

Near infrared face recognition using Zernike moments and Hermite kernels



Sajad Farokhi^{a,b}, Usman Ullah Sheikh^{a,*}, Jan Flusser^b, Bo Yang^b

^a Digital Signal and Image Processing Research Group, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

^b Institute of Information Theory and Automation, Academy of Science of the Czech Republic, 18208 Prague 8, Czech Republic

ARTICLE INFO

Article history:

Received 21 June 2014

Received in revised form 7 April 2015

Accepted 11 April 2015

Available online 21 April 2015

Keywords:

Face recognition
Near infrared
Zernike moments
Hermite kernel
Decision fusion

ABSTRACT

This work proposes a novel face recognition method based on Zernike moments (ZMs) and Hermite kernels (HKs) to cope with variations in facial expression, changes in head pose and scale, occlusions due to wearing eyeglasses and the effects of time lapse. Near infrared images are used to tackle the impact of illumination changes on face recognition, and a combination of global and local features is utilized in the decision fusion step. In the global part, ZMs are used as a feature extractor and in the local part, the images are partitioned into multiple patches and filtered patch-wise with HKs. Finally, principal component analysis followed by linear discriminant analysis is applied to data vectors to generate salient features and decision fusion is applied on the feature vectors to properly combine both global and local features. Experimental results on CASIA NIR and PolyU NIR face databases clearly show that the proposed method achieves significantly higher face recognition accuracy compared with existing methods.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

With the rapid development of biometric technology, face recognition (FR) has become an active research area in the field of computer vision. Despite research efforts, FR remains a challenge principally because of large intra-class variations stemming from illumination conditions, facial expressions, wearing eyeglasses, head pose, scale and the effects of time lapse. In the last two decades, numerous techniques have been proposed to solve these challenges and to develop a more effective FR system [1,27,41,54,58,63,65,66]. Comprehensive reviews on recent FR methods were presented in [67,42]. Among the many issues in visible FR systems, variation of illumination is regarded as the most challenging for subject identification in cooperative, as well as non-cooperative, user scenarios [2,28,56,59]. Several illumination invariant FR methods have been proposed, including methods based on three-dimensional (3D) shapes of the face, as well as methods based on thermal images measuring body temperature [3,31,37,38,46,50]. However, 3D techniques are costly and require high computational complexity, while thermal images are extremely sensitive to environmental temperature, health conditions, perspiration and are opaque to eyeglasses [4,7,53]. Other methods to compensate for illumination problems have also been introduced, a typical example being presented in [52]. However, the solution to the illumination problem in FR is still not perfect. Recently, illumination variations have been addressed utilizing near infrared (NIR) imagery [10,24,34].

* Corresponding author. Tel.: +60 197055336.

E-mail addresses: fsajad2@live.utm.my (S. Farokhi), usman@fke.utm.my (U.U. Sheikh), flusser@utia.cas.cz (J. Flusser), yang@utia.cas.cz (B. Yang).

Several NIR FR methods have been proposed to achieve an accurate FR system. Li et al. based their advanced method on a local binary pattern (LBP) and statistical learning algorithms, achieving substantial improvement in their results [33]. However, their method suffered from the high sensitivity of LBP to noise and minor pose variations and alignment errors [9,13,25,35].

Zhang et al. utilized Gabor filtering to enhance image features, and subsequently extracted discriminative features using a directional binary code (DBC). DBC captures more spatial information than LBP, and thus gives greater recognition accuracy [64]. The authors were successful in proposing an accurate FR system. However, more challenges, such as wearing eye-glasses, still remain.

To improve recognition performance, both NIR and visible images and other types of modality, such as voice and fingerprints, have been attempted [57]. Shen et al. proposed a method based on boosted DBC features using both visible and NIR images [43]. DBC and the AdaBoost algorithm were used for feature extraction and classification, respectively, while the performance was improved by consequent information fusion on a decision level. Other similar approaches can be found in [20,51,55,61]. However, these remain beyond the scope of this paper.

Nighttime FR at long distance was recently investigated by Maeng et al. [36]. Their preprocessing step was based on the difference of Gaussians, then scale invariant feature transform (SIFT) and multi-scale local binary pattern (MLBP) were used for feature extraction. Their results indicated that SIFT has better accuracy compared to MLBP. Other improvements in this domain have also been presented in [26,36,40].

A broad review of infrared FR methods can be found in [18,19]. However, as reported by Ghiass et al., all papers are limited by failing to examine the performance of the proposed method in the simultaneous presence of all considered challenges. Most related works in the NIR domain have focused solely on the illumination problem, with scant attention paid to facial expression, wearing eyeglasses, variations in head pose and scale and the effects of time lapse. These factors are all known to introduce crucial problems in FR systems [6].

Farokhi et al. [12,13] systematically studied the possibility of NIR FR in the presence of the above mentioned challenges. In [12], they proposed a method based on Zernike moments (ZMs) to cope with face rotation and noise. Although their results were an improvement on those achieved by LBP [33], the authors concluded that an accurate system could not be based on single-type features. Instead, it should utilize a combination of both global and local features (fusion) to make the final decision. A similar conclusion was drawn in [15], also combining the global and local facial features presented in [32,62], and outperformed the individual use of these features. Farokhi et al. used this principle in NIR FR [13] in which ZMs were used to extract global features while undecimated discrete wavelet transform (UDWT) was used to extract local features. Their system achieved good recognition performance. However, UDWT is inefficient in terms of memory usage and computational time.

In this work, we follow this successful approach and replace UDWT by other local features that provide at least as good a discrimination power as UDWT while decreasing the computational time. We propose Hermite kernels (HKs) as filters to extract local features and ZMs as global features. Presented results of extensive experiments on the CASIA NIR and the PolyU NIR face databases show that our method achieves a higher recognition rate, compared to existing works in the presence of the most common challenges (face image modification). Our method is capable of overcoming and improving many of the shortcomings of the existing state-of-the-art techniques.

The paper is organized as follows: In Sections 2 and 3, brief reviews of ZMs and HKs are given. The proposed method is then described in Section 4. Experimental results and performance analysis are then presented in Section 5, and finally the paper is concluded in Section 6.

2. Zernike moments

ZMs were first introduced by Teague in the early 1980s, and have been applied in many research works [16,23,45,47]. ZMs belong to the family of so-called *radial moments*, whose basis functions are, in polar coordinates, products of a 1D polynomial in the radial direction and a harmonic function in the angular direction. As follows from the Fourier Shift Theorem, radial moments change under image rotation only in their phase, while the amplitude remains constant (the same is true also for quaternion moments, see [21]). This favorable behavior makes ZMs useful in coping with image rotations. On the contrary, the moments orthogonal on a rectangle, such as Legendre, Chebyshev, Gegenbauer and Krawtchouk [16,22], change under rotation in a much more complicated manner. This renders them quite difficult to utilize for constructing invariants. The ZMs have shown good performance in FR systems in the presence of facial expression, image rotation, and noise [11,12,14,30,44]. However, they cannot handle partial occlusions properly due to their global nature. The ZM of order p with repetition q of a function $f(r, \theta)$, where (r, θ) are polar coordinates, is defined by the following equation:

$$Z_{pq} = \frac{p+1}{\pi} \int_{\theta=0}^{2\pi} \int_{r=0}^1 V_{pq}^*(r, \theta) f(r, \theta) r dr d\theta, \quad |r| \leq 1, \quad (1)$$

where the symbol “*” denotes a complex conjugate and V_{pq} denotes a Zernike polynomial of order p and repetition q , which can be written as follows:

$$V_{pq}(r, \theta) = R_{pq}(r) e^{iq\theta}. \quad (2)$$

R_{pq} is a real-valued radial polynomial of the form

$$R_{pq}(r) = \sum_{k=0}^{\frac{p-|q|}{2}} (-1)^k \frac{(p-k)!}{k! \left(\frac{p+|q|}{2} - k\right)! \left(\frac{p-|q|}{2} - k\right)!} r^{p-2k}, \quad (3)$$

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

3. Hermite kernels

The Hermite kernel is a type of kernel function that has been used widely in image processing and computer vision [17,19]. Mathematically, it is a product of two 1D-scaled Hermite polynomials and rotationally symmetric Gaussian functions. The formula describing the HK in 2D space is as follows:

$$B_{p,q}(x, y, \sigma) = H_p\left(\frac{x}{\sigma}\right) H_q\left(\frac{y}{\sigma}\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (4)$$

where σ is the standard deviation of a Gaussian function and $H_p(x)$ is the Hermite polynomial that can be expressed as follows:

$$H_p(x) = \sum_{k=0}^{\lfloor p/2 \rfloor} \frac{(-1)^k p!}{k! (p-2k)!} (2x)^{p-2k}. \quad (5)$$

The HKs up to order 4 are shown in Fig. 1.

HKs can be applied on an image as a filter to generate directional features. As shown in Fig. 1, the application of basis functions of order two is similar to that of horizontal, diagonal and vertical high-pass filters. To generate directional features, we apply the basis functions together as filters of r by r size on an image $f(x, y)$. We then obtain the effective directional features by convolving the image with the following operator:

$$H_B(x, y, \sigma, f) = [B_{1,1}(x, y, \sigma) * f(x, y), B_{2,0}(x, y, \sigma) * f(x, y), B_{0,2}(x, y, \sigma) * f(x, y)]. \quad (6)$$

The filtered face images are shown in Fig. 2.

As can be seen from Fig. 2, the proposed operator is competent in encoding rich edge information and generating directional features of the facial images. This simple scheme can be extended by adding more filters of higher orders and/or multiple orientations of the basis filters. In such a case, the scheme benefits from the property of steerability of the HK [60], which means that a filter of arbitrary orientation is synthesized as a linear combination of a set of basis filters. The image filtered by an arbitrary oriented filter becomes only a linear combination of images filtered by basic filters, thus facilitating very efficient implementation of the filtering operation.

4. Proposed face recognition system using ZMs and HKs (ZMHK)

The proposed FR system is composed of several steps, as shown in Fig. 3. In this section, feature extraction and decision fusion strategy are explained in detail.

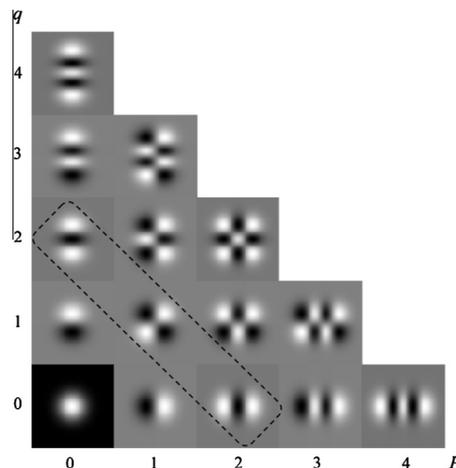


Fig. 1. Hermite kernels ($B_{p,q}$) up to order 4. The basis functions for the kernel of order 2 are marked by the dashed rectangle.

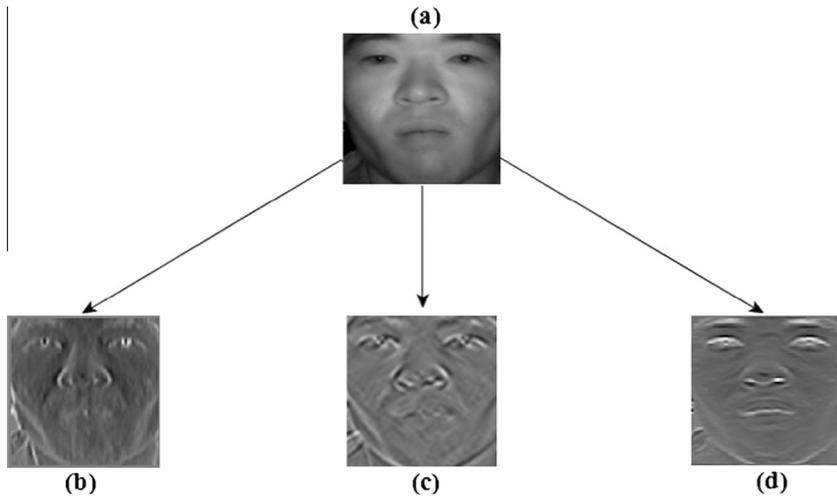


Fig. 2. (a) Original image. (b)–(d) The results using Hermite kernels $B_{2,0}$, $B_{1,1}$, and $B_{0,2}$, respectively ($r = 13$, $\sigma = 2$).

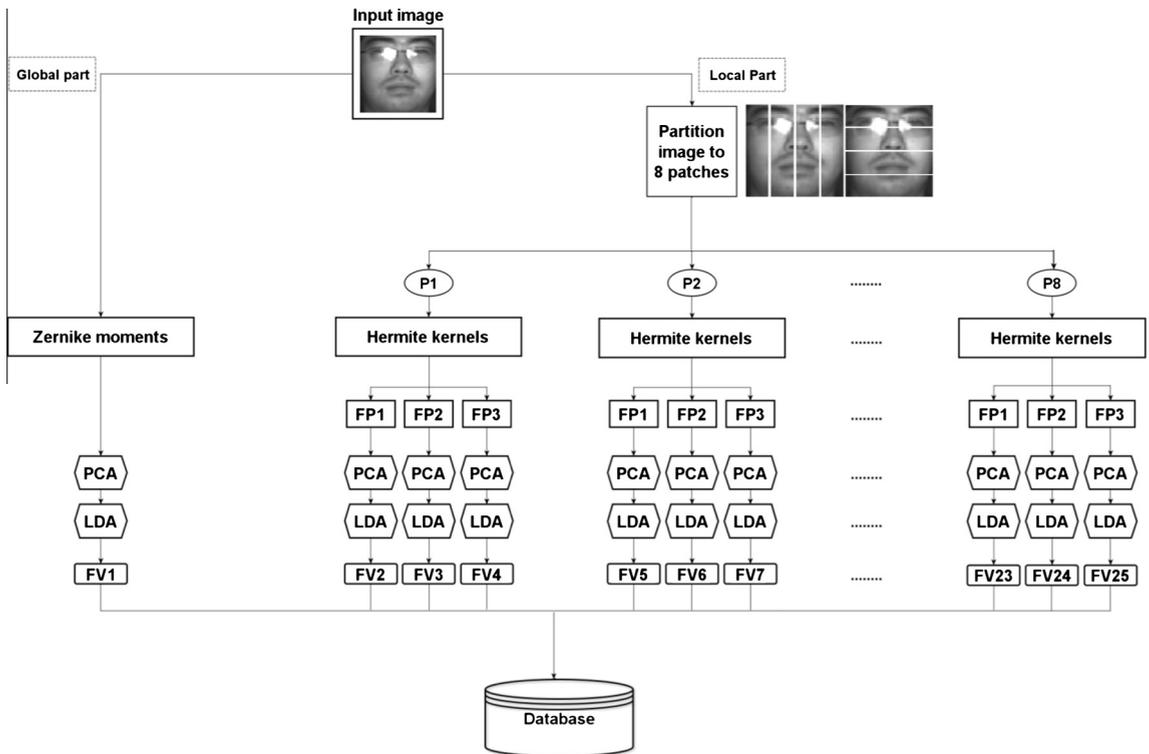


Fig. 3. Schematic of the proposed face recognition system (training phase).

4.1. Feature extraction

Fig. 3 illustrates the proposed feature extraction, separated into global and local feature extractions. In the global part, salient facial features are represented by ZMs, while local facial features are calculated from the response of the HKs on the image patches. Dimension reduction techniques are applied on both the extracted global and local features. Prior to feature calculation, the image is preprocessed and normalized, as discussed in Section 5.1.

4.1.1. Global feature extraction using ZMs

To extract global features, we first calculate the ZMs for a normalized image up to order 10 as clarified in our previous work [13]. Since ZMs are in general complex valued, both imaginary and real parts of the ZMs are incorporated into a data

vector. The ZMs, up to order 10 and their dimensionality (i.e., the number of independent components), are shown in Table 1. In calculating the dimensionality, $Z_{p,q} = (Z_{p,-q})^*$ and $Z_{p,p}$ are always real. The dimension of the data is the sum of all the ZMs generated from order 0 to 10, resulting in a dimension of 66. Therefore the M images in our training set result in a data matrix size of $M \times 66$.

In the second step, principal component analysis (PCA) [48], applied to the whole data matrix of $M \times 66$ size, is used to de-correlate the features. New features are found as projections onto the eigenvectors of the covariance matrix. All the features are maintained, and the PCA applies no dimension reduction. Finally, the de-correlated features serve as an input into linear discriminant analysis (LDA) [5] to enhance their discrimination power. Unlike PCA, LDA is a supervised method which maximizes the separability between the individual classes (faces) in the training set. This separability is measured by means of intra-class and inter-class scatter matrices. More sophisticated dimensionality reduction methods exist; such as that proposed in [68] (references to many others can be found herein) which uses regularized optimization coupled with a sparsity constraint. These methods usually consider each pixel as a single feature, and are applied directly on such data. However, since we inserted the calculation of ZMs and HKs between the image acquisition and dimensionality reduction steps, the use of a simple PCA + LDA scheme is sufficient.

4.1.2. Local feature extraction using HKs

The process of local feature extraction is more extensive than global feature extraction. In the first step, the normalized images are decomposed into four horizontal (16×64 pixels) and four vertical (64×16 pixels) patches, as in Fig. 3.

For the second step, HKs are applied on each patch and the values of the filtered patches are used as data vectors. For each patch, three filtered patches (FP) are produced (we use three second-order HKs). Hence $8 \times 3 = 24$ filtered patches are generated in this phase. Thus, for M images in our training set, 24 data matrices of size $M \times 1024$ are generated. In the third step, PCA is applied to each data matrix of size $M \times 1024$ to de-correlate the features. Since 24 data matrices are utilized in the training step, 24 principal eigenvectors related to each data matrix are learned in this stage, and the size of each data matrix is reduced to $M \times M - 1$. In the final step, the transformed data are used as input to the LDA in order to decrease the dimension and to enhance the discrimination power of the system. Finally, 24 data matrices of size $M \times 66$ are obtained and used as local features.

4.2. Proposed decision fusion

As explained above, we generated 25 feature vectors for each image (24 local feature vectors and one global feature vector). Originally, local and global feature vectors are of different size (dimensionality), but PCA + LDA reduction transforms all vectors to the same size (depending on the number of persons in the database). We could use these 25 vectors as inputs for 25 independent minimum-distance classifiers, and then make a decision by majority vote. This obvious approach, however, does not take into account possible different confidence levels of the individual classifiers. To overcome this drawback, we propose another fusion strategy, as described below.

Initially, the distance between the test face feature vector and feature vectors of the individuals in the database is calculated for each class. Then, assuming there are c classes (persons) with m_i samples per class ($i = 1, 2, \dots, c$), the minimum distances between each feature vector of the test image and the feature vectors of the database images are calculated using the following formula:

$$D_{v,i} = \min_{1 \leq p \leq m_i} d(F_v(\text{test}), F_{v,i,p}(\text{database})) \quad (7)$$

$$v = 1, \dots, 25, \quad i = 1, \dots, c,$$

where $F_v(\text{test})$ is the v th feature vector related to the test images, $F_{v,i,p}(\text{database})$ is the v th feature vector of the p th sample in the i th class, and $d(\cdot)$ stands for a distance measure (we use Euclidean distance) between two feature vectors. Hence, the distance matrix

Table 1
Zernike moments (up to order 10) and their dimensionality.

Order	Zernike moments ($Z_{p,q}$)	Dimensionality of the specified order
0	$Z_{0,0}$	1
1	$Z_{1,1}, Z_{1,-1}$	2
2	$Z_{2,0}, Z_{2,2}, Z_{2,-2}$	3
3	$Z_{3,1}, Z_{3,-1}, Z_{3,3}, Z_{3,-3}$	4
4	$Z_{4,0}, Z_{4,2}, Z_{4,-2}, Z_{4,4}, Z_{4,-4}$	5
5	$Z_{5,1}, Z_{5,-1}, Z_{5,3}, Z_{5,-3}, Z_{5,5}, Z_{5,-5}$	6
6	$Z_{6,0}, Z_{6,2}, Z_{6,-2}, Z_{6,4}, Z_{6,-4}, Z_{6,6}, Z_{6,-6}$	7
7	$Z_{7,1}, Z_{7,-1}, Z_{7,3}, Z_{7,-3}, Z_{7,5}, Z_{7,-5}, Z_{7,7}, Z_{7,-7}$	8
8	$Z_{8,0}, Z_{8,2}, Z_{8,-2}, Z_{8,4}, Z_{8,-4}, Z_{8,6}, Z_{8,-6}, Z_{8,8}, Z_{8,-8}$	9
9	$Z_{9,1}, Z_{9,-1}, Z_{9,3}, Z_{9,-3}, Z_{9,5}, Z_{9,-5}, Z_{9,7}, Z_{9,-7}, Z_{9,9}, Z_{9,-9}$	10
10	$Z_{10,0}, Z_{10,2}, Z_{10,-2}, Z_{10,4}, Z_{10,-4}, Z_{10,6}, Z_{10,-6}, Z_{10,8}, Z_{10,-8}, Z_{10,10}, Z_{10,-10}$	11

$$\mathbf{D}_1 = \begin{bmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,c} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,c} \\ \vdots & \vdots & \cdots & \vdots \\ D_{25,1} & D_{25,2} & \cdots & D_{25,c} \end{bmatrix}_{25 \times c} \quad (8)$$

shows, in each of its rows, the distances of the given test image feature vector to all database classes.

The proposed fusion is a two-stage process. First, we combine all local features, and then we combine the result with the global features. The fusion is based on the distance matrix (8), and is performed “column-wise.” This means we are trying to find the “fused distance” between the test image and each class.

We define the weight λ_v of the v th local feature vector as $\lambda_v = \frac{D'_{v,1}}{D'_{v,2}}$ where $D'_{v,1}$ and $D'_{v,2}$ are the first and second minimum distances over all c classes, respectively, i.e., two minimal numbers in the v th row of the distance matrix (8). The weight λ_v (which is always non-negative and less or equal to one) reflects the confidence we have in the v th local feature vector, the lower the λ_v , the higher the confidence level. For each class, we calculate a weighted average

$$D_{L,i} = \frac{1}{A} \sum_{v=1}^{24} \lambda_v D_{v,i}, \quad i = 1 : c, \quad (9)$$

where $A = \sum_{v=1}^{24} \lambda_v$. Hence, we obtain a new distance matrix \mathbf{D}_2

$$\mathbf{D}_2 = \begin{bmatrix} D_{L,1} & D_{L,2} & \cdots & D_{L,c} \\ D_{G,1} & D_{G,2} & \cdots & D_{G,c} \end{bmatrix}_{2 \times c}, \quad (10)$$

where the first row is obtained from Eq. (9), and the second row is equal to the last row of \mathbf{D}_1 , i.e., $D_{G,i} = D_{25,i}$.

Now we can fuse the local and global features together using basically the same algorithm as for the fusion of the local features. We calculate a new vector $D_{E,i}$ as

$$D_{E,i} = \lambda_L D_{L,i} + \lambda_G D_{G,i}, \quad i = 1 : c, \quad (11)$$

where λ_L and λ_G are the weights of $D_{L,i}$ and $D_{G,i}$ respectively, defined row-wise as a ratio of the first and second minimum in \mathbf{D}_2 ,

$$\lambda_L = \frac{D'_{L,1}}{D'_{L,2}} \quad (12)$$

$$\lambda_G = \frac{D'_{G,1}}{D'_{G,2}}. \quad (13)$$

Doing so, we end up with a vector \mathbf{D}_3

$$\mathbf{D}_3 = [D_{E,1}, D_{E,2}, \cdots, D_{E,c}]_{1 \times c}, \quad (14)$$

which is immediately used for the decision. The test image is assigned to the class with minimum $D_{E,i}$ ($i = 1 : c$), and the label of this class is the output of the FR system.

5. Experimental results and performance analysis

In this section, we investigate the performance of the proposed method using CASIA NIR and PolyU NIR face databases. We also compare our method with several other popular FR methods that also use directional features. This comparative study is carried out against the following FR schemes:

- Gabor Wavelet + Fisher LDA, which is called Gabor Fisher Classifier (GFC) [8].
- LBP + LDA (LBPL) [33].
- Gabor Wavelet + DBC (GDBC) [64].
- Wavelet Scattering (WS) [29].
- ZMs + UDWT (ZMUDWT) [13].

Descriptions of the settings used for performance evaluation are explained in Table 2. In all experiments, we applied our method in three ways: global features based on ZMs only (denoted as ZM), local HK features only (denoted as HK) and both global and local parts together (denoted as ZMHK). The global part is basically identical (except for the PCA + LDA reduction) to the popular methods described in [39,44].

In the first part of this section, we briefly describe the database and preprocessing. Then we explain the experiments carried out to evaluate and compare the performance of different methods.

Table 2

Experimental and parameter set-up of the benchmark methods used in the performance evaluation.

Method	Specification
GFC [8]	First Gabor wavelets of five scales and eight orientations are used to extract features. Then Fisher linear discriminant analysis is performed on these features.
LBPL [33]	The image is first divided into 64, 8×8 blocks and then a local binary pattern histogram is calculated for each block. Finally, linear discriminant analysis is used to decrease the dimension of features.
GDDB [64]	First an image is divided into four 32×32 blocks. Gabor wavelets of five scales and eight orientations are then applied on the image blocks. Next, directional binary code is applied along 0° , 45° , 90° and 135° directions to produce features.
WS [29]	The order of wavelet scattering is 2. The numbers of scales and orientations are 3 and 6, respectively.
ZMUDWT [13]	Zernike moments are used as global features, whereas local features are generated by undecimated discrete wavelet transform. Recognition is done by decision fusion. The order of Zernike moments is 10. The decomposition level for performing undecimated discrete wavelet transform is 3, and the wavelet basis is "Db 3".
ZMHK (Proposed method)	The selected order of Zernike moments is 10. The values of r and σ are empirically set to 13 and 2 due to the best performance of systems using these parameters. The Euclidean criterion is used for implementing proposed decision fusion.

5.1. Preprocessing and database

The face images of the CASIA NIR database [33] and PolyU NIR face database [49] (Fig. 4(a)) are used in our experiments. From both databases, a total of only 70 and 100 subjects were selected, respectively – excluding those which did not present special challenges. As a result, the sizes of the gallery set and probe set for the CASIA NIR database are 350 (five images per subject) and 490 (seven images per subject), respectively. The sizes of the gallery set and probe set for the PolyU NIR face database are 600 (six images per subject) and 800 (eight images per subject), respectively. For both the CASIA NIR and PolyU NIR face databases, the gallery set includes normal images without any challenge. The probe set for the CASIA NIR database contains images with facial expressions, wearing eyeglasses and a moderate head pose, while the probe set for the PolyU NIR face database includes NIR images with a sharp head pose, scale variations and time lapse. There is no overlap between the gallery set and the probe set. The database specifications are described in Table 3. The flow of the preprocessing is as follows:

1. Face images are aligned by placing the eyes at a fixed position (Fig. 4(b)).
2. Face images are cropped to remove hair and background (Fig. 4(c)).
3. Each image is resized to 64×64 with 256 gray levels to decrease the computational time (Fig. 4(d)).

5.2. Experimental results on CASIA NIR database

In the first experiment, we used the CASIA NIR database, which includes normal images, images with facial expression, wearing eyeglasses and a moderate head pose. To create our gallery set (training set), we selected a random subset of 70

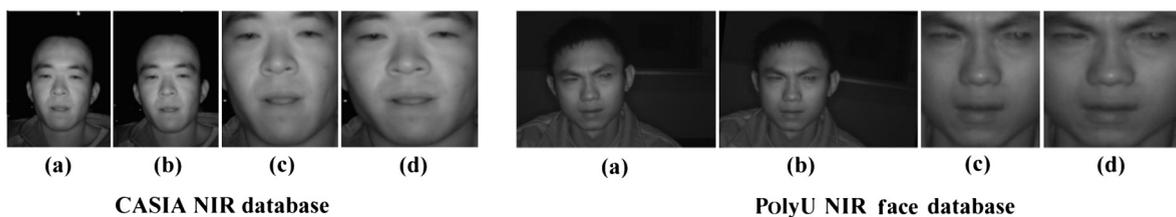


Fig. 4. Proposed preprocessing method for CASIA NIR and PolyU NIR face databases. (a) Input image, (b) normalized image with fixed eyes position, (c) cropped image, and (d) size and gray-level normalization.

Table 3

Summary of the CASIA NIR and PolyU NIR face databases.

	Database	
	CASIA NIR	PolyU NIR Face
Infra-red wavelength	850 nm	850 nm
No. of subjects	197	335
Number of still images per subject	20	100
Distance	50 cm and 100 cm	80 cm and 120 cm
Resolution	640×480	768×576
Format	BMP	JPG

persons with three normal images (without any challenge). For the probe set, three random images with different variations were chosen (Fig. 5) while employing a closed universe assumption, i.e., each probe image should have a corresponding match in the gallery. Then, the average recognition rate with over 10 random splits was calculated. Assuming that the success rates are normally distributed, we also calculated confidence intervals, which are one of the most useful criteria for evaluating result reliability. A narrower confidence interval indicates higher stability of a particular method. The resulting mean, standard deviation and confidence interval with 95% significance level are shown in Table 4 and Fig. 6. The following conclusions are made.

As is apparent from Fig. 6 and Table 4, the performance of the proposed method is better compared to the remaining tested methods. There is no overlap between the confidence intervals of the proposed method and those of the other methods, except for ZMUDWT. The proposed method performs statistically significantly better than the other methods, since it is based on both global and local features along with a sophisticated fusion rule, whereas other methods, except for ZMUDWT, are based solely on local features. Comparing GDBC and LBPL; the LBPL method achieves better performance than GDBC, contradicting the findings of previous research [64] which showed that GDBC performs better than LBP. The main reason for this discrepancy is that LDA was not used for dimension reduction in GDBC; whereas LBPL is based on a combination of LBP and LDA techniques which ensures that the resulting features are more salient than the raw features in GDBC.

5.3. Experimental results on the PolyU NIR face database

In the second experiment, the PolyU NIR face database was used. While the gallery set includes frontal face images, the probe set includes images with scale changes, pose variations and time lapse (see Fig. 7 for some examples). One hundred subjects were used in this experiment, otherwise the configuration would be the same as in the previous experiment.

Comparison of results obtained by the different methods presented in Table 5 and Fig. 8 show that our method again performs better than the benchmark methods, and both global and local features are mutually complementary and can handle image variations effectively. Further analysis shows that the recognition rate of the proposed method is 6% better than that of LBPL, since we use images with head pose variations which decreases the accuracy of LBP. This experiment highlights the sensitivity of LBP to head pose, which has been reported previously [9,13].

In this experiment, the ZMHK method achieves an average accuracy of 87%, compared to 91% in the previous experiment using CASIA NIR database. This is due to the presence of more pose variations in the PolyU NIR face database compared to CASIA NIR database.

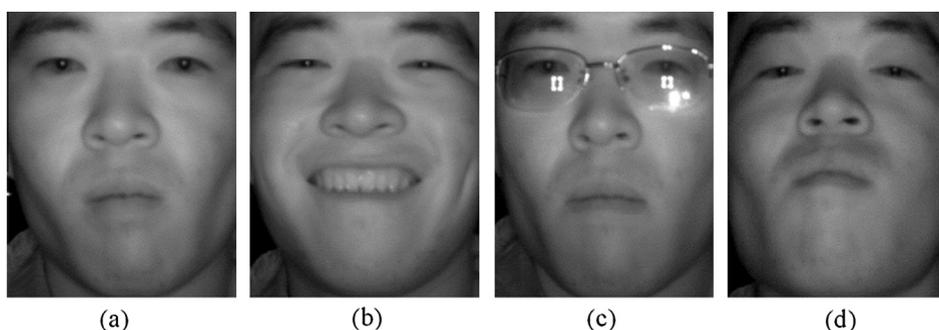


Fig. 5. (a) Sample of a normal image used as a gallery image. Sample of images with (b) facial expression, (c) wearing eyeglasses, and (d) variation in head pose used as probe images.

Table 4

Accuracy of different methods in the presence of different challenges based on the CASIA NIR database.

Method	Mean \pm standard deviation	Confidence interval
GFC	80.26 \pm 3.61	[78.02, 82.49]
LBPL	86.75 \pm 2.93	[84.93, 88.56]
GDBC	80.14 \pm 4.39	[77.41, 82.86]
WS	82.79 \pm 2.31	[81.35, 84.22]
ZMUDWT	90.25 \pm 1.21	[89.50, 91.00]
ZM (Global part)	78.29 \pm 3.32	[76.23, 80.34]
HK (Local part)	87.86 \pm 2.46	[86.33, 89.38]
ZMHK (Proposed method)	91.47 \pm 2.13	[90.15, 92.79]

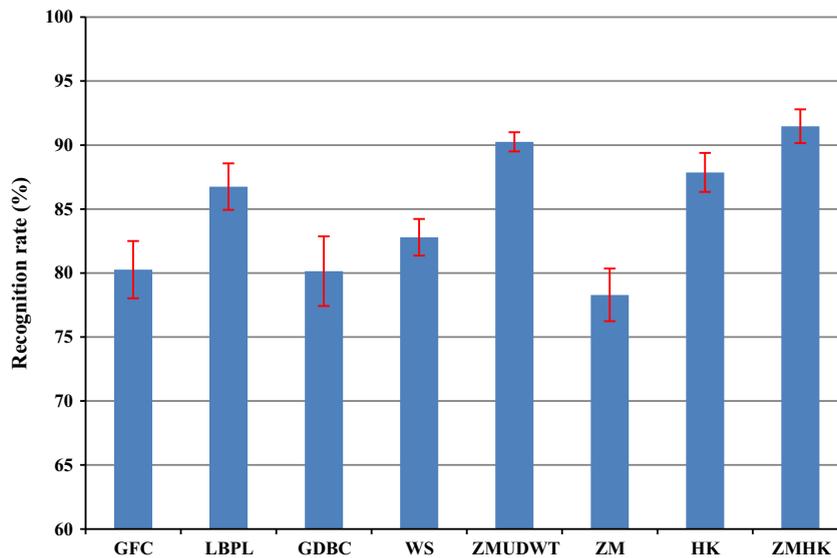


Fig. 6. Recognition results of different methods for the CASIA NIR database.

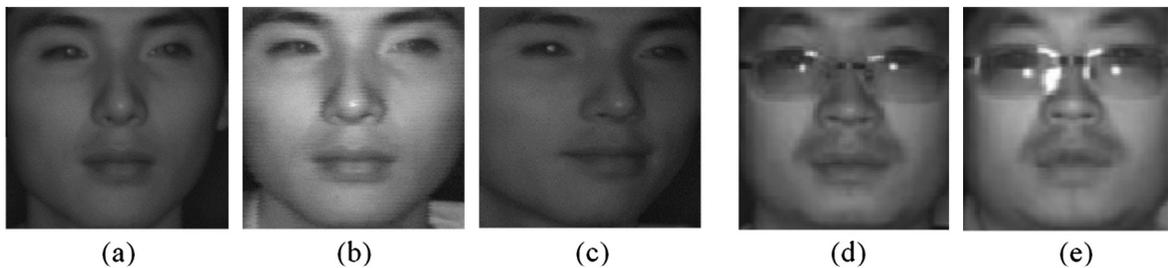


Fig. 7. (a, d) Sample of normalized images used as gallery images. Sample of images with variations in (b) scale, (c) head pose, and (e) time lapse used as probe images.

Table 5

Accuracy of different FR methods on PolyU NIR face database in the presence of various challenges.

Method	Mean \pm standard deviation	Confidence interval
GFC	76.79 \pm 2.31	[75.35, 78.22]
LBPL	81.57 \pm 4.47	[78.79, 84.34]
GDBC	75.62 \pm 3.56	[73.41, 77.82]
WS	80.16 \pm 3.64	[77.90, 82.41]
ZMUDWT	85.01 \pm 1.78	[83.90, 86.11]
ZM (Global part)	76.14 \pm 2.18	[74.78, 77.49]
HK (Local part)	84.26 \pm 2.53	[82.69, 85.82]
ZMHK (Proposed method)	87.22 \pm 2.11	[85.91, 88.52]

5.4. Comparison of the recognition time

Most FR systems are utilized in real-time applications, such as security access control, where the result of the FR system must be available immediately upon request. As any delay will significantly degrade user experience, the recognition time of the FR method is as important as its accuracy. We performed an evaluation of the execution time for existing methods compared to our proposed method with the parameter set-up as described in Table 2. To ensure fair comparison, all methods were implemented using MATLAB R2013a, and the evaluation was performed by running the code profiler for all implementations. To evaluate the execution time, we measured the recognition time (i.e. the time required to extract the features and classify one new probe image) in elapsed CPU seconds on an Intel Core i5 2.5 GHz CPU with 4 GB of RAM. Each test was repeated five times, and the mean recognition time and standard deviation are reported in Table 6.

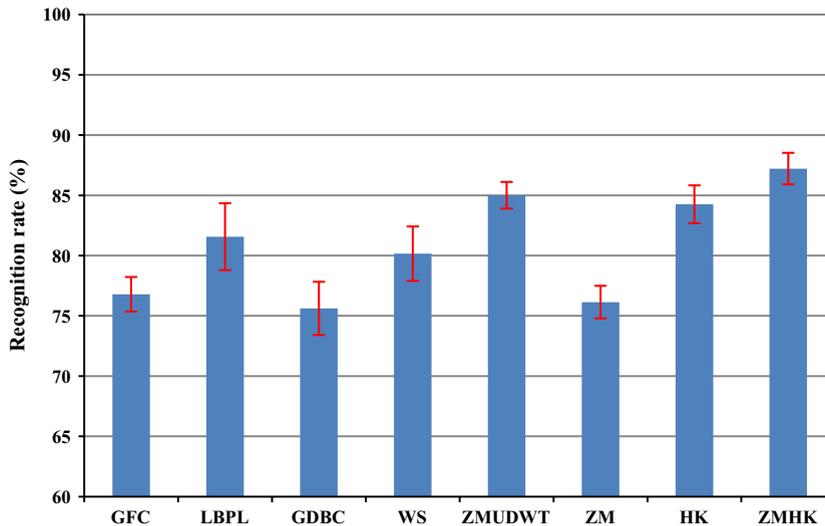


Fig. 8. Recognition results of different methods for the PolyU NIR face database.

Table 6

Comparison of recognition time for different FR methods.

Method	Mean recognition time (s) \pm std
GFC	0.14 \pm 0.031
LBPL	0.05 \pm 0.004
GDBC	0.24 \pm 0.006
WS	0.34 \pm 0.012
ZMUDWT	0.28 \pm 0.008
ZM (Global part)	0.09 \pm 0.012
HK (Local part)	0.13 \pm 0.004
ZMHK (Proposed method)	0.19 \pm 0.005

Comparison of the execution time of our method (ZMHK) with the next most accurate method (ZMUDWT) indicates that, besides having the highest face recognition accuracy, the execution time of the proposed ZMHK method is 32% faster (difference of -0.09 s) compared to ZMUDWT. We note that, as both ZMHK and ZMUDWT use the fusion of global and local features to improve the accuracy of recognition, they also increase computation time. Thus, it is not surprising that the execution times for both ZMHK and ZMUDWT are higher than other methods (except for wavelet scattering). However, in practical applications, the proposed ZMHK method provides the best trade-off between accuracy and time complexity (execution time).

6. Conclusion

In this paper, we have proposed a novel method for highly accurate NIR FR in order to tolerate deformations caused by facial expression, wearing eyeglasses, variations of head pose and scale and the effects of time lapse. The proposed method is based on the combination of global features extracted by the calculation of ZMs, and local features extracted from partitioned image patches by HKs. By applying PCA followed by LDA on both global and local features, multiple feature vectors were obtained and then combined with decision fusion to fully exploit global and local features. Employing HKs as local feature extractors and proposing a fusion rule for combining global and local features of different dimensionality, are the major contributions of this paper.

The performance of the proposed method was compared to other popular FR methods in the presence of the most common challenges in NIR FR systems. The CASIA NIR and the PolyU NIR face databases were used to validate the performance. The experimental results obtained showed that the proposed ZMHK method improved the FR accuracy of the ZMUDWT and also outperformed other existing FR methods significantly. We have also shown that the proposed ZMHK method has a faster execution time compared to ZMUDWT.

Acknowledgments

The authors would like to thank Universiti Teknologi Malaysia (UTM) and the Institute of Information Theory and Automation (UTIA) for their support in the research and development of this work. Also, the Institute of Automation and

Chinese Academy of Sciences (CASIA) for providing the CASIA NIR database to carry out our experiment. Also, thanks to the Biometric Research Center at Hong Kong Polytechnic University for providing the PolyU NIR face database. Sajad Farokhi is a researcher at UTM under the postdoctoral fellowship scheme for the project “Accurate Near Infrared Face Recognition” (Grant No. QJ130000.21A2.01E81). This work is supported by the Ministry of Education, Malaysia (FRGS Grant No. FRGS/1/2014/TK03/UTM/02/12, R/J130000.7823.4F461), the Ministry of Science, Technology and Innovation, Malaysia (MOSTI), (Science Fund Grant No. 01-01-06-SF1197, R/J130000.7923.4S081) and is partially supported by the Czech Science Foundation (Grant No. GA13-29225S). A part of this work was undertaken when S. Farokhi visited UTIA.

References

- [1] M.F.A. Abdullah, M.S. Sayeed, K. Sonai Muthu, H.K. Bashier, A. Azman, S.Z. Ibrahim, Face recognition with symmetric local graph structure (SLGS), *Expert Syst. Appl.* 41 (2014) 6131–6137.
- [2] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (1997) 721–732.
- [3] J. Bai, Y. Ma, J. Li, H. Li, Y. Fang, R. Wang, H. Wang, Good match exploration for thermal infrared face recognition based on YWF-SIFT with multi-scale fusion, *Infrared Phys. Technol.* 67 (2014) 91–97.
- [4] G. Bebis, A. Gyaourova, S. Singh, I. Pavlidis, Face recognition by fusing thermal infrared and visible imagery, *Image Vis. Comput.* 24 (2006) 727–742.
- [5] 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 (1997) 711–720.
- [6] 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 (2009) 750–762.
- [7] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, *Comput. Vis. Image Underst.* 101 (2006) 1–15.
- [8] L. Chengjun, H. Wechsler, Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition, *IEEE Trans. Image Process.* 11 (2002) 467–476.
- [9] Y. Dong, L. Rongbing, C. RuFeng, W. Rui, L. Dong, Z.L. Stan, Outdoor face recognition using enhanced near infrared imaging, in: *International Conference on Biometrics*, Seoul, Korea, 27–29 August, 2007, pp. 1–9.
- [10] J. Dowdall, I. Pavlidis, G. Bebis, Face detection in the near-IR spectrum, *Image Vis. Comput.* 21 (2003) 565–578.
- [11] S. Farokhi, S.M. Shamsuddin, J. Flusser, U.U. Sheikh, Assessment of time-lapse in visible and thermal face recognition, *World Acad. Sci. Eng. Technol.* 62 (2012) 540–545.
- [12] S. Farokhi, S.M. Shamsuddin, J. Flusser, U.U. Sheikh, M. Khansari, J.-K. Kourosh, Rotation and noise invariant near-infrared face recognition by means of Zernike moments and spectral regression discriminant analysis, *J. Electron. Imaging.* 22 (2013) 013030-1–013030-11.
- [13] S. Farokhi, S.M. Shamsuddin, J. Flusser, U.U. Sheikh, M. Khansari, J.-K. Kourosh, Near infrared face recognition by combining Zernike moments and undecimated discrete wavelet transform, *Digit. Signal Process.* 31 (2014) 13–27.
- [14] S. Farokhi, S.M. Shamsuddin, U.U. Sheikh, J. Flusser, Near infrared face recognition: a comparison of moment-based approaches, in: *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications*, Penang, Malaysia, 10–12 November, 2014, pp. 129–135.
- [15] S. Farokhi, U.U. Sheikh, J. Flusser, S.M. Shamsuddin, H. Hashemi, Evaluating feature extractors and dimension reduction techniques for near infrared face recognition systems, *J. Teknol.* 70 (2014) 23–33.
- [16] J. Flusser, T. Suk, B. Zitova, *Moments and Moment Invariants in Pattern Recognition*, Wiley, Chichester, 2009.
- [17] W.T. Freeman, E.H. Adelson, The design and use of steerable filters, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (1991) 891–906.
- [18] R.S. Ghiass, O. Arandjelović, A. Bendada, X. Maldague, Infrared face recognition: a comprehensive review of methodologies and databases, *Pattern Recognit.* 47 (2014) 2807–2824.
- [19] 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*, Dallas, Texas, USA, 4–9 August, 2013, pp. 1–10.
- [20] D. Goswami, C. Chi-Ho, D. Windridge, J. Kittler, Evaluation of face recognition system in heterogeneous environments (visible vs NIR), in: *IEEE International Conference on Computer Vision Workshops*, Barcelona, USA, 6–13 November, 2011, pp. 2160–2167.
- [21] L. Guo, M. Dai, M. Zhu, Quaternion moment and its invariants for color object classification, *Inform. Sci.* 273 (2014) 132–143.
- [22] B. Honarvar Shakibaei Asli, J. Flusser, Fast computation of Krawtchouk moments, *Inform. Sci.* 288 (2014) 73–86.
- [23] K.M. Hosny, A systematic method for efficient computation of full and subsets Zernike moments, *Inform. Sci.* 180 (2010) 2299–2313.
- [24] Z. Jun-Yong, Z. Wei-Shi, L. Jian-Huang, S.Z. Li, Matching NIR face to VIS face using transduction, *IEEE Trans. Inform. Forensics Sec.* 9 (2014) 501–514.
- [25] M. Jun, G. Yumao, W. Xiukun, L. Tsauyoung, Z. Jianying, Face recognition based on local binary patterns with threshold, in: *IEEE International Conference on Granular Computing (GrC)*, San Jose, California, USA, 14–16 August, 2010, pp. 352–356.
- [26] D. Kang, H. Han, A.K. Jain, S.-W. Lee, Nighttime face recognition at large standoff: cross-distance and cross-spectral matching, *Pattern Recognit.* 47 (2014) 3750–3766.
- [27] J. Kang, D. Nyang, K. Lee, Two-factor face authentication using matrix permutation transformation and a user password, *Inform. Sci.* 269 (2014) 1–20.
- [28] A.Z. Kouzani, F. He, K. Sammut, Towards invariant face recognition, *Inform. Sci.* 123 (2000) 75–101.
- [29] C. Kuang-Yu, L. Cheng-Fu, C. Chu-Song, H. Yi-Ping, Applying scattering operators for face recognition: a comparative study, in: *21st International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, 11–15 November, 2012, pp. 2985–2988.
- [30] S.M. Lajevardi, Z.M. Hussain, Higher order orthogonal moments for invariant facial expression recognition, *Digit. Signal Process.* 20 (2010) 1771–1779.
- [31] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, An efficient 3D face recognition approach using local geometrical signatures, *Pattern Recognit.* 47 (2014) 509–524.
- [32] D. Li, X. Tang, W. Pedrycz, Face recognition using decimated redundant discrete wavelet transforms, *Mach. Vis. Appl.* 23 (2012) 391–401.
- [33] S.Z. Li, R. Chu, S. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (2007) 627–639.
- [34] S.Z. Li, D. Yi, Face recognition using near infrared images, in: S.Z. Li, A.K. Jain (Eds.), *Handbook of Face Recognition*, Springer, 2011, pp. 383–400.
- [35] N. Loris, L. Alessandra, B. Sheryl, Local binary patterns variants as texture descriptors for medical image analysis, *Artif. Intell. Med.* 49 (2010) 117–125.
- [36] H. Maeng, S. Liao, D. Kang, S.-W. Lee, A.K. Jain, Nighttime face recognition at long distance: cross-distance and cross-spectral matching, in: *Asian Conference on Computer Vision*, Daejeon, Korea, 5–9 November, 2012, pp. 1–14.
- [37] Mamta, M. Hanmandlu, A new entropy function and a classifier for thermal face recognition, *Eng. Appl. Artif. Intell.* 36 (2014) 269–286.
- [38] Y. Ming, Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis, *Image Vis. Comput.* 35 (2015) 14–22.
- [39] A. Nabatchian, E. Abdel-Raheem, M. Ahmadi, Human face recognition using different moment invariants: a comparative study, in: *Congress on Image and Signal Processing (CISP '08)*, Sanya, Hainan, China, 27–30 May, 2008, pp. 661–666.
- [40] F. Omri, S. Fofou, M. Abidi, NIR and visible image fusion for improving face recognition at long distance, in: *6th International Conference on Image and Signal Processing (ICISP)*, Cherbourg, France, 30 June–2 July, 2014, pp. 549–557.
- [41] C. Pagano, E. Granger, R. Sabourin, G.L. Marcialis, F. Roli, Adaptive ensembles for face recognition in changing video surveillance environments, *Inform. Sci.* 286 (2014) 75–101.

- [42] M. Ramji, Illumination invariant face recognition: a survey of passive methods, *Proc. Comput. Sci.* 2 (2010) 101–110.
- [43] L. Shen, J. He, S. Wu, S. Zheng, Face recognition from visible and near-infrared images using boosted directional binary code, in: *International Conference on Intelligent Computing*, Zhengzhou, China, 11–14 August, 2012, pp. 404–411.
- [44] C. Singh, N. Mittal, E. Walia, Face recognition using Zernike and complex Zernike moment features, *Pattern Recognit. Image Anal.* 21 (2011) 71–81.
- [45] C. Singh, E. Walia, R. Upneja, Accurate calculation of Zernike moments, *Inform. Sci.* 233 (2013) 255–275.
- [46] D.A. Socolinsky, A. Selinger, J.D. Neuheisel, Face recognition with visible and thermal infrared imagery, *Comput. Vis. Image Underst.* 91 (2003) 72–114.
- [47] M.R. Teague, Image analysis via the general theory of moments, *JOSA* 70 (1980) 920–930.
- [48] M. Turk, A. Pentland, Face recognition using eigenfaces, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, HI, USA, 3–6 June, 1991, pp. 586–591.
- [49] PolyU-NIRFD. <http://www.comp.polyu.edu.hk/~biometrics/polyudb_face.htm>.
- [50] E. Vezzetti, F. Marcolin, G. Fracastoro, 3D face recognition: an automatic strategy based on geometrical descriptors and landmarks, *Robot. Auton. Syst.* 62 (2014) 1768–1776.
- [51] R. Wang, J. Yang, D. Yi, S. Li, An analysis-by-synthesis method for heterogeneous face biometrics, in: *Third International Conference on Biometrics*, Alghero, Italy, 2–5 June, 2009, pp. 319–326.
- [52] Y.-h. Wang, X.-j. Ning, C.-x. Yang, Q.-f. Wang, A method of illumination compensation for human face image based on quotient image, *Inform. Sci.* 178 (2008) 2705–2721.
- [53] W.K. Wong, H. Zhao, Eyeglasses removal of thermal image based on visible information, *Inform. Fusion* 14 (2013) 163–176.
- [54] Y. Xu, X. Li, J. Yang, D. Zhang, Integrate the original face image and its mirror image for face recognition, *Neurocomputing* 131 (2014) 191–199.
- [55] Y. Xu, J. Yang, D. Zhang, A. Zhong, Bimodal biometrics based on a representation and recognition approach, *Opt. Eng.* 50 (2011) 037202-1–037202-7.
- [56] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, Z. Lai, Using the idea of the sparse representation to perform coarse-to-fine face recognition, *Inform. Sci.* 238 (2013) 138–148.
- [57] Y. Xu, Q. Zhu, D. Zhang, Combine crossing matching scores with conventional matching scores for bimodal biometrics and face and palmprint recognition experiments, *Neurocomputing* 74 (2011) 3946–3952.
- [58] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu, H. Liu, Using the original and 'symmetrical face' training samples to perform representation based two-step face recognition, *Pattern Recognit.* 46 (2013) 1151–1158.
- [59] Z. Xuan, K. Josef, M. Kieron, Illumination invariant face recognition: a survey, in: *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, Washington, D.C, USA, 27–29 September, 2007, pp. 1–8.
- [60] B. Yang, J. Flusser, T. Suk, Steerability of Hermite kernel, *Int. J. Pattern Recognit. Artif. Intell.* 27 (2013) 1–25.
- [61] D. Yi, R. Liu, R. Chu, Z. Lei, S.Z. Li, Face matching between near infrared and visible light images, in: *Proceedings of IAPR/IEEE International Conference on Biometrics*, Seoul, Korea, 27–29 August, 2007, pp. 523–530.
- [62] S. Yu, S. Shiguang, C. Xilin, G. Wen, Hierarchical ensemble of global and local classifiers for face recognition, *IEEE Trans. Image Process.* 18 (2009) 1885–1896.
- [63] Y. Zhanfeng, Z. Wenyi, C. Rama, Pose-encoded spherical harmonics for face recognition and synthesis using a single image, *EURASIP J. Adv. Signal Process.* 2008 (2008) 1–18.
- [64] B. Zhang, L. Zhang, D. Zhang, L. Shen, Directional binary code with application to PolyU near-infrared face database, *Pattern Recognit. Lett.* 31 (2010) 2337–2344.
- [65] L. Zhang, B. Denney, J. Lu, A collaborative approach for face verification and attributes refinement, *Inform. Sci.* 281 (2014) 620–634.
- [66] X. Zhang, Y. Gao, M.K. Leung, Automatic texture synthesis for face recognition from single views, in: *18th International Conference on Pattern Recognition*, Hong Kong, China, 20–24 August, 2006, pp. 1151–1154.
- [67] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, *ACM Comput. Surv.* 35 (2003) 399–458.
- [68] L. Zhihui, X. Yong, Y. Jian, T. Jinhui, D. Zhang, Sparse tensor discriminant analysis, *IEEE Trans. Image Process.* 22 (2013) 3904–3915.