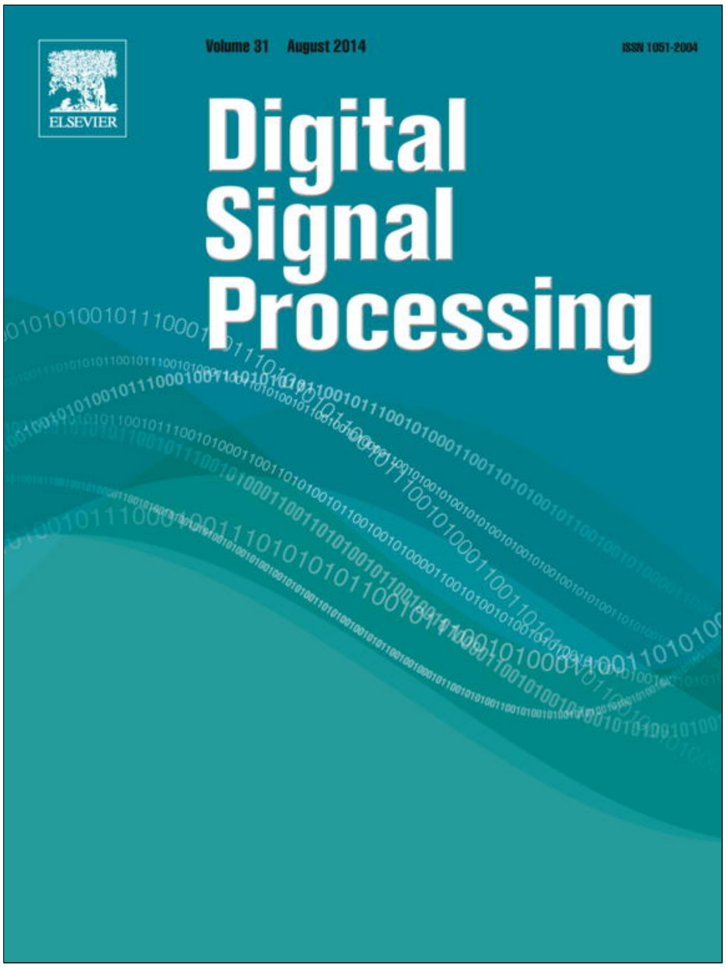
Volume 31 August 2014
ISSN 1051-2004
ELSEVIER
Digital Signal Processing

Digital Signal Processing 31 (2014) 13–27

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Near infrared face recognition by combining Zernike moments and undecimated discrete wavelet transform

Sajad Farokhi [a,*], Siti Mariyam Shamsuddin [a], U.U. Sheikh [b], Jan Flusser [c], Mohammad Khansari [d], Kourosh Jafari-Khouzani [e]

[a] UTM Big Data Centre, Faculty of Computing, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[b] Digital Signal and Image Processing Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Johor, Malaysia
[c] Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, 182 08, Prague, Czech Republic
[d] Faculty of New Sciences and Technologies, University of Tehran, 14399-55941, Tehran, Iran
[e] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 02114, Boston, MA, USA

ARTICLE INFO

Article history:
Available online 29 April 2014

Keywords:
Zernike moments
Undecimated discrete wavelet transform
Decision fusion
Near infrared
Face recognition

ABSTRACT

This study proposes a novel near infrared face recognition algorithm based on a combination of both local and global features. In this method local features are extracted from partitioned images by means of undecimated discrete wavelet transform (UDWT) and global features are extracted from the whole face image by means of Zernike moments (ZMs). Spectral regression discriminant analysis (SRDA) is then used to reduce the dimension of features. In order to make full use of global and local features and further improve the performance, a decision fusion technique is employed by using weighted sum rule. Experiments conducted on CASIA NIR database and PolyU-NIRFD database indicate that the proposed method has superior overall performance compared to some other methods in the presence of facial expressions, eyeglasses, head rotation, image noise and misalignments. Moreover its computational time is acceptable for on-line face recognition systems.

1. Introduction

Face recognition (FR) has received much attention over the past decades due to its potential value for many applications as well as wide challenges such as illumination variations, facial expression, head rotation, eyeglasses and misalignments which result in a significant decrease of the accuracy of the best known techniques [1,2]. On the other hand, collecting sufficient prototype images which can cover all challenges is practically impossible. Hence proposing an accurate face recognition system which is robust to most of the variations is still a challenging problem in the field of face recognition. Many face recognition techniques have been developed over the past few decades, some of which can be found in [3–5]. Most of them have focused on visible face recognition due to the fact that face recognition is one of the primary activities of the human visual system. The main limitation of visible face recognition, however, is the high dependency of system performance on external light, angle of light and even skin color [6,7]. Various methods have been developed and introduced to solve the illumination problem by proposing illumination invariant face recognition [8–10].

Recently, researchers have investigated the use of near infrared (NIR) imagery for face recognition with satisfactory results [11–13]. Three advantages of near infrared images in comparison with visible imagery can be expressed as follows. First, near infrared images are scarcely influenced by natural light. Hence it is possible to take images under very dark illumination whereas visible cameras have deficiency in this case [12,14]. Second, the eye is not sensitive to near infrared illumination and thus can be used in a more flexible and possibly covert manner [9]. Third, face detection, based on the position of the eyes, can be made more accurate than visible images due to "bright pupil" effect which simplifies eye localization and face detection consequently [9,15]. As a result, an automatic and accurate face recognition system based on near infrared spectrum can be implemented more easily than visible imagery.

Different face representations have been proposed in near infrared domain which can be roughly classified into two main categories: global-based and local-based. In the global-based face representation, each dimension of the feature vector corresponds

* Corresponding author.
E-mail addresses: fsajad2@live.utm.my (S. Farokhi), mariyam@live.utm.my (S.M. Shamsuddin), usman@fke.utm.my (U.U. Sheikh), flusser@utia.cas.cz (J. Flusser), m.khansari@ut.ac.ir (M. Khansari), kjafari@nmr.mgh.harvard.edu (K. Jafari-Khouzani).

http://dx.doi.org/10.1016/j.dsp.2014.04.008
1051-2004/

to some holistic characteristics in the face and hence encodes the global information embodied in every part of facial images. In contrast, in the local-based face representations each dimension of feature vector contains the detailed information corresponding to a certain local region of facial images [16].

An advanced design of the active NIR camera and progressive local-based method to propose illumination invariant face recognition was introduced by Li et al. [17]. In this method, first, face images are produced by active near infrared (NIR) imaging system regardless of visible light in the environment. Second, local binary pattern (LBP) features are used to compensate monotonic transform and propose illumination invariant face recognition. Finally, statistical learning algorithms are used to extract most discriminative features. The main weakness of this study, however, is the high sensitivity of LBP features to noise, head rotation and misalignments [18,19].

He et al. offered a global feature-based method based on discrete wavelet transform (DWT) and two dimensional principal component analysis (2DPCA) [20]. Coefficients of low frequency component which contribute to global information of images are used as features for the proposed method. Despite its advantages, there are several weaknesses in this approach as follows. First, the number of images in training is high and the accuracy of a system with a small number of images in the training set is not examined. Second, discrete wavelet transform is used in this paper which is not translation invariant. In other words, shift of the image by an odd number of pixels may change the whole coefficients of wavelet transform [21,22]. Hence, its accuracy will be highly affected when misalignments occur [23].

Zhang et al. introduced a novel local-based face recognition method, namely directional binary code (DBC) to capture the directional edge information of NIR facial images [24]. They showed that DBC achieves better performance than LBP. But they did not consider the effect of noise and misalignments which are the grand challenges in face recognition systems. Some other prior works can be found in [25–28].

The researches to date have tended to develop a face recognition system based on local features which are believed very robust to eyeglasses, pose variations and facial expressions. However, many studies have shown that both local and global features are crucial for highly accurate face recognition systems [16,23,29,30]. The underlying reason is that local and global features play different roles in face recognition scenarios. While local features contain more detailed local information of images and correspond to finer representation, global features are proper for coarse representation because they reflect the information of the whole face.

Inspired by the works presented in [23] which is based on combination of local and global features, and trying to avoid the translation sensitivity of DWT, in this paper, a novel face recognition method based on the integration of undecimated discrete wavelet transform (UDWT), Zernike moments (ZMs) and spectral regression discriminant analysis (SRDA) is proposed. We expect better performance by combining local and global features.

Though the basic idea of the proposed method is somewhat similar to other methods which are based on a combination of local and global features, this study has made the following transcendent contributions:

- Unlike previous studies which are based on a combination of Fourier, wavelet or Gabor wavelet transform, this paper exploits undecimated discrete wavelet transform (UDWT) as local features and Zernike moments as global features. Our experimental results confirm the effectiveness of the proposed method in the near infrared domain compared to similar transformations.
- Unlike classic methods which use discrete wavelet transform (DWT), we use UDWT to enhance the performance of system and to highlight the effectiveness of UDWT. We show experimentally that the generated features by UDWT are more robust to facial expressions, noise and misalignments in comparison with generated features by DWT.
- Unlike the method in [23] which employs fixed weights for feature vectors, in this paper a new method for weighting process is proposed.
- Unlike the method in [23] which uses Principal Component Analysis (PCA) coupled with Linear Discriminant Analysis (LDA) as dimension reduction, in this paper a Spectral regression discriminant analysis (SRDA) is used for this purpose.
- Unlike the previous methods in NIR domain which examined proposed method's performance in the context of some of challenges [31], in this study comprehensive experiments and analysis encompassing all of challenges are done.
- The proposed method has high accuracy in the presence of eyeglasses whereas our previous work [28] has deficiency in this case.
- The proposed face recognition system is extensively evaluated on CASIA NIR database and PolyU-NIRFD database and the results show that our proposed method shows good performance for most common challenges such as facial expression, eyeglasses, head rotation, noise and misalignments. Therefore, it is an accurate method which can be the core for real-world face recognition systems.

The remainder of the paper is organized as follows. In Sections 2 and 3 a brief review of undecimated discrete wavelet transform and Zernike moments are provided. Spectral regression discriminate analysis is discussed in Section 4. The proposed algorithm is provided in Section 5. Experimental results are presented in Section 6. Finally, Section 7 concludes this paper.

2. Undecimated discrete wavelet transform

In the last two decades, wavelets have become very popular due to their flexibility, locality and their high ability to analyze image at different resolutions or scales [22,32]. They have been successfully used in many fields, such as, image processing, signal analysis and pattern recognition. The two dimensional wavelet transforms can be carried out as a set of filter banks including a low-pass and high-pass filter, each followed by downsampling by a factor of two ($\downarrow 2$). One of the major problems with discrete wavelet transform (DWT) is that it suffers from translation-sensitivity. In other words, a simple shift of the input signal may change all coefficients of wavelet transform [22,32]. This deficiency is visualized in Fig. 1.

UDWT arises as a good solution to deal with the DWT's translation sensitivity. It can be considered to be an approximation to the continuous wavelet transform which removes a downsampling process from the DWT to produce over complete representation. From the implementation point of view in the context of filter banks, the filtering process is done without any downsampling (decimation), so all bands keep the same size as the original image. The implementation process of UDWT is shown in Fig. 2. The shift-invariance is achieved by two main parts. In the first part, filter coefficients of selected wavelet are upsampled by a factor of $2^{(j-1)}$ in the jth level. In the second part, the coefficients of the approximation are convolved with an upsampled version of the original filters to generate coefficients of approximation and details in the next levels. The result is shift-invariant wavelet transform. It has been already used for face virtual pose generation, translation-invariant feature extraction and object tracking [33–35].

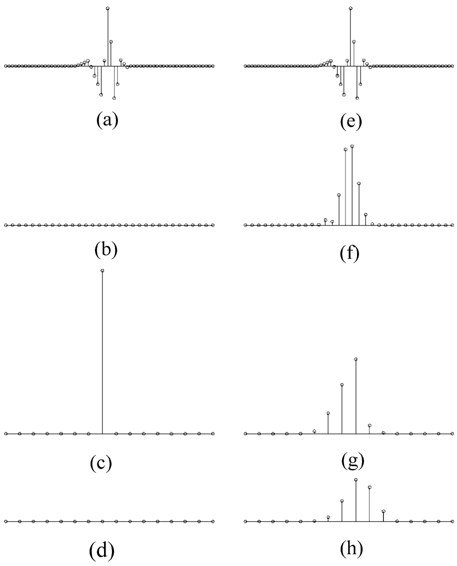

Fig. 1. Illustration of the shift sensitivity of discrete wavelet transform: (a) original signal, (b, c, d) coefficients of discrete wavelet transform subbands related to the original signal, (e) shifted signal "a" to right (one unit), (f, g, h) new coefficients of discrete wavelet transform subbands related to shifted signal [36,37].

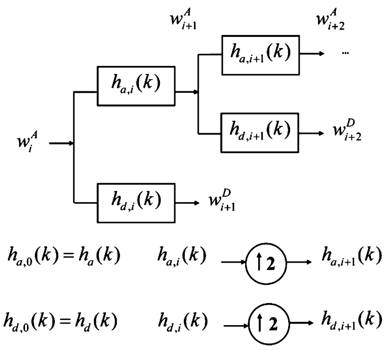

Fig. 2. Implementation of UDWT. $h_{a,i+1}$ and $h_{d,i+1}$ are filters in each level which are up-sampled versions of the previous [38].

3. Zernike moments

Zernike moments (ZMs) as powerful feature extractors have received a lot of attention in pattern recognition field due to their good performance in recognition of circular shapes such as face and their high robustness to noise [39]. Their history dated back to 1980 when Teague introduced ZMs based on the theory of orthogonal polynomials in image analysis [40]. The discrete approximation of ZMs of order p with repetition q is defined by the following equation for an image function $f(x, y)$:

$$Z_{pq} = \lambda_Z(p, R, C) \sum_{i=0}^{R-1} \sum_{k=0}^{C-1} R_{pq}(r_{ik}) e^{-jq\theta_{ik}} f(i, k) \tag{1}$$

where $N = \max(R, C)$,

$$\lambda_Z(p, R, C) = \frac{2(p+1)}{\pi(R-1)(C-1)} \tag{2}$$

$$x_i = \frac{\sqrt{2}}{N-1} i - \frac{1}{\sqrt{2}}, \qquad y_k = \frac{\sqrt{2}}{N-1} k + \frac{1}{\sqrt{2}} \tag{3}$$

$$r_{ik} = \sqrt{x_i^2 + y_k^2}, \qquad \theta_{ik} = \tan^{-1}\left(\frac{y_k}{x_i}\right) \tag{4}$$

and R_{pq} is the real-valued radial polynomial which is expressed as follows:

$$R_{pq} = \sum_{k=0}^{\frac{p-|q|}{2}} (-1)^k \frac{(p-k)!}{k!(\frac{p+|q|}{2} - k)!(\frac{p-|q|}{2} - k)!} r^{p-2k} \tag{5}$$

where $p \geq q$ and $p - |q|$ is even.

Zernike moments have proved a good performance in face recognition [41,42] but if they are used solo, they cannot handle properly local occlusions. That is why a combination with appropriate local features is desirable.

4. Spectral regression discriminant analysis

4.1. Definition

Spectral Regression Discriminant Analysis (SRDA) is one of the advanced techniques designed for high dimensional discriminant analysis [43]. In this technique, discriminant analysis is cast into regression framework by using spectral graph analysis that accelerates computation and simplifies the use of the regularization technique which can cope with small sample size (*SSS*) problem. It shares the same objective function of the original Linear Discriminant analysis (LDA). Especially noteworthy is that unlike LDA, in SRDA only a set of regularized least square problem is solved and no eigenvector computation is involved, which leads to considerable save of both time and memory.

4.2. Algorithmic procedure

Suppose $x_1, \ldots, x_m \in \mathbb{R}^n$ be a set of data points that belong to c classes and m_k be the number of samples in the kth class $(\sum_{k=1}^{c} m_k = m)$. The algorithmic procedure of SRDA is given as follows:

4.2.1. Response generation

Let $\quad y_k = [\underbrace{0, \ldots, 0}_{\sum_{i=1}^{k-1} m_i}, \underbrace{1, \ldots, 1}_{m_k}, \underbrace{0, \ldots, 0}_{\sum_{i=k+1}^{c} m_i}]^T, \quad k = 1, \ldots, c$

and $y_0 = [1, 1, \ldots, 1]^T$ be a vector of ones. Now y_0 is taken as the first vector and then the Gram–Schmidt process is used to orthogonalize $\{y_k\}$. Since y_0 is in the subspace spanned by $\{y_k\}$, $c - 1$ vectors are obtained as follows:

$$\{\bar{y}_k\}_{k=1}^{c-1} \quad (\bar{y}_i^T y_0 = 0,\ \bar{y}_i^T \bar{y}_j = 0,\ i \neq j) \tag{6}$$

4.2.2. Regularized least squares

In the second step, first a new element "1" is appended to each x_i and is still denoted by x_i for simplicity. Then $c - 1$ vectors $\{a_k\}_{k=1}^{c-1} \in \mathbb{R}^{n+1}$ which are the basis vectors of SRDA are calculated by the regularized least squares problem as follows:

$$a_k = \arg\min_a \left(\sum_{i=1}^{m} (a^T x_i - \bar{y}_i^k)^2 + \alpha \|a\|^2 \right) \tag{7}$$

where $\bar{y}_i^k$ denotes the ith element of $\bar{y}^k$ and α is a regularization parameter that controls the smoothness of the estimator.

4.2.3. Embedding to $(c - 1)$ dimensional subspace

Suppose $A = [a_1, \ldots, a_{c-1}]$ be an $(n+1) \times (c-1)$ transformation matrix, in the last step the samples can be embedded into $(c - 1)$-dimensional subspace as follows:

$$x \to z = A^T \begin{bmatrix} x \\ 1 \end{bmatrix} \tag{8}$$

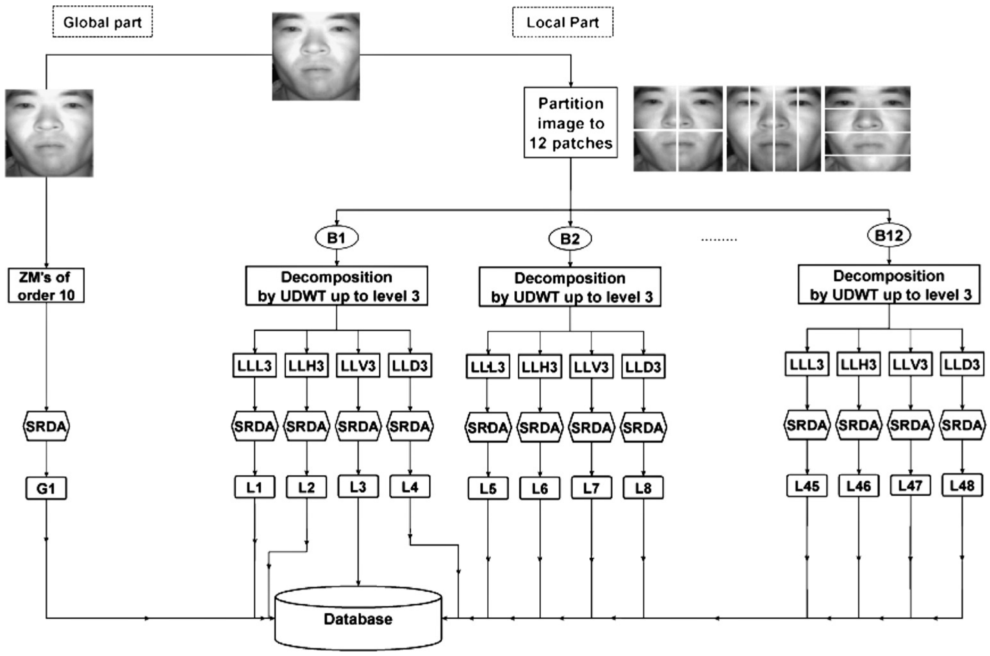

Fig. 3. The block diagram of the proposed face recognition system (training phase).

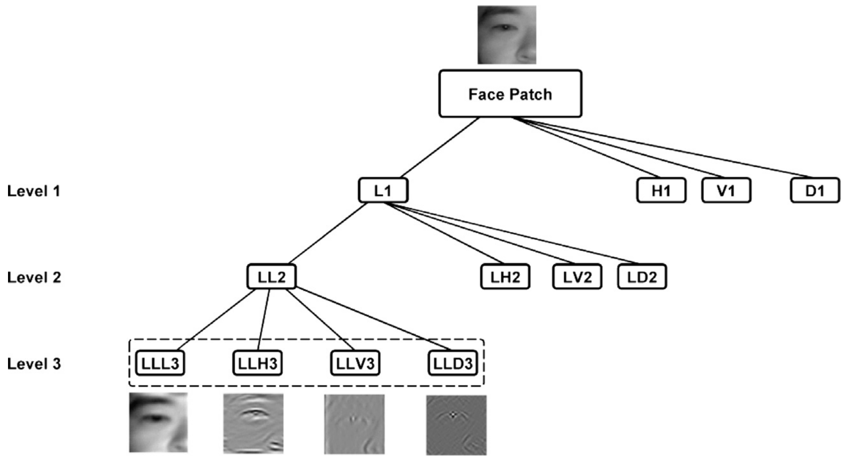

Fig. 4. Wavelet decomposition tree and spatial-frequency used in the proposed system.

5. The proposed system

The proposed face recognition system includes different components which are discussed in detail in this section. The block diagram of the proposed face recognition system is shown in Fig. 3. This is followed by feature extraction, decision fusion and weights computing phases which are expressed in detail.

5.1. Feature extraction

Feature extraction is the most important phase in face recognition methods which considerably affect the algorithm's performance. As shown in Fig. 3, facial features in the proposed approaches are classified into local and global features. In this subsection, the feature extraction process which is composed of local and global feature extraction procedures is discussed in detail.

5.1.1. Local feature extraction

Local features contain the local information of facial images which are dependent on position and orientation. The process of local feature extraction can be summarized as follows:

- In the first step of the local part, an image is partitioned into 12 patches to produce stable and meaningful information (Fig. 3).
- In the second step, every patch is decomposed to 3 levels using UDWT. As shown in Fig. 4, in the first level L_1, H_1, V_1 and D_1 are generated. Due to the high sensitivity of high frequency H_1, V_1 and D_1 to facial expressions and noise and to speed up the algorithm, they are not used and generated. Moreover, it has been shown that these subbands result in low performance for classification [44]. Accordingly the generated subbands in level 3 are used in our proposed method due to the best performance of generated subbands in comparison with other subbands in level 1 and 2 (Table 4). Because UDWT does not have downsampling, the resolution of decomposed patches is the same as the original patches.
- Since a large number of data are generated in the previous step, using all of the data will cause the system to be computational expensive. Hence a dimension reduction technique such as PCA or LDA should be used. PCA cannot present good results since the number of images in training set is not high. The LDA cannot be used due to "small sample size" problem.

Regularized LDA (RLDA) also needs a large memory to store a matrix with a size of $1024 \times 1024 = 1\,048\,576$ which imposes a costly computational complexity on the process [45]. Hence to reduce the burden of using all of the generated coefficients, in the last part SRDA is applied to decrease the dimension of data and produce the salient features. Finally 48 (12×4) distinct data vectors related to 12 patches are produced.

5.1.2. Global feature extraction

Global features contain the holistic information about facial images. Accordingly, the global information is encoded by means of ZMs. The process of global feature extraction can be summarized as follows:

- ZMs up to order 10 are calculated for an image to generate global features. Since ZMs are complex valued, imaginary part, real part and magnitude of ZMs are used as a data vector and they are concatenated together.
- Low classification accuracy might have stemmed from the simultaneous usage of the generated data vector that encompasses both low and high discriminative features. Therefore, in the last step, data vectors are sent to SRDA to remove the low-discriminable features and to enhance the discrimination power of the system.

5.2. Decision fusion

The last part of the proposed method is decision fusion. It consists of 5 steps which are described below:

5.2.1. Calculation of confidence score

Suppose we have c classes with m_i samples per class (m_i is the number of samples in ith class) in our database, then the confidence score $S_{v,i}$ of the system decision for vth feature vector and ith class is defined as follows:

$$S_{v,i} = \frac{1}{\min\limits_{1 \le p \le m_i} d(FV_v(test), FV_{v,i,p}(database))}, \quad v = 1, \ldots, 49, \; i = 1, \ldots, c \tag{9}$$

where FV_v (test) is the vth feature vector related to test images, $FV_{v,i,p}$ (database) is the vth feature vector of pth sample of ith class related to database and $d(.)$ stands for the distance function (Euclidean distance) between two feature vectors.

5.2.2. Normalization of confidence score

In this step, the values of confidence score are normalized as follows. Suppose the confidence score to vth feature vector of ith class be $S_{v,i}$ and there are c classes in our database and the normalized values are denoted as $N_{v,i}$, which are calculated as follows:

$$N_{v,i} = \frac{S_{v,i}}{\sum_{i=1}^{c} S_{v,i}} \quad \Rightarrow \quad \sum_{i=1}^{c} N_{v,i} = 1 \tag{10}$$

5.2.3. Formation of decision profile matrix

In this phase the decision profile matrix is formed as follows:

$$\mathbf{D_1}(\mathbf{P}) = \begin{bmatrix} \overbrace{\begin{matrix} N_{L,1,1} & N_{L,1,2} & \cdots & N_{L,1,c} \end{matrix}}^{\sum_{i=1}^{c} N_{L,1,i}=1} \\ \begin{matrix} N_{L,2,1} & N_{L,2,2} & \ldots & N_{L,2,c} \\ \vdots & \vdots & \vdots & \vdots \\ N_{L,48,1} & N_{L,48,2} & \cdots & N_{L,48,c} \\ N_{G,49,1} & N_{G,49,c} & \cdots & N_{G,49,c} \end{matrix} \end{bmatrix}_{49 \times c} \tag{11}$$

where $N_{L,v,i}$ is the normalized confidence value of vth local feature vector related to ith class ($v = 1, \ldots, 48$ and $i = 1, \ldots, c$) and $N_{G,i}$ is the normalized confidence value of global feature vector related to ith class. Since we have 48 local feature vectors and 1 global feature vector the dimension of $\mathbf{D_1}(\mathbf{P})$ is 49 by c.

5.2.4. Combined decision

In the last part, combined decision is applied. It consists of two main parts which are described as follows:

- In the first part, the normalized confidence values of local part are combined with a weighted sum rule for every class which is formulated as follows:

$$N_{L,i} = \sum_{v=1}^{48} \lambda_v N_{L,v,i}, \quad i = 1, \ldots, c \tag{12}$$

where $N_{L,v,i}$ is the normalized confidence value of vth local feature vector related to ith class and λ_v is the weight of vth local feature vector. Hence c values corresponding to c classes are resulted in this step and a new decision profile matrix is as follows:

$$\mathbf{D_2}(\mathbf{P}) = \begin{bmatrix} N_{L,1} & N_{L,2} & \cdots & N_{L,c} \\ N_{G,1} & N_{G,2} & \cdots & N_{G,c} \end{bmatrix} \tag{13}$$

where the first row is related to local feature vector and the second row is related to global feature vector.

- In the second part, the normalized confidence value of global part ($N_{G,i}$) and the summation of normalized confidence value of local part ($N_{L,i}$) are combined again with a weighted sum rule which is formulated as follows:

$$N_{E,i} = (\lambda_L) N_{L,i} + (1 - \lambda_L) N_{G,i} \tag{14}$$

where λ_L is the weight of $N_{L,i}$ and balances the importance of local and global information. The reason for using this strategy is that the performance of local and global information are quite different [16]. In a nutshell the decision profile matrix is resulted as follows:

$$\mathbf{D_3}(\mathbf{P}) = \begin{bmatrix} N_{E,1} & N_{E,2} & \cdots & N_{E,c} \end{bmatrix} \tag{15}$$

5.2.5. Decision strategy

The last stage of this part is the decision strategy. In this step, the label of the highest $N_{E,i}$ ($i = 1, \ldots, c$) is selected as the output of system.

5.3. Weights computing for local features, local part and global part

In this section, the procedure of weight computing is described. In the first part the weights of local feature vectors (λ_v) which are used in decision fusion part are derived. This is followed by weight computation of local part (λ_L) and global part (λ_G).

5.3.1. Weight computation of local features

Suppose we have c classes with m_i samples per class. Let FV_v (test) be the vth feature vector of test sample and $FV_{v,i,p}$ (database) is the vth feature vector of pth sample related to ith class. The weight computation method for local part consists of three steps which are described as follows:

- In the first step, the minimum distances between each local feature vector of test image and local feature vectors of database images are calculated. It can be described by the following formula:

$$D_{v,i} = \min_{1\le p\le m_i} d\big(FV_v(test), FV_{v,i,p}(database)\big),$$
$$v = 1,\ldots,48,\ i = 1,\ldots,c \tag{16}$$

- In the second step the distances related to each local feature vector over all classes are included in a vector which can be written as follows:

$$\mathbf{D_v} = (D_{v,1}, D_{v,2}, \ldots, D_{v,c}) \tag{17}$$

- In the third step the values of $\mathbf{D_v}$ are sorted increasingly for each v and a vector called D'_v is formed as follows:

$$\mathbf{D'_v} = \big(D'_{v,1}, D'_{v,2}, \ldots, D'_{v,c}\big) \tag{18}$$

where $D'_{v,1} \le D'_{v,2} \le D'_{v,3} \le \ldots \le D'_{v,c}$ for each v.

- In the fourth step the weight of vth feature vector (λ'_v) is calculated by the following formula:

$$\lambda'_v = \frac{D'_{v,2}}{D'_{v,1}}, \quad v = 1,\ldots,48 \tag{19}$$

- Finally the values of λ'_v are normalized by the following equation:

$$\lambda_v = \frac{\lambda'_v}{\sqrt{\lambda_1'^2 + \lambda_2'^2 + \ldots + \lambda_{48}'^2}}, \quad v = 1,\ldots,48 \tag{20}$$

5.3.2. Weight computation of local part and global part

Since we calculated 48 weights which are related to local feature vectors, we take the average of these weights as a representative of local part. Hence in this part the weight of the local part λ_L and weight of global part λ_G are defined as follows:

$$\lambda_L = mean(\mathbf{v}) \tag{21}$$

$$\lambda_G = 1 - \lambda_L \tag{22}$$

where $\mathrm{v} = [\lambda_1, \lambda_2, \ldots, \lambda_{48}]$ is a vector including weights of local feature vectors which were calculated previously. It should be noted that Eqs. (20)–(21) were found heuristically. Other normalizations and choices of λ_L are also possible.

6. Experimental results and performance analysis

In this section, we investigate the performance of the proposed method using CASIA NIR database [17] and PolyU-NIRFD database [24]. Comparative study is carried out against some existing face recognition schemes:

- Linear Discriminant Analysis (LDA) [3].
- Principal Component Analysis (PCA) [4].
- Kernel Principal Component Analysis (KPCA) [5].
- Pseudo Zernike Moments (PZMs) + Radial Basis Function Neural Network (PZMRBF) [46].
- Orthogonal Locality Preserving Projection (OLPP) which is also called orthogonal Laplacianface [47].
- Local binary pattern (LBP) + Fisherface (LBPF) [17].
- The method using decimated redundant discrete wavelet transform and Fisherface (DRDWTF) [23].

Furthermore, one more experiment based on discrete wavelet transform + SRDA (DWTSRDA) is conducted to highlight the contributions of UDWT. Descriptions of the settings of the aforementioned methods for performance evaluation are summarized in Table 1. In all experiments, we applied our method in three versions, global features based on Zernike moments only (denoted as ZMSRDA), local UDWT features only (denoted as UDWTSRDA) and both global and local part together (DF). The global part is basically identical (except SRDA reduction) with popular simple methods referred in [41,42].

Table 1
Descriptions of settings for different methods used in performance evaluation.

Method	Specification
LDA	Fisherface technique is utilized.
PCA	The eigenspace distance measure is Mahalanobis.
KPCA	Polynomial is used as a kernel and the polynomial parameter is set to 0.7.
PZMRBF	It is combination of Pseudo Zernike Moments and Radial Basis Function (RBF) neural network. The order of PZM is 10. The number of input layers equals to the dimension of feature vectors which are generated by PZM and the number of output layers equals to the number of classes. Spread value of radial basisfunctions is set as 0.9.
OLPP	Supervised OLPP is used. The weight metric criterion is cosine.
LBPF	$\mathrm{LBP}^{U_2}_{8,1}$ is used. The image is first divided into 64 blocks of size 8 by 8 and then an LBP histogram is calculated for each block. Finally Fisherface technique is used to decrease the dimension of features.
DRDWTF	This method is based on decimated redundant discrete wavelet transform. The decomposition level is 3 and the wavelet basis is 'Db 4'. The method is based on a combination of local and global features in decision step. The Fisherface technique is used for dimension reduction.
DWTSRDA	All of the parameter settings are the same as the local part of the algorithm (UDWTSRDA).

In the first part of this section, we briefly describe the database and preprocessing. This is followed by the experiments carried out to evaluate the performance of different methods and comparison between them.

The following sets of experiments are carried out:

- Testing the performance of the proposed method for parameter settings.
- Testing the performance of system against facial expression variations.
- Testing the influence of eyeglasses on the performance of the system.
- Testing the performance of the algorithms in the presence of head rotation in the x and y-axis.
- Testing the robustness of the system against additive zero-mean white Gaussian noise.
- Testing the performance in the presence of misalignments.
- Measurement of computation time.

6.1. Image normalization and database

The face images of CASIA NIR database [17] and PolyU-NIRFD database[1] [24] (Fig. 5) are used in our experiments. The database specifications are described in Table 2. The sizes of the training set, gallery set and probe set for CASIA NIR are 700, 800 and 800 respectively. The sizes of the training set, gallery set and probe set for PolyU-NIRFD are 500, 300 and 500 respectively. There is no overlap between training set, gallery set and probe set. The flow of preprocessing is as follows.

- Face images are aligned by placing the two eyes at fixed position (Fig. 5(b)).
- Face images are cropped to remove hair and background (Fig. 5(c)).
- Face images are resized to 64 × 64 with 256 gray levels to decrease computational time (Fig. 5(d)). This resizing is decided experimentally (data not shown) as choosing a larger size does

[1] http://www4.comp.polyu.edu.hk/~biometrics/polyudb_face.htm.

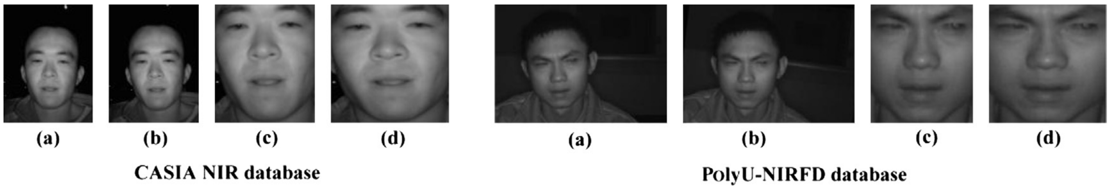

Fig. 5. The proposed preprocessing method: (a) raw image, (b–c) preprocessing steps, (d) the normalized images.

Table 2
Summary of the CASIA NIR database and PolyU-NIRFD database [17].

	Database	
	CASIA NIR database	PolyU-NIRFD database
Acquisition device	Home-brew camera with 850 nm wavelength	JAI camera with 850 nm wavelength
No. of subjects	197	335
Number of still images per subject	20	100
Distance	50 centimeters and 100 centimeters	80 centimeters and 120 centimeters
Resolution	640 × 480	768 × 576
Format	BMP	JPG

not significantly increase accuracy but increases computational time. The resized image still retains the useful information for face recognition.

6.2. *Testing the performance of the proposed system for parameter settings*

In this section, we assess the performance of system to enhance the performance of different components in the proposed system by setting different parameters with various allowed values. Three experiments were conducted in this section. 600 normal face images of 100 subjects (6 images per person) were chosen randomly from the training set of CASIA NIR database and used in our experiments. The normal faces are frontal faces without facial expression, head rotation and eyeglasses. Each test is repeated 20 times and the results are averaged. The following sets of experiments were conducted in this section:

- Testing the performance of ZMs with different orders.
- Performing best decomposition level and best subband analysis for UDWT.
- Testing the performance of the system with different values of α in SRDA and pick the optimum one to improve the performance of SRDA.

6.2.1. *Testing different orders of Zernike moments*

In the first experiment, the influences of different orders regarding to recognition rate and computational time (recognition time of all 300 test images) are checked and the results are shown in Table 3, respectively. As can be seen in this table, there is an improvement in the performance of the system when order of ZMs increases. The underlying reason is that, when higher orders are used, more salient features with higher discrimination power are generated which may increase the recognition rate. Further analysis shows that the satisfactory result is acquired with order 10 and the recognition rate of ZMs does not increase considerably when the order is higher than 10. Although higher others may result better accuracy, but the computational time of the system increases highly when the order is higher than 10 (Fig. 7). As a result, order 10 which may give a suitable trade-off between recognition rate and time computational time is used in the proposed algorithm and the feature vectors of order 0 to 10 are used for our further experiments.

Table 3
Face recognition results and cumulative time for testing 300 normal images on CASIA NIR database for ZMs.

Order	Cumulative dimensionality up to the specified order	Mean ± Std-Dev	Computational time (seconds)
0	1	10.29 ± 3.51	2.04
1	3	38.39 ± 3.64	2.48
2	6	50.28 ± 3.88	4.09
3	10	57.24 ± 3.77	4.97
4	15	60.34 ± 3.41	7.01
5	21	63.73 ± 2.54	9.53
6	28	69.37 ± 2.70	12.51
7	36	74.11 ± 2.43	15.32
8	45	80.27 ± 1.88	17.04
9	55	84.22 ± 2.18	20.22
10	**66**	**87.78 ± 1.90**	**23.65**
11	78	88.62± 2.79	28.73
12	91	89.12± 1.87	34.93

6.2.2. *Best decomposition level and best subband analysis*

In the second part, we search for the most discriminant decomposition level by evaluating the performance of its selected subbands. Because it is more time-consuming when the decomposition level L is larger than 3, $L = 3$ is used in the experiment. Hence, different subbands of UDWT at different levels are evaluated and the results are tabulated in Table 4. Daubechies 3 ("Db 3") is used as a basis due to its symmetry, orthonormal nature and compact support [48]. It is interesting to observe that level 3 attains the highest correct classification rates as compared to other levels. The same result can be found in [23,49]. In fact, with further wavelet decomposition, finer scale information which may include salient features is generated; however, the computational time of the system increases. Consequently, level 3 which may give a suitable trade-off between scale information and computational time is used in the proposed algorithm.

6.2.3. *Parameter selection for SRDA*

α (Eq. (7)) is an essential parameter that controls the smoothness of the estimator in SRDA algorithm. In this experiment the impact of parameter on the performance of our algorithm is checked. Two individual experiments for local and global part of the algorithm are evaluated. Different values of α are tested and the results are considered. As shown in Table 5 there is a significant improvement in the recognition rate when α is decreased and the optimum value is 0.01 with the highest recognition rate.

Table 4
Comparison of the performances of the system using different decomposition levels and subbands (Mean ± Std-Dev).

Level	Subband	UDWT
1	L1	93.51 ± 0.82
1	H1	91.04 ± 0.76
1	V1	93.34 ± 0.89
1	D1	81.67 ± 2.09
2	LL2	95.24 ± 1.46
2	LH2	92.19 ± 1.69
2	LV2	92.97 ± 0.90
2	LD2	93.40 ± 2.04
3	**LLL3**	**96.73 ± 0.87**
3	**LLH3**	**92.59 ± 0.84**
3	**LLV3**	**94.50 ± 0.97**
3	**LLD3**	**95.58 ± 0.68**

Table 5
Effect of different values of α on the face recognition rate (Mean ± Std-Dev).

α	Local part (UDWTSRDA)	Global part (ZMSRDA)
0.400	95.70 ± 0.93	94.38 ± 0.80
0.100	96.70 ± 0.75	95.93 ± 0.71
0.070	98.36 ± 0.81	96.33 ± 0.41
0.040	99.05 ± 0.81	96.97 ± 0.74
0.010	**99.05 ± 0.81**	**97.40 ± 0.65**
0.009	99.05 ± 0.81	97.40 ± 0.65
0.008	99.05 ± 0.81	97.40 ± 0.65

6.3. Testing the performance of the algorithms in facial expression variations

Facial expressions are one of the grand challenges in face recognition methods [50]. In this section some experiments are performed on the face images with different facial expressions to evaluate the robustness of the different methods to variations in facial expressions. 120 frontal normal images of 40 subjects belonging to gallery set are randomly chosen and used as gallery images and 120 random images having facial expression variations are chosen from probe set and used as probe images. Some samples of normal images and images with facial expressions from CASIA NIR and PolyU-NIRFD are shown in Fig. 6. The test is repeated 20 times and the results are averaged to get the representative values. Tables 6 and 7 show the mean recognition rates of different methods along with standard deviation and confidence interval with 95% significant level. Confidence interval is one of most useful criterion to evaluate the reliability of results. Smaller confidence intervals indicate the high precision of the method. We use the following formula to calculate confidence interval

$$\left[\bar{x} - z^* \times \frac{\delta}{\sqrt{n}}, \bar{x} + z^* \times \frac{\delta}{\sqrt{n}}\right] \tag{23}$$

where $\bar{x}$ stands for mean, z^* is the score for level of confidence which is 1.96 when the significance level is 95%, n is the number of observations which is 20 in our case and δ stands for standard deviation.

The following conclusions can be made:

- As shown in Tables 6 and 7, the local part of the proposed method (UDWTSRDA) has better performance in comparison with DWTSRDA counterpart in terms of the recognition rate. The underlying reason lies in the fact that UWDT generates the full resolution subbands without any downsampling. Accordingly, more discriminative information is achieved compared to DWTSRDA.
- Since both local and global parts of the algorithm have high robustness to facial expressions the performance of the proposed method is not severely affected by the expression variations in both databases. Further analysis shows that there is no overlap between the confidence intervals of the proposed method and those of other methods. This narrow width of confidence intervals of the proposed method shows the high accuracy, precision and robustness of the proposed method to various training and test samples.
- The proposed method outperforms the DRDWTF method counterpart which is also based on local and global features. This result may be explained by a number of different factors. First, ZMs which are used as global features in our method have better performance in comparison with generated global features by DRDWT in the presence of facial expressions. Second, SRDA as a sophisticated technique, improves the accuracy of the system compared with classical Fisherface which is used in DRDWTF as a dimension reduction technique.
- Comparing ZMSRDA and PZMRBF, it can be seen that the ZMSRDA achieves better performance than PZMRBF. This seems to contradict [41] which showed that PZMs work better than ZMs. The explanation is that in ZMSRDA, the discrimination power of the features is enhanced by SRDA. Obviously, the resulted features are more salient than raw features in PZMRBF.
- What is interesting in Tables 6 and 7 is that OLPP performs the best among other appearance-based methods including LDA, PCA, and KPCA. The underlying reason is that OLPP has more discriminative power than other methods due to orthogonal basis function in OLPP which alleviates the problem of data reconstruction better compared to other appearance-based methods.

6.4. Testing the performance of the algorithms for images with eyeglasses

Wearing glasses generally affects the performance of FR systems and it has been introduced as one of the remaining challenging

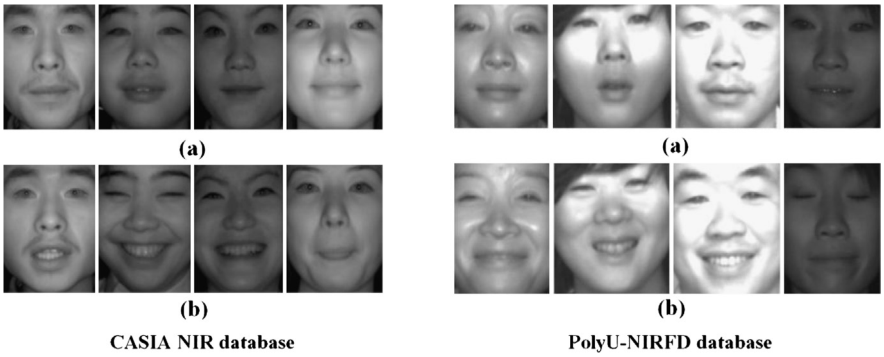

Fig. 6. (a) Normalized images used as gallery images, (b) images with facial expression from CASIA NIR database and PolyU-NIRFD used as probe images.

Table 6
Performance of different methods in the presence of different challenges (CASIA NIR database).

Algorithm	Mean	Standard deviation	Confidence interval
Face recognition results under different facial expressions			
LDA	91.12	2.03	[90.23, 92.00]
PCA	75.45	3.82	[73.77, 77.12]
KPCA	83.20	2.21	[82.23, 84.16]
PZMRBF	86.76	1.24	[86.21, 86.21]
OLPP	93.16	1.01	[92.71, 93.60]
LBPF	94.62	1.30	[94.05, 95.18]
DRDWTF	92.54	1.13	[92.04, 93.03]
DWTSRDA	91.04	1.29	[90.47, 91.60]
ZMSRDA (global part)	92.59	1.32	[92.01, 93.16]
UDWTSRDA (local part)	93.41	0.71	[93.09, 93.72]
DF (decision fusion)	96.58	0.74	[96.25, 96.90]
Face recognition results of different methods for images with eyeglasses			
LDA	89.16	1.62	[88.45, 89.87]
PCA	82.95	2.68	[81.77, 84.12]
KPCA	86.25	1.78	[85.46, 87.03]
PZMRBF	70.24	1.71	[69.49, 70.98]
OLPP	87.75	2.16	[86.80, 88.69]
LBPF	96.41	1.01	[95.96, 96.85]
DRDWTF	94.33	0.69	[94.02, 94.63]
DWTSRDA	95.16	0.95	[94.74, 95.57]
ZMSRDA (global part)	72.08	3.29	[70.63, 73.52]
UDWTSRDA (local part)	97.29	1.37	[96.68, 97.89]
DF (decision fusion)	98.50	0.95	[98.08, 98.91]
Face recognition results of different methods for images with head rotation in the x-axis			
LDA	64.37	4.01	[62.61, 66.12]
PCA	39.29	4.57	[37.28, 41.29]
KPCA	58.12	2.37	[57.08, 59.15]
PZMRBF	73.12	2.08	[72.20, 74.03]
OLPP	71.62	2.58	[70.48, 72.75]
LBPF	70.16	4.82	[68.04, 72.27]
DRDWTF	72.66	3.66	[71.05, 74.26]
DWTSRDA	71.87	4.25	[70.00, 73.73]
ZMSRDA (global part)	74.75	1.66	[74.02, 75.47]
UDWTSRDA (local part)	80.25	2.67	[79.07, 81.42]
DF (decision fusion)	85.82	1.21	[85.28, 86.35]

Table 7
Performance of different methods in the presence of different challenges (PolyU-NIRFD database).

Algorithm	Mean	Standard deviation	Confidence interval
Face recognition results under different facial expressions			
LDA	93.25	1.26	[92.69, 93.80]
PCA	91.66	1.48	[91.01, 92.30]
KPCA	83.20	2.21	[82.23, 84.16]
PZMRBF	88.66	2.01	[87.77, 89.54]
OLPP	95.29	1.48	[94.64, 95.93]
LBPF	94.58	1.28	[94.01, 95.14]
DRDWTF	93.70	1.76	[92.92, 94.47]
DWTSRDA	93.22	1.04	[92.76, 93.67]
ZMSRDA (global part)	93.91	1.11	[93.42, 94.39]
UDWTSRDA (local part)	95.83	0.90	[95.43, 96.22]
DF (decision fusion)	97.58	0.71	[97.26, 97.89]
Face recognition results of different methods for images with head rotation in the y-axis			
LDA	30.50	4.15	[28.68, 32.31]
PCA	27.87	1.61	[27.16, 28.57]
KPCA	34.08	2.26	[33.08, 35.07]
PZMRBF	45.24	1.72	[44.48, 45.99]
OLPP	35.12	4.42	[33.18, 37.05]
LBPF	35.62	3.70	[33.99, 37.24]
DRDWTF	37.00	4.92	[34.84, 39.15]
DWTSRDA	29.16	1.98	[28.29, 30.02]
ZMSRDA (global part)	47.16	3.72	[45.52, 48.79]
UDWTSRDA (local part)	40.12	4.00	[38.36, 41.87]
DF (decision fusion)	54.01	2.01	[53.12, 54.89]

issues in FR algorithms [50]. To determine the effect of eyeglasses on the performance of systems another experiment is designed. Since the subject never wears glasses in the target imagery, 120 random images of 40 subjects (3 images per person) without eyeglasses are used as gallery images and 120 random images with eyeglasses are used as probe images. We just use the images from CASIA NIR database since PolyU-NIRFD database does not include such a scenario. The images are chosen from gallery set and probe set of CASIA NIR database respectively. Some samples of images are shown in Fig. 7. The test is repeated 20 times and the results

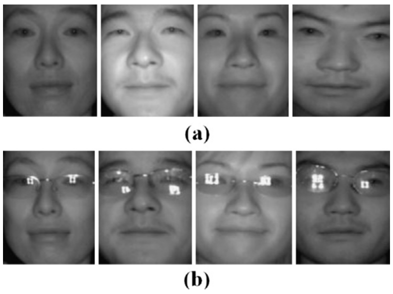

Fig. 7. (a) Normalized images without eyeglasses used as gallery images, (b) images with eyeglasses from CASIA NIR database used as probe images.

are averaged to get the representative values. The face recognition results are tabulated in Table 6 from which the following conclusions can be made.

- In Table 6, there is a clear trend of decrease in the recognition rate of appearance-based methods. A possible explanation for this is that the performance of appearance-based method largely relies on the representation of the training samples. Hence the recognition accuracy of appearance-based methods degraded sharply when the eyeglasses versus no-eyeglasses scenario is applied in experiments.
- In this experiment, wearing eyeglasses was found to cause considerable degradation in the performance of global part of system (ZMSRDA). The observed degradation in the recognition performance of ZMSRDA (global part) is attributed to the global structure of this approach which affects the values of all moments in the presence of eyeglasses. Hence the performance of ZMSRDA degrades sharply when local change such as eyeglasses occurs. The results here prove the high sensitivity of ZMs to eyeglasses which were mentioned as a crucial problem of ZMs in [28]. The same analysis applies in case of PZMRBF, too.
- LBPF has a good performance in this experiment. This finding is in agreement with the result proposed by Li et al. [17] which shows the high stability of the LBPF method to eyeglasses due to the local nature of the proposed method.
- Since local part of our method has high robustness in the presence of eyeglasses the accuracy of system is not affected in this case.

6.5. *Testing the performance of system in the presence of head rotation in the x-axis and the y-axis*

Head rotation is one of the most common challenges in face recognition systems which affects the performance of algorithms significantly. Handling head rotation is more difficult than other challenges. Different methods have been introduced to address the problem caused by head rotation [51,52]. Three types of head rotations are shown in Fig. 8. The rotations around x and y-axis cause the major difficulty in recognition because they significantly change the appearance and the visible part of the face. On the other hand, the in-plane rotation (head rotation in z-axis) is not a serious problem because such transformation only rotates the image without any occlusions and without changes of the visible parts. Moreover, the physical constraints often limit the in-plane rotation to small angles only.

In the current implementation, the proposed method tolerates the in-plane rotation up to 6 degrees (data not shown). If we know we have to deal with this kind of rotation, we can simply switch off all local features and use only the ZM magnitudes, which are rotation invariant. The method is then equivalent to a simple recognition system [28] and is completely rotation invariant. Hence, here we concentrated to the tests of the performances in the presence of head rotation in x-axis and y-axis. Two separate experiments are done in this section. In the first experiment the performance of the proposed method in the presence of head rotations in the x-axis is investigated. The images of CASIA NIR database with small degree of head rotations in the x-axis are used as probe images for this evaluation (Fig. 9). In the second experiment the performance of the proposed method in the presence of head rotation in the y-axis is examined. The facial images of PolyU-NIRFD with higher degree of head rotations in y-axis (both left and right) are used as probe images for this experiment (Fig. 9). It is necessary to mention that in the PolyU-NIRFD database the angles of head rotations in y-axis are sharper than in CASIA NIR database. The databases do not provide the angle values but we estimate that in CASIA NIR database the angles are less than 20 degrees while in PolyU-NIRFD database the angles used are even more than 45 degrees. In all of the experiments 120 random images of 40 subjects (3 images per person) without head rotations are used as gallery images and 120 random images with head rotations are used as probe images. Each test is repeated 20 times and the results are averaged to get the representative values. Tables 6 and 7 compare the results obtained from the analysis of head rotations with different methods.

From Tables 6 and 7, the following conclusions can be drawn.

- As expected, head rotation affects the recognition accuracy of methods considerably which highlights the importance of compensating head rotation in face recognition systems. This result may be explained by the fact that head rotations in x-axis and y-axis change in the visual appearances of the face image significantly. Hence the performance of the methods especially appearance-based method degrades sharply under head rotations. Further analysis shows that the performance of the methods based on PolyU-NIRFD database decreases more severely than those of the methods based on CASIA NIR database. The underlying reason is that the images in PolyU-NIRFD database include sharper yaw and roll angles. Hence the appearances of images are changed more significantly which affect the results as well.
- Although the proposed method cannot cope with head rotation in x-axis and y-axis, this problem can be successfully resolved by brute force only (i.e. by expanding the training set and including many rotated images) or by an application of 3D imaging technologies.
- No significant differences are found between the performances of ZMSRDA and PZMRBF in the presence of head rotation in x and y-axis.
- Both DRDWTF and our proposed method (DF) which are based on combination of local and global features have better accuracy in the presence of head rotation in comparison with other methods in both databases.
- Strong evidence of improvement in the recognition rate of system based on decision fusion can be seen in this experiment. As can be seen in Table 6 the recognition accuracy of the proposed method based on CASIA NIR database boosts almost 5 points from our local framework (UDWTSRDA). Further analysis based on PolyU-NIRFD database (Table 7) shows that the recognition accuracy of the proposed method increases almost 7 points from our global framework (ZMSRDA). This result proves that both local and global features are indeed mutually complementary which can handle pose variations effectively.

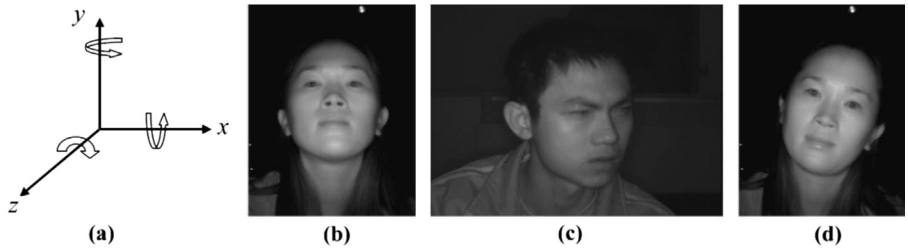

Fig. 8. (a) The direction of rotation, (b) rotation in x-axis, (c) rotation in y-axis, (d) rotation in z-axis.

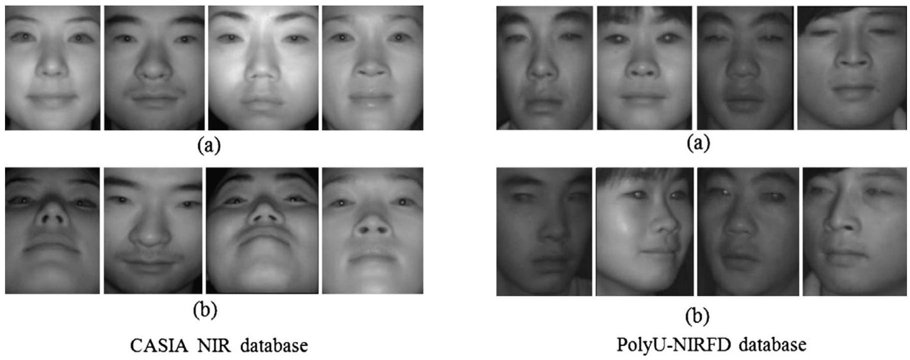

Fig. 9. (a) Sample of normalized images used as gallery images, (b) sample of images with head rotation in x-axis and y-axis used as probe images.

Fig. 10. Samples of images with different levels of white Gaussian noise. From left to right signal to noise ratio (SNR) is: no noise, 21 dB, 18 dB and 16 dB respectively.

Table 8
Effect of noise on the performance of different methods (Mean ± Std-Dev).

Algorithm	Signal to noise ratio (SNR) level			
	No noise	21	18	16
LDA	97.42 ± 0.91	96.05 ± 0.39	94.93 ± 0.81	89.87 ± 1.25
PCA	94.63 ± 1.02	86.17 ± 0.68	61.27 ± 1.37	44.36 ± 0.88
KPCA	95.66 ± 1.31	88.87 ± 1.01	66.06 ± 1.45	47.22 ± 1.04
PZMRBF	95.58 ± 0.71	92.11 ± 0.52	84.51 ± 1.11	75.11 ± 1.29
OLPP	98.31 ± 0.86	88.86 ± 0.92	67.17 ± 1.76	51.06 ± 1.68
LBPF	99.27 ± 0.43	3.46 ± 0.34	2.17 ± 0.31	1.77 ± 0.12
DRDWTF	99.42 ± 0.74	94.51 ± 1.08	85.75 ± 2.42	74.67 ± 1.57
DWTSRDA	99.36 ± 0.81	96.20 ± 0.72	91.52 ± 1.48	82.98 ± 1.82
ZMSRDA (global part)	98.90 ± 0.64	95.91 ± 0.77	89.65 ± 1.35	77.06 ± 1.48
UDWTSRDA (local part)	99.73 ± 0.51	99.11 ± 0.94	98.23 ± 0.69	97.11 ± 0.81
DF (decision fusion)	99.96 ± 0.25	99.75 ± 0.27	99.36 ± 0.35	97.70 ± 0.54

6.6. Testing the robustness against additive zero mean white Gaussian noise

Unlike visible imaging, noise is a serious problem in NIR imaging. To assess the effects of noise on the performance of algorithms and to evaluate the usability of different methods for non-cooperative systems such as video surveillance scenarios where the resolution of images is low, we conduct another experiment based on noisy images. 600 normal images of 60 subjects belong to CASIA NIR database (10 images per person) without eyeglasses, head rotation and facial expression are chosen from the gallery and probe set and used for this experiment. Some samples of degraded images are shown in Fig. 10. While training, 5 normal images without noise are used as gallery images. For probe images, the images are corrupted with additive zero mean white Gaussian noise with different signal to noise ratio (SNR) level. 20 realizations of noisy images are generated and their accuracies are averaged to get the representative value. The mean recognition rates along with standard deviations of different methods based on additive zero mean white Gaussian noise with different SNRs are shown in Table 8.

From this table, we have the following conclusions.

- It is apparent from Table 8 that our method performs significantly better than other methods in the presence of noise. As shown in Fig. 10 when SNR is 16, it is not easy to recognize the face images even by human eyes. However our proposed method does an excellent job in this case which shows the

Table 9
Performance comparison of different methods against misalignments using the robustness value as an evaluation measurement.

Methods	r_0	Translation		Rotation		Scale	
		r^*	R	r^*	R	r^*	R
LDA	97.42	94.80	0.97	73.97	0.75	81.59	0.83
PCA	94.63	91.07	0.96	80.45	0.85	79.90	0.84
KPCA	95.66	90.48	0.94	72.04	0.75	73.57	0.76
PZMRBF	95.58	94.41	0.98	93.27	0.97	87.32	0.91
OLPP	98.31	95.49	0.97	82.17	0.83	89.52	0.90
LBPF	99.27	97.32	0.98	96.73	0.97	97.33	0.98
DRDWTF	99.42	98.84	0.99	91.19	0.91	98.65	0.99
DWTSRDA	99.36	96.91	0.97	91.37	0.91	98.65	0.99
ZMSRDA (global part)	98.90	97.88	0.98	96.83	0.97	91.28	0.92
UDWTSRDA (local part)	99.73	99.18	0.99	96.71	0.96	99.07	0.99
DF (decision fusion)	99.96	99.51	0.99	98.11	0.98	99.46	0.99

high robustness of the proposed method to noise even when the noise strength is high.

- What is interesting is that there is a significant difference between the recognition accuracy of DWTSRDA and UDWTSRDA (local part) which shows the high robustness of UDWT to heavy noise. This finding is in agreement with the presented findings in [53] which showed the high robustness of UDWT for object shape prediction in the presence of noise. The reason for this superiority is the redundancy of UDWT which leads to full resolution low frequency components in comparison with DWT which generates decimated components.
- From this data we can see that the performance of our method is almost 23% higher than that of DRDWTF when SNR is 16. It seems possible that these results are due to the decimation process in DRDWTF which decreases the resolution of images and deteriorates the accuracy of system in the presence of noise.
- The recognition rate of LBPF is the lowest among the other recognition accuracies. The high sensitivity of local binary pattern to noise can be seen in this experiment. The underlying reason is that LBP thresholds exactly at the value of the central pixel. Hence original LBP tends to be sensitive to noise which limits its usability for applications which encounter with low resolution images. Our results here prove the high sensitivity of LBP to noise which were mentioned in the literature [18].

6.7. Evaluation of robustness against misalignments

Misalignment is one of the inevitable challenges in face recognition systems which typically results from inaccurate estimation of facial landmarks (eyes, nose, mouth, etc.). Many solutions have been proposed to solve this problem which could be roughly classified into three categories: invariant features, misalignment modeling and alignment returning. To determine the effects of misalignments Chen et al. in [54] calculated the standard deviation of manual labels on 30 randomly selected images from 25 individuals. A 3.2 pixel standard deviation is reported in their study. To quantify how misalignments influence the performance of algorithms, we conduct another experiment to evaluate the performance of algorithms in the presence of random translation, rotation and scale. To measure the degradation degree of the recognition method against different perturbations, we use a robustness concept, R which is introduced in [55] as follows:

$$R = \frac{r^*}{r_0} \tag{24}$$

where r^* is the recognition rate of system after specific perturbation and r_0 is the recognition rate of the system without any perturbation. R is a value between 0 and 1. A larger R implies less sensitivity or high robustness to a specific perturbation. 300 random normal images of 60 subjects (5 images per person) without facial expressions, eyeglasses, head rotation and misalignments are used as gallery images and 300 normal images including misalignments are used as probe images. To simulate the spatial misalignments a random translation, rotation and scaling are added to probe images separately and used in our experiments. The image translation is set as integer within $[-2, 2]$ pixel for both vertical and horizontal directions, the rotation is set randomly within $[-3°, 3°]$ and the scaling is set randomly within $[0.95, 1.05]$. To simulate the misalignments brought by the automatic face alignment process, we also apply the mixed spatial misalignments. Hence a horizontal translation $t_x \in [-1, 1]$ and vertical translation $t_y \in [-1, 1]$, a rotation $r \in [-3°, 3°]$ and a scaling $s \in [0.95, 1.05]$ are added to the original image and the results are considered. The average recognition rates of different methods with their related robustness values (Eq. (24)) are tabulated in Table 9, Table 10 and shown in Fig. 11 from which the following observations can be made.

Table 10
Performance comparison of different methods against mixed misalignments using the robustness value as an evaluation measurement.

Methods	Mixed misalignments		
	r_0	r^*	R
LDA	97.42	42.63	0.43
PCA	94.63	68.58	0.72
KPCA	95.66	45.32	0.47
PZMRBF	95.58	84.05	0.87
OLPP	98.41	60.72	0.61
LBPF	99.27	82.19	0.82
DRDWTF	99.42	86.75	0.87
DWTSRDA	99.36	71.85	0.72
ZMSRDA (global part)	98.90	87.59	0.88
UDWTSRDA (local part)	99.73	92.12	0.92
DF (decision fusion)	99.96	96.80	0.96

- The high degradation of the LDA, PCA, KPCA and OLPP especially in the presence of mixed misalignments shows the misalignment sensitivity of the subspace learning techniques which have been used widely in face recognition systems in the past three decades. The results here bolster the results presented in [56].
- As shown in Table 9 and Table 10, UDWTSRDA (local part) which is based on UDWT has better accuracy to misalignments in comparison with DWTSRDA which is based on DWT. This result is consistent with the literature [23] which mentioned the misalignments as a deficiency of DWT.
- Since DRDWTF uses translation invariant wavelet transform, it has good accuracy in the presence of misalignments. The findings of the current study are consistent with those of Li et al. [23] who found that DRDWTF has good robustness to misalignments.

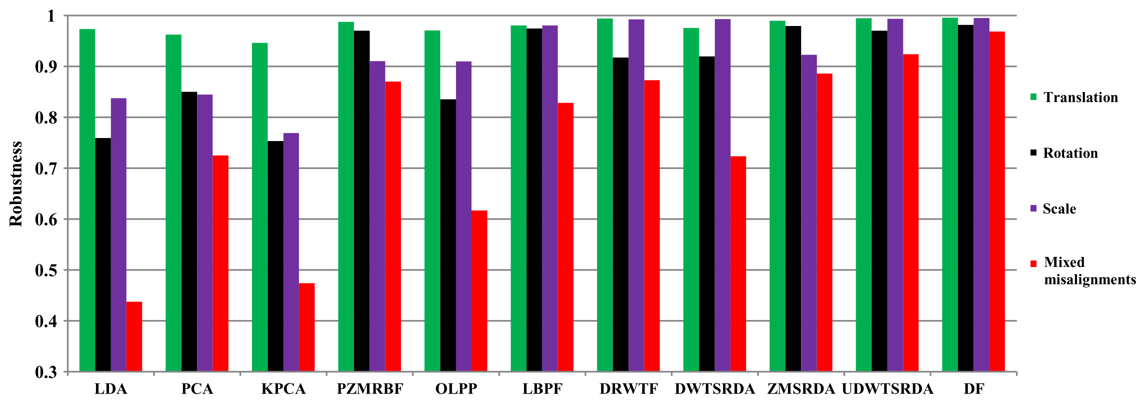

Fig. 11. The robustness value of different methods in the presence of misalignments.

Table 11
Comparison of recognition time for different algorithms.

Algorithm	Recognition time (s)
LDA	0.18
PCA	0.19
KPCA	0.24
PZMRBF	0.39
OLPP	0.20
LBPF	0.32
DRDWTF	0.47
DWTSRDA	0.38
ZMSRDA (global part)	0.32
UDWTSRDA (local part)	0.45
DF (decision fusion)	0.52

- From Table 10 we can see that the accuracy of LBPF decreased highly in the presence of mixed misalignments. This finding corroborates the ideas of Yi et al. [19] who mentioned misalignment as one of the deficiencies of LBP.
- Strong evidence of LDA sensitivity to mixed misalignments was found in this experiment. As can be seen in Table 10, the robustness value of the LDA technique is the lowest among other methods. It seems possible that these results are due to divergence of samples from the same class which enlarges the within-class scatter and reduces between-class scatter to some degree. The same result can be found in [55,56].
- Both ZMSRDA and PZMRBF have good performance in the presence of misalignments especially in the presence of rotation due to rotation invariance property of ZMs and PZMs.
- The current study found that mixed misalignments in probe images affect the accuracy of systems more highly than separate misalignments.
- Since both local and global parts of the proposed method have high robustness to misalignments, the highest value of robustness is achieved by the proposed method (DF) which shows the effectiveness of extracted features by the proposed method.

6.8. Measurement of computation time

To evaluate the computational load of the proposed method and other methods, in this section, the recognition time of each algorithm is calculated and shown in Table 11. The recognition time (per second) is the time needed to extract the features and classify one new probe image. All of our experiments were conducted using MATLAB R2013a on a Core 2 Duo 2.50-GHz Windows 7 machine with 4 GBytes of memory. As can be seen the time required for recognition of the new probe image by the proposed method is higher than that of the others. However, it is still allowing on-line processing, because the persons are supposed to approach the system one by one with at least one-second intervals.

7. Conclusion

We human beings, employ both global and local face features to identify faces. In this paper, we presented a novel algorithm based on the combination of local features extracted using UDWT and global features extracted using ZMs to compensate facial expressions, eyeglasses, head rotation, image noise and misalignments for near infrared face recognition. In the proposed method, local features are extracted from partitioned image patches by UDWT and the global features are extracted by calculation of ZMs. By applying SRDA on both local and global features, multiple feature vectors are obtained. Finally they are combined with a weighted sum rule to make a full use of local and global features.

This paper also compared the performance of the proposed method with popular face recognition algorithms in the presence of the most common challenges in face recognition systems. The CASIA NIR database and PolyU-NIRFD database were used to validate the performance of the proposed method and to compare with other known methods. Experimental results show that the combination of UDWT and ZMs, greatly improve the face recognition accuracy. The following conclusions can be drawn from the present study:

- Both local and global features are crucial for proposing a robust face recognition system and the combination of local and global information in decision part, improves the performance of the system significantly.
- ZMs are powerful feature extractor which can be used as global features in FR methods.
- UWDT has better performance than DWT especially in the presence of noise and misalignments. Hence its usage in FR method can improve the performance.

The study has gone some way towards enhancing our understanding the nature of different challenges in face recognition systems. Moreover, it makes several noteworthy contributions to automatic face recognition in the near infrared domain which is suitable for identification purpose in face recognition systems. Now we can precisely describe how different challenges affect the performance of algorithms. The important limitation of this study is related to interfacing with legacy systems that may only have visible imagery and may need matching visible imagery to NIR imagery. We are currently exploring this problem in theory and practice and future works will be on enhancing solutions to overcome these limitations.

Acknowledgments

The authors would like to thank Universiti Teknologi Malaysia (UTM) for the support in Research and Development, UTM Big Data Centre and Soft Computing Research Group (SCRG) for the inspiration in making this study a success, the Institute of Automation, Chinese Academy of Sciences (CASIA) for providing CASIA NIR database to carry out this experiment, Biometric Research Centre (UGC/CRC) for providing PolyU-NIRFD database and Institute of Information Theory and Automation (UTIA) for providing MATLAB codes. This work is supported by the Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS-4F347) and partially by the Czech Science Foundation under the grant No. P103/11/1552.

References

[1] W. Zhao, R. Chellappa, P.J. Philips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[2] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Trans. Circuits Syst. Video Technol. 14 (1) (2004) 4–20.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[4] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 13 (1) (1991) 71–86.
[5] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[6] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 721–732.
[7] S. Moritz, A.H. Jorgen, G.E. Erik, Physics-based modelling of human skin colour under mixed illuminants, Robot. Auton. Syst. 35 (3) (2001) 131–142.
[8] D.V. Jadhav, R.S. Holambe, Rotation, illumination invariant polynomial kernel Fisher discriminant analysis using Radon and discrete cosine transforms based features for face recognition, Pattern Recognit. Lett. 31 (9) (2010) 1002–1009.
[9] Pradeep Buddharaju, Ioannis T. Pavlidis, Panagiotis Tsiamyrtzis, M. Bazakos, Physiology-based face recognition in the thermal infrared spectrum, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 613–626.
[10] D.A. Socolinsky, A. Selinger, J.D. Neuheisel, Face recognition with visible and thermal infrared imagery, Comput. Vis. Image Underst. 91 (1) (2003) 72–114.
[11] X. Zou, J. Kittler, K. Messer, Face recognition using active near-IR illumination, in: Proc. British Machine Vision Conference, Oxford, UK, 2005, pp. 1–11, Citeseer.
[12] S.Z. Li, R. Chu, M. Ao, et al., Highly accurate and fast face recognition using near infrared images, in: D. Zhang, A. Jain (Eds.), International Conference on Biometrics, Springer, Hong Kong, 2005, pp. 151–158.
[13] S.Z. Li, Z. Lun, L. ShengCai, et al., A near-infrared image based face recognition system, in: Proc. 7th International Conference on Automatic Face and Gesture Recognition, IEEE, Southampton, UK, 2006, pp. 455–460.
[14] S. Zhao, R.-R. Grigat, An automatic face recognition system in the near infrared spectrum, in: P. Perner, A. Imiya (Eds.), Machine Learning and Data Mining in Pattern Recognition, Springer, Berlin/Heidelberg, 2005, pp. 437–444.
[15] J. Dowdall, I. Pavlidis, G. Bebis, Face detection in the near-IR spectrum, Image Vis. Comput. 21 (7) (2003) 565–578.
[16] S. Yu, S. Shiguang, C. Xilin, G. Wen, Hierarchical ensemble of global and local classifiers for face recognition, IEEE Trans. Image Process. 18 (8) (2009) 1885–1896.
[17] S.Z. Li, R. Chu, S. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 627–639.
[18] L. Nanni, A. Lumini, S. Brahnam, Local binary patterns variants as texture descriptors for medical image analysis, Artif. Intell. Med. 49 (2) (2010) 117–125.
[19] D. Yi, R. Liu, R. Chu, et al., Outdoor face recognition using enhanced near infrared imaging, in: Advances in Biometrics, Springer, Seoul, Korea, 2007, pp. 1–9.
[20] Y. He, G.F. Feng, F. Liu, H. He, Near infrared face recognition based on wavelet transform and 2DPCA, in: Proc. 2010 International Conference on Intelligent Computing and Integrated Systems, IEEE, Guilin, China, 2010, pp. 359–362.
[21] S. Mallat, Wavelets for a vision, Proc. IEEE 84 (4) (1996) 604–614.
[22] D.-Q. Dai, H. Yan, S. Yat-Sen, Wavelets and face recognition, in: K. Delac, M. Girgic (Eds.), Face Recognition, I-Tech, Vienna, Austria, 2007, pp. 59–74.
[23] D. Li, X. Tang, W. Pedrycz, Face recognition using decimated redundant discrete wavelet transforms, Mach. Vis. Appl. 23 (2) (2012) 391–401.
[24] B. Zhang, L. Zhang, D. Zhang, L. Shen, Directional binary code with application to PolyU near-infrared face database, Pattern Recognit. Lett. 31 (14) (2010) 2337–2344.
[25] H. Maeng, S. Liao, D. Kang, et al., Nighttime face recognition at long distance: cross-distance and cross-spectral matching, in: Proc. ACCV, Daejeon, Korea, 2012, pp. 1–14.
[26] Y. Zheng, Near infrared face recognition using orientation-based face patterns, in: Proc. International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE Fraunhofer IGD, Darmstadt, Germany, 2012, pp. 1–4.
[27] H. Maeng, H.-C. Choi, U. Park, et al., NFRAD: near-infrared face recognition at a distance, in: Proc. 2011 International Joint Conference on Biometrics (IJCB), IEEE, Washington D.C., 2011, pp. 1–7.
[28] S. Farokhi, S.M. Shamsuddin, J. Flusser, et al., Rotation and noise invariant near-infrared face recognition by means of Zernike moments and spectral regression discriminant analysis, J. Electron. Imaging 22 (1) (2013) 1–11.
[29] D. Graham, Perceiving and Remembering Faces, Academic Press, New York, 1981.
[30] H. Ellis, M. Jeeves, F. Newcombe, A. Young, Aspects of Face Processing, Springer/Nijhoof, Dordrecht, 1986.
[31] R.S. Ghiass, O. Arandjelovic, H. Bendada, X. Maldague, Infrared face recognition: a literature review, in: Proceedings of the International Joint Conference on Neural Networks, IEEE, Dallas, TX, 2013, pp. 1–10.
[32] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1999.
[33] M.H. Rohban, H.R. Rabiee, M. Khansari, Face virtual pose generation using multi resolution subspaces, in: Proc. International Symposium on Telecommunications, IEEE, Tehran, Iran, 2008, pp. 629–633.
[34] D. Li, H. Luo, Z. Shi, Redundant DWT based translation invariant wavelet feature extraction for face recognition, in: Proc. 19th International Conference on Pattern Recognition (ICPR), IEEE, Tampa, FL, USA, 2008, pp. 1–4.
[35] M. Khansari, H.R. Rabiee, M. Asadi, M. Ghanbari, Object tracking in crowded video scenes based on the undecimated wavelet features and texture analysis, EURASIP J. Adv. Signal Process. 2008 (1) (2008) 1–18.
[36] J. Neumann, G. Steidl, Dual-tree complex wavelet transform in the frequency domain and an application to signal classification, Int. J. Wavelets Multiresolut. Inf. Process. 8 (5) (2003) 1–26.
[37] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multiscale transforms, IEEE Trans. Inf. Theory 38 (2) (1992) 587–607.
[38] J.-L. Starck, F. Murtagh, J.M. Fadili, Sparse Image and Signal Processing, Cambridge University Press, New York, 2010.
[39] J. Flusser, T. Suk, B. Zitova, Moments and Moment Invariants in Pattern Recognition, Wiley, Chichester, 2009.
[40] M.R. Teague, Image analysis via the general theory of moments, J. Opt. Soc. Amer. 70 (8) (1980) 920–930.
[41] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, Human face recognition using different moment invariants: a comparative study, in: Proc. Congress on Image and Signal Processing (CISP'08), IEEE, Sanya, Hainan, China, 2008, pp. 661–666.
[42] C. Singh, N. Mittal, E. Walia, Face recognition using Zernike and complex Zernike moment features, Pattern Recognit. Image Anal. 21 (1) (2011) 71–81.
[43] D. Cai, X. He, J. Han, SRDA: an efficient algorithm for large-scale discriminant analysis, IEEE Trans. Knowl. Data Eng. 20 (1) (2008) 1–12.
[44] H. Kemal Ekenel, B. Sankur, Multiresolution face recognition, Image Vis. Comput. 23 (5) (2005) 469–477.
[45] J. Ye, T. Wang, Regularized discriminant analysis for high dimensional, low sample size data, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, Philadelphia, PA, USA, 2006, pp. 454–463.
[46] J. Haddadnia, K. Faez, M. Ahmadi, An efficient human face recognition system using pseudo Zernike moment invariant and radial basis function neural network, Int. J. Pattern Recognit. Artif. Intell. 17 (1) (2003) 41–62.
[47] D. Cai, X. He, J. Han, H.J. Zhang, Orthogonal laplacianfaces for face recognition, IEEE Trans. Image Process. 15 (11) (2006) 3608–3614.
[48] I. Daubechies, The wavelet transform, time–frequency localization and signal analysis, IEEE Trans. Inf. Theory 36 (5) (1990) 961–1005.
[49] C.-C.L. Liu, D.-Q.D. Dai, H.Y. Yan, Local discriminant wavelet packet coordinates for face recognition, J. Mach. Learn. Res. 8 (1) (2007) 1165–1195.
[50] J.R. Beveridge, G.H. Givens, P.J. Phillips, B.A. Draper, Factors that influence algorithm performance in the face recognition grand challenge, Comput. Vis. Image Underst. 113 (6) (2009) 750–762.
[51] S. Du, R. Ward, Face recognition under pose variations, J. Franklin Inst. 343 (6) (2006) 596–613.
[52] S.K. Zhou, R. Chellappa, Image-based face recognition under illumination and pose variations, J. Opt. Soc. Amer. A 22 (2) (2005) 217–229.
[53] M. Khansari, H. Rabiee, M. Asadi, et al., A quantization noise robust object's shape prediction algorithm, in: Proceedings of the 13th European Signal Processing Conference (EUSIPCO'05), Antalya, Turkey, 2005, pp. 1–4, Citeseer.
[54] L. Chen, L. Zhang, L. Zhu, et al., A novel facial feature point localization algorithm using probabilistic-like output, in: Proceedings of the Asian Conference on Computer Vision, Jeju Island, Korea, 2004, pp. 1–6.
[55] S. Shan, Y. Chang, W. Gao, B. Cao, Curse of misalignment in face recognition: problem and a novel misalignment learning solution, in: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, Korea, 2004, pp. 314–320.
[56] H. Wang, S. Yan, T. Huang, et al., Misalignment robust face recognition, IEEE Trans. Image Process. 19 (4) (2010) 1087–1096.

Sajad Farokhi received his PhD degree (2014) in computer science from Universiti Teknologi Malaysia. His research areas include computer vision, machine learning, pattern recognition, biometrics and image processing.

Siti Mariyam Shamsuddin received a BSc and MSc in mathematics from New Jersey and a PhD in pattern recognition and artificial intelligence from Universiti Putra Malaysia, Malaysia. Currently, she is head of Soft Computing Research Group and a full professor of computer science at Universiti Teknologi Malaysia, Johor, Malaysia. Her research interests include soft computing and its applications, pattern recognition, and geometric modeling.

Usman Ullah Sheikh received his BEng degree (2003) in electrical and mechatronics engineering, the MEng degree (2005) in telecommunications engineering and PhD degree (2009) in image processing and computer vision from Universiti Teknologi Malaysia. His research work is mainly on computer vision and embedded systems design. He is currently a senior lecturer at Universiti Teknologi Malaysia working on image processing for intelligent surveillance systems. He is a member of IET and IEEE and has published works in many IEEE conferences.

Jan Flusser received the M.Sc. degree in mathematical engineering from the Faculty of Nuclear Science and Physical Engineering, Czech Technical University, Prague, in 1985, the Ph.D degree in computer science from the Czechoslovak Academy of Sciences in 1990, and the DrSc. degree in technical cybernetics from the Czech Technical University in 2001. He is the director of the Institute of Information Theory and Automation of the ASCR, Prague, Czech Republic, and a full professor of Computer Science at the Czech Technical University, Prague, and at the Charles University, Prague. He has authored and coauthored more than 200 research publications in these areas, including the book Moments and Moment Invariants in Pattern Recognition (Wiley, 2009), 50 papers in prestigious journals and 20 tutorials at international conferences (ICIP'05, ICIP'07, EUSIPCO'07, CVPR'08, FUSION'08, SPPRA'09, SCIA'09, ICIP'09, SPPRA'10 and CGIM'10) and invited/keynote talks (ICCS'06, COMPSTAT'06, WIO'06, DICTA'07, CGIM'08). His publications have received about 5000 citations. His main areas are moments and moment invariants, image registration, multichannel blind deconvolution, and super-resolution imaging.

Mohammad Khansari received his B.S, M.S and PhD degrees in Computer Engineering all from Sharif University of Technology, Tehran, Iran, in 1996, 1998 and 2008 respectively. He is the co-author of four books with Prof. Rabiee in Free/Open Source Software topics and has more than forty papers in international conferences and journals. His main research interests are network science and complex networks, multimedia and healthcare sensor networks, multimedia delivery over peer-to-peer networks, Free/Open Source Software specially localization, customization, mining of software repositories and diffusion. Currently, he is the faculty member of Faculty of New Sciences and Technologies, University of Tehran and the director of IT and Cyberspace Center of University of Tehran.

Kourosh Jafari-Khouzani received his PhD degree in Computer Science from Wayne State University in 2006. He is a research fellow in Athinoula A. Martinos Center for Biomedical Imaging at the Department of Radiology, Massachusetts General Hospital, Boston, MA, USA. His main area of research interest is medical image analysis and neuroimaging.