a single band.

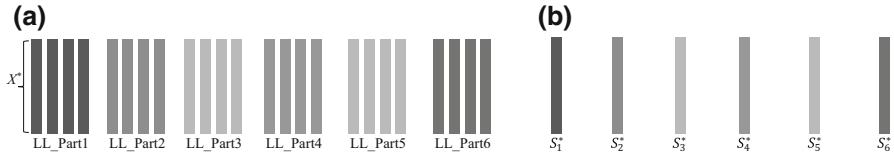
2 The FastICA is used for:

- (a) Decorrelating the high-order statistics beside decorrelating the second-order moments. Most of the information about the local characteristics of the facial images is contained in the high-order statistics. Hence, ICA basis is used for better representing the facial images [8, 18, 41, 59].
- (b) Improving the convergence rate and reducing the computational complexity [8, 27].

This leads to the resultant ICA features are:

- (a) Less sensitive to the facial variations arising from different facial expressions, illuminations, rotations, different poses, etc. [8, 18, 59].
- (b) Independent. The ICA architecture-I produces spatially localized representation compared with PCA and ICA architecture-II that produce global representations, which are sensitive to any distortion in the faces. These traits of ICA architecture-I enhance the system performance [8, 18, 58].

The first step in the feature extraction is applying 2D DMWT to each partition of each pose in the databases. Samples of 2D DMWT results are shown in Fig. 10. As shown in Fig. 10, each partition is divided into four main sub-bands. Hence, each sub-band has  $32 \times 32$  dimensions for ORL database,  $64 \times 64$  dimensions for both YALE and FERET databases,  $128 \times 128$  dimensions for FEI database, while each sub-band in LFW database has  $16 \times 16$  dimensions. Moreover, each sub-band is further divided into four sub-images each with  $16 \times 16$  dimensions for ORL database,  $32 \times 32$  dimensions for both YALE and FERET databases,  $64 \times 64$  dimensions for FEI database, and  $8 \times 8$  dimensions for LFW database. From Fig. 10a, b, most of the relevant information is localized in the upper left sub-band, which is low–low (LL) frequency sub-band. Therefore, the LL sub-band is maintained and all the other sub-bands are eliminated. Hence, the resultant feature tensor for each partition has dimensions of  $16 \times 16 \times 4$ ,  $32 \times 32 \times 4$ ,  $64 \times 64 \times 4$ , and  $8 \times 8 \times 4$  for ORL, YALE and FERET, FEI, and LFW databases, respectively. The following procedures are applied to the retained feature matrices to obtain better facial representations:



**Fig. 11** Two representations.  $S_1^* \dots S_6^*$  are denoted as sub-images with maximum  $\ell_2$ -norm. **a** First representation. **b** Second representation

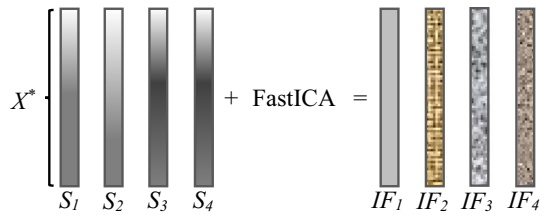
- (a) Convert each sub-image in the retained LL sub-band of each partition to 1-D vector of  $256 \times 1$ ,  $1024 \times 1$ ,  $4096 \times 1$ ,  $64 \times 1$  for ORL, YALE and FERET, FEI, and LFW databases, respectively.
- (b) Repeat [1] to all sub-images in each partition of each pose in the databases. There are two representations as shown in Fig. 11:
  - (A) The first representation shown in Fig. 11a is using all four sub-images of LL sub-band with  $X^* \times 4 \times 6$  dimensions for each pose. Where  $X^*$  is varied according to which database is involved.
  - (B) The second representation, Fig. 11b, uses only one sub-image that has maximum  $\ell_2$ -norm. Hence, each pose has  $X^* \times 6$  dimensions.

Producing new features from the original one is a very beneficial step in order to reduce the dimensionality and achieve better facial representation. In order to achieve that, the following techniques are applied to the resultant matrices corresponding to the first and second representations:

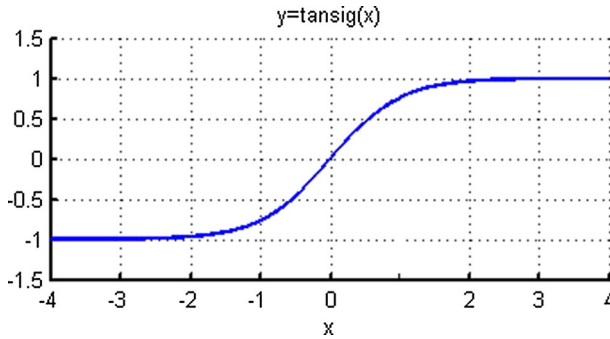
- (A) First representation shown in Fig. 11a.
  - *Technique 1*
    - (i) Apply 2D FastICA to each resultant LL sub-band matrix of each part that has  $X^* \times 4$  dimensions, see Fig. 11a.
    - (ii) Find  $\ell_2$ -norm for each row of the resultant ICA matrix, independent features  $IF_l$  as shown in Fig. 12, to reduce the dimensionality of the data and constrain all the energy in a single column. Hence, the resultant features for each pose of each person has  $X^* \times 6$  dimensions.
  - *Technique 2* Is the same as *Tech.1* except that we used Mixing matrix  $A$  expressed in Eq. (11) instead of using *resultant Independent features* as an alternative way of facial representation [8,27,58].
  - *Technique 3*
    - (i) Apply 2D FastICA to each LL sub-band matrix of each part of each pose as in *Tech.1-i*.
    - (ii) Choose the  $IF_l$  column, see Fig. 12, that has maximum  $\ell_2$ -norm for better feature representation and further dimensionality reduction. The dimensions of the resulting matrix features of each pose are the same as dimensions expressed in *Tech.1*.
  - *Technique 4*
    - (i) Apply 2D FastICA to the combined feature matrix of all partitions of LL sub-band, Fig. 11a, which has  $X^* \times 24$  dimensions.



**Fig. 12** Applying FastICA to retained LL sub-band.  $S_l$  represents sub-images of LL sub-band.  $IF_l$  represents independent features. Where  $l = 1, 2, 3,$  and  $4$



- (ii) Find  $\ell_2$ -norm for each row of the resultant ICA pose matrix for further dimensionality reduction/feature compaction. Here, the resultant features matrix for each pose has of  $X^* \times 1$  dimensions.
- *Technique 5* It is the same as *Tech.4* except using Mixing matrix  $A$  presented in Eq. (11) instead of using *resultant independent features*. As we mentioned, it is considered as an alternative way of representing the input signal [8,27,58].
- *Technique 6*
  - (i) Apply 2D FastICA to the combined feature matrix as performed in *Tech.4*.
  - (ii) Choose the  $IF_l$  column, Fig. 12, that has maximum  $\ell_2$ -norm for better facial representation and further data compaction. Thus, the resultant features for each pose have  $X^* \times 1$  dimensions.
- (B) Second representation shown in Fig. 11b.
  - *Technique 1*
    - (i) Apply 2D FastICA to the matrix resulted from second representation shown in Fig. 11b that has  $X^* \times 6$  dimensions.
    - (ii) Find  $\ell_2$ -norm for each row of the resultant features to reduce feature dimensions and concentrate the energy in one column. The resultant features for each pose have  $X^* \times 1$  dimensions.
  - *Technique 2*
    - (i) The mixing matrix  $A$ , used in Eq. (11), resulted from *Tech.1* is used here instead of FastICA signal matrix.
    - (ii) Apply *Tech.1-ii* to mixing matrix. The dimensions of the resultant feature matrix of each pose are of  $X^* \times 1$ .
  - *Technique 3*
    - (i) Apply the first step of *Tech.1*.
    - (ii) The  $IF_l$  column, shown in Fig. 12, that has maximum  $\ell_2$ -norm is selected. The dimensions of the output features resulted are the same as dimensions of *Tech.1* and *Tech.2*.
- (c) Repeat step [1–2] for all poses and for all persons in the databases.
- (d) The final compacted independent feature, which has dimensions varied from *Technique* to another according to which database gets involved, is fed to the recognition step for classification, which is based on BPTA.



**Fig. 13** Hyperbolic tangent sigmoid transfer function. The activation function is expressed as  $\text{tansig}(x) = \frac{2}{1 + e^{-2x}} - 1$  and considered as  $\text{tanh}(x)$ . Its output range is between  $(-1, 1)$

### 3.3 Recognition

In this phase, the unknown image will be compared with the training images (stored database). Therefore, building a good classifier is a very important step in order to anticipate a good performance and accuracy. Thus, the recognition phase consists of two important modes: Training and Testing Mode.

A NN is a very powerful tool used to classify the input signals. The NN based on BPTA is used in this paper in both training and testing modes. The BPTA is a supervised learning algorithm; thus, it is necessary to choose the desired output for each database used in this paper. Each desired output for each person should be different from other desired outputs of other persons. However, for the same person the desired output should be the same for all poses used in the training mode and for all partitions. For example, the desired output for person one of YALE database is written as:  $[1-1-1-1-1-1-1-1-1-1-1-1-1-1-1]$ . Where number (1) represents the output (1) is active and number  $(-1)$  represents the output is inactive. Note that the output might not be equal to 1 or  $-1$  but may reach to these values since the activation function used in the BPTA is hyperbolic tangent sigmoid transfer function shown in Fig. 13 [24]. The configuration of the NN has three layers, namely input, hidden, and output layers. The number of the neurons in the input and the output layers is always fixed according to the dimensions of the input vectors and the number of the persons in the database, while the choice of the number of hidden neurons is somehow flexible.

In the training mode, the NN is configured and the classifier is built using the described training features. Each technique has a classifier different from other techniques since each one has its own resultant features and then each one will have its own results as shown in the experimental results section.

In the testing mode, the same procedures of the training mode are employed. First, each test pose is passed through preprocessing steps. Second, 2D DMWT is applied to each partition. Then only LL sub-band is retained and the remaining sub-bands are eliminated. Next, convert all sub-bands to 1-D form. After that all discussed techniques of the two representations are applied to the resultant features. Finally, the resulting matrix features that have dimensions vary from technique to another are fed to NN for classification.

### 3.4 Decision

In this section, we will explain how the decision is made. The following example is presented according to the first representation. Let's assume we have an unknown pose (tested pose) corresponding to the person 1. Since we have 6 different features corresponding to 6 different partitions resulted from Tech.1–Tech.3, each one is tested individually with the training features and assigned to the specific person. Let's assume that the outputs from the NN for these six partitions are (111511), which mean the following:

- (1): The tested features of these partitions are matched with the corresponding trained features and assigned to **person 1, which is correct**.
- (5): The tested features of the fourth partition are matched with one of the trained partitions of person 5 and **person 5 is assigned, which is incorrect**.

In general, if there are more than three tested partitions matched with the corresponding trained features of the same person, the testing pose is correctly matched. In the above example, the tested pose is correctly matched with **person 1**. If there is no majority matched, the testing pose is incorrectly matched. Also, if the partition 1–partition 4 are assigned to the same person, the program will stop and the decision is made since the majority is achieved. This leads to less computational complexity.

## 4 Experimental Results

The experimental results of the proposed techniques based on two representations are presented and compared with a large number of the state-of-the-art methods. Five databases, namely ORL, YALE, FERET, FEI, and LFW containing different facial variations, such as facial expressions, light conditions, rotations, are used to test the proposed techniques. *K*-fold cross-validation (CV) is employed to analyze the experimental results. Various values of *K* are chosen,  $K = 2$ ,  $K = 3$ , and  $K = 5$ . The recognition rates reported in Tables 2, 3, 4, 5 and 6 are the average of the rates obtained across the *K*-folds CV. The NN simulations of the proposed techniques are presented and compared with the existing approaches. The proposed techniques achieved 100% recognition rates when tested with the training poses.

### 4.1 Experimental Results for the ORL Database

The ORL database consists of 10 different poses of each of 40 different persons with the resolution of  $112 \times 92$  pixels in the BMP format. The poses of all subjects have different facial variations, such as facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses/ no glasses). The images were taken against a dark homogeneous background with the persons in an upright, frontal position [14]. Table 2 summarizes the results of the proposed techniques based on two representations compared with the state-of-the-art approaches.

**Table 2** Experimental results for the ORL database

Proposed techniques		Other methods	
First representation		PCA–CDA [47]	95%
<i>Tech.1</i>	100%	RV–LDA [29]	97%
<i>Tech.2</i>	99%	OLPP [51]	93.5%
<i>Tech.3</i>	98.5%	SNMFSPM [40]	98.15%
<i>Tech.4</i>	99.25%	DWT–PCA [44]	96.75%
<i>Tech.5</i>	98.75%	DWT–ICA [41]	97.5%
<i>Tech.6</i>	98.25%	LBPP–DCT [13]	95.5%
Second representation		GWT–DCT [2]	97.25%
<i>Tech.1</i>	99%	DCT–PCA [10]	96%
<i>Tech.2</i>	98.75%	SADL [61]	97.5%
<i>Tech.3</i>	98.25%	DWT–MPSO [33]	98.33%

**Table 3** Experimental results for the YALE database

Proposed techniques		Other methods	
First representation		Fast $\ell_1$ -PCA [31]	83%
<i>Tech.1</i>	100%	OLPP [51]	98.2%
<i>Tech.2</i>	99.17%	LPP [7]	97.14%
<i>Tech.3</i>	98.47%	SNMFSPM [40]	97.69%
<i>Tech.4</i>	99.58%	DWT–PCA [44]	96.55%
<i>Tech.5</i>	98.89%	DWT–ICA [41]	97.25%
<i>Tech.6</i>	98.1%	GWT–DCT [2]	96.95%
Second representation		DCT–PCA [10]	96.34%
<i>Tech.1</i>	99.17%	SADL [61]	94.67%
<i>Tech.2</i>	98.75%	DWT–MPSO [33]	98.29%
<i>Tech.3</i>	97.18%	NFLS-II [46]	82.42%

The proposed techniques improved the recognition accuracy compared with other methods. Also, the proposed techniques achieved higher recognition rates when compared with the results reported in [3–5, 7].

## 4.2 Experimental Results for the YALE Database

There are 15 persons in the YALE database, each with 11 different poses with resolution of  $320 \times 243$  pixels in GIF format. All of them have different facial configurations, such as center-light, glasses, happy, left-light, no glasses, normal, right-light, sad, sleepy, surprised, and wink [15]. Table 3 shows the results of the proposed techniques based two representations and the comparison with the other methods. As before, the results achieved using the proposed techniques were better than the results reported in [3–5].

**Table 4** Experimental results for the FERET database

Proposed techniques		Other methods	
First representation		OLPP [51]	92.1%
<i>Tech.1</i>	99.6%	DWT-PCA [44]	96.4%
<i>Tech.2</i>	98.86%	DWT-ICA [41]	97.33%
<i>Tech.3</i>	98.3%	GWT-DCT [2]	96.8%
<i>Tech.4</i>	99.2%	DCT-PCA [10]	95.7%
<i>Tech.5</i>	98.66%	SADL [61]	94.33%
<i>Tech.6</i>	98.21%	LFLM-SIFT [19]	92.4%
Second representation			
<i>Tech.1</i>	99.05%		
<i>Tech.2</i>	98.4%		
<i>Tech.3</i>	98%		

**Table 5** Experimental results for the FEI database

Proposed techniques		Other methods	
First representation		OLPP [51]	98.9%
<i>Tech.1</i>	99.61%	DWT-PCA [44]	96.25%
<i>Tech.2</i>	99%	DWT-ICA [41]	97.21%
<i>Tech.3</i>	98.5%	GWT-DCT [2]	96.65%
<i>Tech.4</i>	99.4%	DCT-PCA [10]	95.85%
<i>Tech.5</i>	98.83%	VQ-KFCG [45]	93.81%
<i>Tech.6</i>	98.23%	LFLM-SIFT [19]	85.3%
Second representation		DWT-MPSO [33]	97.08%
<i>Tech.1</i>	99.2%	NFLS-II [46]	93%
<i>Tech.2</i>	98.6%	PCNC [35]	94.17%
<i>Tech.3</i>	98.2%		

### 4.3 Experimental Results for the FERET Database

The FERET database used in this paper contains 200 individual, each with 11 different poses with resolution  $256 \times 384$  pixels in TIFF format. The poses in this database have different facial variations, such as light conditions, rotations, background, facial expressions, and glasses/no glasses [48,49]. The results of both representations are summarized in Table 4.

The results exhibit the same behavior as in the previous databases when compared with the existing approaches. Also, the proposed techniques accomplished higher accuracy compared to those reported in [3–5,45].

### 4.4 Experimental Results for the FET Database

The FEI database contains Brazilian faces recorded at the artificial intelligence laboratory of FEI in São Bernardo do Campo, São Paulo, Brazil. The FEI consists of 200 individuals each with 14 different poses with the resolution of  $640 \times 480$  pixels

**Table 6** Experimental results for the LFW database

Proposed techniques		Other methods	
First representation		DWT-PCA [44]	94.8%
<i>Tech.1</i>	98.91%	DWT-ICA [41]	95.72%
<i>Tech.2</i>	98.06%	GWT-DCT [2]	95.3%
<i>Tech.3</i>	97.33%	DCT-PCA [10]	94.5%
<i>Tech.4</i>	98.54%	CNN-SAE [17]	98.24%
<i>Tech.5</i>	97.87%	LDA-SID [37]	95.65%
<i>Tech.6</i>	97.11%	GWT-PCA [16]	96%
Second representation		HPEN-HD-Gabor-JB [63]	95.25%
<i>Tech.1</i>	98.31%		
<i>Tech.2</i>	97.65%		
<i>Tech.3</i>	96.8%		

in the JPG format. This is a colored database with facial images rotated up to 180 degrees, distinct appearance, hairstyle, and adornments [55]. The gray scale version of this database is used in this paper. The results, which are presented in Table 5, of the proposed techniques based two representations demonstrate the same attitudes in comparison with the state-of-the-arts approaches.

#### 4.5 Experimental Results for the LFW Database

Labeled faces in the wild (LFW), a large database of photographs designed for use in unconstrained face recognition, contains 13,233 images of 5749 persons. 1680 people have two or more image, while the remaining 4069 people have just a single image in the database. Each pose is labeled with the name of the person. These facial images were collected from the Internet; therefore, all poses of all persons have different facial variations, such as pose, occlusion, illumination, facial expression variations, etc. This is a color database with the size of  $250 \times 250$  pixels in the JPG format [25]. The cropped and gray scale version of this database is used in this paper. Table 6 summarizes the results of the proposed techniques compared with the state-of-the-arts methods.

## 5 Discussion

As we mentioned the purpose of applying partitioning step in the preprocessing phase was to reduce the effect of unnecessary information, which is common among all poses and persons and to highlight the local features. Also, partitioning helps to improve the recognition accuracy since each partition contained local facial features that are in common with the other two partitions. This led to increase the matching probability for each pose and hence the overall recognition accuracy improved.

In contrast to the DWT, the DMWT has several favorable features through achieving a good reconstruction (orthogonality), better performance (linear phase symmetry), high order of approximation (vanishing moments), and compact support. These desirable features cannot be achieved at the same time in the scalar wavelet, while in MWT provide more degree of freedom and give perfect performance in signal and image applications [53]. Based on a number of decomposition levels required and the resulted sub-bands that mathematically expressed as  $3 \times L^* + 1$  for DWT and  $4 + 12 \times L^*$  for DMWT, DMWT achieves high dimensionality reduction than DWT. Where  $L^*$  is the number of decomposition levels required. Therefore, the proposed techniques accomplished less storage compared to those reported in [41, 44] that used DWT.

Also, the results of the proposed techniques based two representations are higher than the results reported in [19] even though they partitioned the input facial images to  $n$  overlapping partitions. These achievements are due to applying efficient integrated tools (2D DMWT and 2D FastICA) to the partitioned images. The dimensionality reduction and efficient feature extraction were obtained by retaining only the LL sub-band of 2D DMWT. Furthermore, producing independent and more discriminating features were achieved through applying 2D FastICA to the resultant of DMWT representations.

As explained, FastICA has several advantages compared with ICA, such as FastICA has faster convergence than traditional ICA since it does not require to select a step size compared with ICA-based gradient algorithm. Also, the independent non-Gaussian signal using FastICA is found by any arbitrary non-linear function  $\theta(t)$ , while other ICA algorithms required to evaluate the PDF of the selected non-linear function [27, 58]. Therefore, FastICA is more efficient of estimating the statistical components. Hence, the key of using FastICA in this paper is to find the bases images that are statistically independent. These bases are maintaining the structural information of the facial images. The ICA representations are more robust to the different variations caused by different facial expressions, pose, and illuminations, which can be considered as forms of noises with respect to the original images [8, 18, 59]. Therefore, combining these favorable properties of DMWT and the efficient representations of FastICA led us to achieve a notable dimensionality reduction as well as an improvement in the recognition rates.

Each partition has possessed its own features (local features). Since each partition is in common with other two partitions, as seen in Fig. 9, the matching probability for each pose is significantly increased if features of each partition are participated in both training and testing phase. Since techniques 1–3 in the first representation are applied to DMWT features of each partition, there are six resultant feature representations for each pose. After testing these six features of each pose with the trained stored features of whole database and after applying voting technique, we can see that the experimental results of Techniques 1–3 outperformed other results of Techniques 4–6 and Techniques 1–3 in the first and second representations, respectively, which were employed to all partitions, whole image, at the same time.

In other words, using the features extracted from each partition (6 features for each pose) in both training and testing mode led the proposed Techniques 1–3 in the

**Table 7** The computation time per second

Methods		Proposed methods	DWT-PCA [44]	DWT-ICA [41]	GWT-DCT [2]	DCT-PCA [10]
Computation time	Training mode	366	326	348	357	336
	Testing mode	16	9.5	11.8	14.2	10.36

first representation to accomplish higher recognition rates compared to those obtained using one matrix (features) to represent the whole pose as in Techniques 4–6 and Techniques 1–3 in the first and second representations, respectively. Figure 9b shows how parts are participating with each other. Also, the results of the first representation outperformed the results of the second representation. This is due to the fact that images of faces are more efficiently described by the first representation as compared to the second representation.

Although the Techniques 4–6 and Techniques 1–3 in the first and the second representation, respectively, are applied to the all combined partitions (whole image), their results illustrate an improvement in the recognition rates when compared with the state-of-the-art approaches.

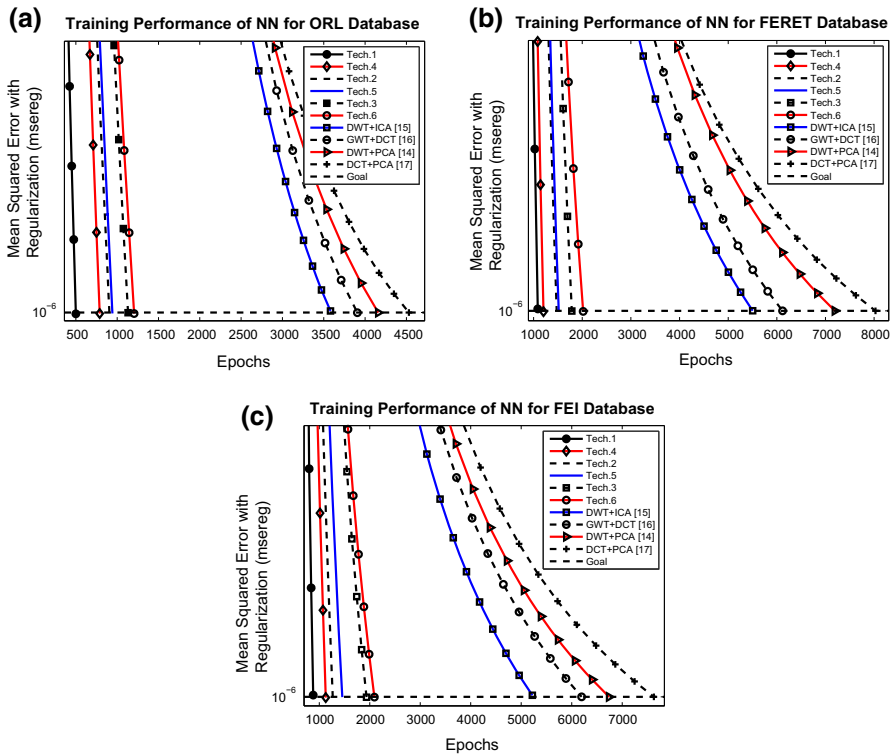
The method was implemented using a Laptop computer that has following properties: Model: Hp Envy m4 Notebook PC, Windows 10, CPU: Intel (R) Core (TM) i7-3632QM @ 2.2 GHz. The processor is a third generation, and RAM: 16 GB. The proposed methods and other approaches were performed using MATLAB 2016a. The computation time, for example of FERET database at  $K = 5$ , taken to perform the proposed method compared with the other approaches is shown in Table 7 below:

As shown in Table 1, the computation time taken by each technique of our proposed system in the training mode was 366 s on average. Hence, each pose in the FERET database required a time of  $366/1600 = 0.228$  s. While in the testing mode, each pose required a time of  $16/600 = 26.67$  ms. Our proposed methods used, on average, comparable times compared with the time taken by other approaches. In Table 7, we did not report the time taken by the NNs in the training mode since NNs are affected by several factors, as explained in Sect. 6.

## 6 Performance of Neural Network

There are several factors that affect the performance of NN in the training mode and lead to impact the overall system accuracy. The first factor is the *Network complexity*, which depends on the number of hidden layers, the number of neurons in each hidden layer, and the type of activation functions used for each interconnection weights. The second one is *Learning complexity*, which depends on choosing training algorithms, initialization parameters, and weight selection. Finally, the *Problem complexity* that depends on how accurate, efficient, and sufficient the databases used in the training mode.

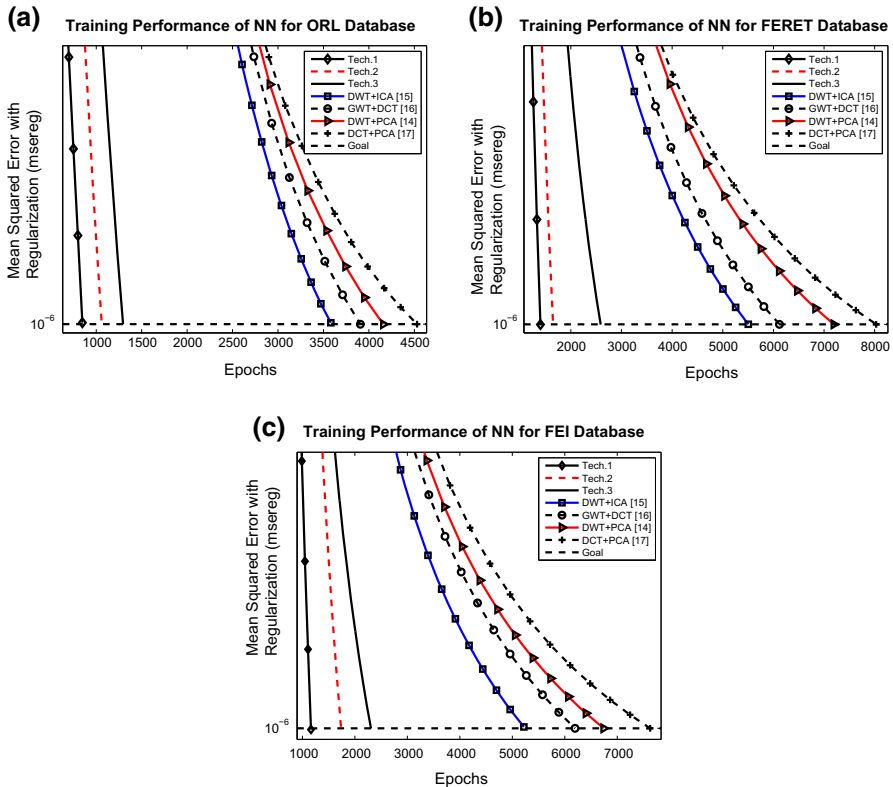




**Fig. 14** NN performance for the proposed techniques compared with the other approaches based first representation. **a** ORL. **b** FERET. **c** FEI

In this paper, we used one hidden layer for all five databases. Hence, 256, 1024, 1024, 4069, and 64 hidden neurons were selected for ORL, YALE, FERET, FEI, and LFW databases, respectively. The activation function used is hyperbolic tangent sigmoid shown in Fig. 13. The databases are trained and tested using BPTA. Mean square error with regularization (msereg) is used to calculate the error. Msereg is finding the mean sum of the square error between the actual output and the target. Figure 14 shows samples of the performance of the NN in the training mode of all discussed techniques based on the first representation in comparison with some of the state-of-the-arts approaches [2, 10, 41, 44].

It is obvious from Fig. 14 that the proposed Techniques 1–3 outperformed and achieved at the goal faster than Techniques 4–6. In addition, the performance of first representation is achieved the goal faster than the second representation that shown in Fig. 15. In addition, all proposed techniques based on the two representations outperformed the existing approaches, such as [2, 10, 41, 44]. This is due to the fact that partitioning step and employing the two efficient tools (DMWT and FastICA) in the feature extraction steps produced features more efficient than those obtained by the existing methods. Hence, the NN convergence is enhanced and the accuracies are improved.



**Fig. 15** NN performance for the proposed techniques compared with the other approaches based second representation. **a** ORL. **b** FERET. **c** FEI

## 7 Conclusion

A new approach applying 2D DMWT followed by FastICA to partitioned facial images for face recognition was proposed in this paper. Dimensionality reduction, efficient feature extraction, and as a consequence, high recognition rates were accomplished in this contribution. The disadvantages and shortcomings of DWT and ICA and other methods have been taken into account while considering the combination of DMWT and FastICA. The facial images were Partitioned into six parts in order to reduce the effect of the mutual information found among all persons in the databases on the overall system performance, and to highlight the local features. 2D DMWT was applied to each part, and only the LL sub-band is retained and all other sub-bands were eliminated. As a consequence, the dimensionality reduction (less storage requirements) was accomplished. Two representations were constructed using DMWT features. Splitting the feature extraction into the two representation methods and the use of FastICA followed by the  $\ell_2$ -norm help to retrieve the useful information from the images and discard the non-discriminating features.

The dimensionality reduction proposed in the system helps in reducing the computational burden and improving the accuracy. The combination of the feature extraction methods (2D DMWT and FastICA followed by  $\ell_2$ -norm) led to more discriminating features being extracted and further led to faster convergence. The final features extracted using the proposed techniques that fed to NN for training/testing have dimensions of  $X^* \times Y^*$ , where  $X^*$  is 256, 1024, 4096, and 64 dimensions for ORL, YALE and FERET, FEI, and LFW databases, respectively.  $Y^*$  is either 6 or 1 according to which technique gets involved and which representation was employed. The proposed system was extensively evaluated using five databases that have different facial variations, such as illuminations, facial expressions, rotations, facial details. The experimental results were analyzed using different values of  $K$ -fold CV. The proposed techniques based on the two representations were not only achieving high recognition rates compared to the other methods, shown in Tables 2, 3, 4, 5 and 6, but also having faster NN convergence than the other approaches, as shown in Figs. 14 and 15.

**Acknowledgements** This work was supported by the Iraqi government scholarship (HCED). The authors acknowledge the valuable comments and feedback from their colleague “Dr. George Atia”.

## References

1. T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
2. S. Ajitha, A.A. Fathima, V. Vaidehi, M. Hemalatha, R. Karthigaiveni, Face recognition system using combined gabor wavelet and DCT approach, in *IEEE International Conference on Recent Trends in Information Technology (ICRITIT)* (2014), pp. 1–6
3. A. Aldhabab, G. Atia, W.B. Mikhael, Supervised facial recognition based on multi-resolution analysis and feature alignment, in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)* (2014), pp. 137–140
4. A. Aldhabab, G. Atia, W.B. Mikhael, High performance and efficient facial recognition using norm of ICA/multiwavelet features, in *Proceedings of the Advances in Visual Computing: 11th International Symposium, Part II, ISVC 2015, Las Vegas, NV, USA, 14–16, December 2015*, ed. By G. Bebis, R. Boyle, B. Parvin, D. Koracin, I. Pavlidis, R. Feris, T. McGraw, M. Elenndt, R. Kopper, E. Ragan, Z. Ye, G. Weber (Springer International Publishing, Cham, 2015), pp. 672–681
5. A. Aldhabab, G. Atia, W.B. Mikhael, Supervised facial recognition based on eigenanalysis of multi-resolution and independent features, in *IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)* (IEEE, 2015), pp. 1–4
6. S. Arlot, A. Celisse et al., A survey of cross-validation procedures for model selection. *Stat. Surv.* **4**, 40–79 (2010)
7. M.E. Ashalatha, M.S. Holi, P.R. Mirajkar, Face recognition using local features by LPP approach, in *IEEE International Conference on Circuits, Communication, Control and Computing*, vol. I4C (2014), pp. 382–386
8. M. Bartlett, J.R. Movellan, T. Sejnowski, Face recognition by independent component analysis. *IEEE Trans. Neural Netw.* **13**(6), 1450–1464 (2002)
9. I.J. Brown, A wavelet tour of signal processing: the sparse way. *Investig. Oper.* **1**, 85 (2009)
10. F.Z. Chelali, A. Djeradi, N. Cherabit, Investigation of DCT/PCA combined with Kohonen classifier for human identification, in *IEEE 4th International Conference on Electrical Engineering (ICEE)* (2015), pp. 1–7
11. K.W. Cheung, L.M. Po, Preprocessing for discrete multiwavelet transform of two-dimensional signals, in *Proceedings IEEE International Conference on Image Processing*, vol. 2 (IEEE, 1997), pp. 350–353
12. Y.T. Chou, J.F. Yang, Intra-facial-feature canonical correlation analysis for face recognition, in *TENCON 2015–2015 IEEE Region 10 Conference* (2015), pp. 1–4. doi:[10.1109/TENCON.2015.7372961](https://doi.org/10.1109/TENCON.2015.7372961)

13. A. Dahmouni, N. Aharrane, K.E. Moutaouakil, K. Satori, Face recognition using local binary probabilistic pattern (LBPP) and 2D-DCT frequency decomposition, in *IEEE 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)* (2016), pp. 73–77
14. O. Database, At&t laboratories Cambridge database of faces (April 1992–1994). <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
15. Y. Database, Ucsd computer vision. <http://vision.ucsd.edu/content/yale-face-database>
16. B. Dhivakar, C. Sridevi, S. Selvakumar, P. Guhan, Face detection and recognition using skin color, in *IEEE 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)* (2015), pp. 1–7
17. C. Ding, D. Tao, Robust face recognition via multimodal deep face representation. *IEEE Trans. Multimed.* **17**(11), 2049–2058 (2015)
18. B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, Recognizing faces with PCA and ICA. *Comput. Vis. Image Underst.* **91**(1), 115–137 (2003)
19. X. Duan, Z.H. Tan, Local feature learning for face recognition under varying poses, in *IEEE International Conference on Image Processing (ICIP)* (2015), pp. 2905–2909
20. S. Farokhi, U.U. Sheikh, J. Flusser, B. Yang, Near infrared face recognition using Zernike moments and Hermite kernels. *Inf. Sci.* **316**, 234–245 (2015). doi:10.1016/j.ins.2015.04.030. Nature-Inspired Algorithms for Large Scale Global Optimization
21. J.S. Geronimo, D.P. Hardin, P.R. Massopust, Fractal functions and wavelet expansions based on several scaling functions. *J. Approx. Theory* **78**(3), 373–401 (1994)
22. M. Girolami, C. Fyfe, Blind separation of sources using exploratory projection pursuit networks, in *International Conference on the Speech and Signal Processing Engineering Applications of Neural Networks* (1996), pp. 249–252
23. R.C. Gonzalez, R.E. Woods, *Digital Image Processing* (Prentice Hall, Upper Saddle River, 2002)
24. P.D.B. Harrington, Sigmoid transfer functions in backpropagation neural networks. *Anal. Chem.* **65**(15), 2167–2168 (1993)
25. G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. Rep. 07-49 (University of Massachusetts, Amherst, 2007)
26. A. Hyvarinen, New approximations of differential entropy for independent component analysis and projection pursuit. *Adv. Neural Inf. Process. Syst.* **10**(2), 273–279 (1998)
27. A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **10**(3), 626–634 (1999)
28. A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis. *Neural Comput.* **9**(7), 1483–1492 (1997)
29. A. Iosifidis, A. Tefas, I. Pitas, On the optimal class representation in linear discriminant analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **24**(9), 1491–1497 (2013)
30. R. Jafri, H.R. Arabnia, A survey of face recognition techniques. *J. Inf. Process. Syst.* **5**(2), 41–68 (2009)
31. M. Johnson, A. Savakis, Fast 11-eigenfaces for robust face recognition, in *IEEE Western New York Image and Signal Processing Workshop (WNYISPW)* (2014), pp. 1–5
32. J. Karhunen, P. Pajunen, E. Oja, The nonlinear PCA criterion in blind source separation: relations with other approaches. *Neurocomputing* **22**(1), 5–20 (1998)
33. T. Khadhraoui, S. Ktata, F. Benzarti, H. Amiri, Features selection based on modified PSO algorithm for 2D face recognition, in *IEEE 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)* (2016), pp. 99–104
34. S. Kundu, P.P. Markopoulos, D.A. Pados, Fast computation of the 11-principal component of real-valued data, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2014), pp. 8028–8032
35. E. Kussul, T. Baydyk, Face recognition using special neural networks, in *IEEE International Joint Conference on Neural Networks (IJCNN)* (2015), pp. 1–7
36. D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
37. Z. Lei, D. Yi, S.Z. Li, Learning stacked image descriptor for face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **PP**(99), 1 (2015)
38. D. Li, H. Zhou, K.M. Lam, High-resolution face verification using pore-scale facial features. *IEEE Trans. Image Process.* **24**(8), 2317–2327 (2015)

39. Y. Linde, A. Buzo, R. Gray, An algorithm for vector quantizer design. *IEEE Trans. Commun.* **28**(1), 84–95 (1980)
40. X. Long, H. Lu, Y. Peng, Sparse non-negative matrix factorization based on spatial pyramid matching for face recognition, in *IEEE 5th International Conference on Intelligent Human–Machine Systems and Cybernetics (IHMSC)*, vol. 1 (2013), pp. 82–85
41. M. Luo, L. Song, S.D. Li, An improved face recognition based on ICA and WT, in *IEEE Asia-Pacific Services Computing Conference (APSCC)* (2012), pp. 315–318
42. S.M. Mahbubur Rahman, S.P. Lata, T. Howlader, Bayesian face recognition using 2D Gaussian–Hermite moments. *EURASIP J. Image Video Process.* **2015**(1), 35 (2015). doi:[10.1186/s13640-015-0090-5](https://doi.org/10.1186/s13640-015-0090-5)
43. S.S. Mudholkar, P.M. Shende, M. Sarode, Biometrics authentication technique for intrusion detection system using fingerprint recognition. *Int. J. Comput. Sci. Eng. Inf. Technol.* **2**(1), 57–65 (2012)
44. M.M. Mukhedkar, S.B. Powalkar, Fast face recognition based on wavelet transform on PCA, in *IEEE International Conference on Energy Systems and Applications* (2015), pp. 761–764
45. S.J. Natu, P.J. Natu, T.K. Sarode, H. Kekre, Performance comparison of face recognition using DCT against face recognition using vector quantization algorithms LBG, KPE, KMCG, KFCG. *Int. J. Image Process.* (IJIP) **4**(4), 377–389 (2010)
46. J.S. Pan, Q. Feng, L. Yan, J.F. Yang, Neighborhood feature line segment for image classifications. *IEEE Trans. Circuits Syst. Video Technol.* **25**(3), 387–398 (2015)
47. K. Papachristou, A. Tefas, I. Pitas, Symmetric subspace learning for image analysis. *IEEE Trans. Image Process.* **23**(12), 5683–5697 (2014)
48. P.J. Phillips, H. Moon, S. Rizvi, P.J. Rauss et al., The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
49. P.J. Phillips, H. Wechsler, J. Huang, P.J. Rauss, The feret database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **16**(5), 295–306 (1998)
50. S.M. Rahman, T. Howlader, D. Hatzinakos, On the selection of 2D krawtchouk moments for face recognition. *Pattern Recogn.* **54**, 83–93 (2016). doi:[10.1016/j.patcog.2016.01.003](https://doi.org/10.1016/j.patcog.2016.01.003)
51. J. Soldera, C.A.R. Behaine, J. Scharcanski, Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. *IEEE Trans. Instrum. Meas.* **64**(9), 2417–2426 (2015)
52. V. Strela, *Multiwavelets: theory and applications*. Ph.D. Thesis (Citeseer, 1996)
53. V. Strela, P. Heller, G. Strang, P. Topiwala, C. Heil, The application of multiwavelet filterbanks to image processing. *IEEE Trans. Image Process.* **8**(4), 548–563 (1999)
54. V. Strela, A.T. Walden, Orthogonal and biorthogonal multiwavelets for signal denoising and image compression, in *Aerospace/Defense Sensing and Controls* (International Society for Optics and Photonics, 1998), pp. 96–107
55. C.E. Thomaz, G.A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis. *Image Vis. Comput.* **28**(6), 902–913 (2010). <http://fei.edu.br/~cet/facedatabase.html>
56. M. Turk, A. Pentland, Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
57. L. Wiskott, J.M. Fellous, N. Kuiger, C. von der Malsburg, Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997)
58. X. Yu, D. Hu, J. Xu, *Blind Source Separation: Theory and Applications* (Wiley, Singapore, 2014)
59. P.C. Yuen, J.H. Lai, Face representation using independent component analysis. *Pattern Recogn.* **35**(6), 1247–1257 (2002)
60. J.J. Zhang, Y.T. Shi, Face recognition systems based on independent component analysis and support vector machine, in *IEEE International Conference on Audio, Language and Image Processing (ICALIP)* (2014), pp. 296–300
61. Z. Zhang, J. Li, R. Zhu, Deep neural network for face recognition based on sparse autoencoder, in *IEEE 8th International Congress on Image and Signal Processing (CISP)* (2015), pp. 594–598
62. X. Zhihua, L. Guodong, Weighted infrared face recognition in multiwavelet domain, in *IEEE International Conference on Imaging Systems and Techniques (IST)* (2013), pp. 70–74
63. X. Zhu, Z. Lei, J. Yan, D. Yi, S.Z. Li, High-fidelity pose and expression normalization for face recognition in the wild, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 787–796