EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH

Open Access

Complementary feature sets for optimal face recognition

Chandan Singh[1], Neerja Mittal[2*] and Ekta Walia[3]

Abstract

In face recognition tasks, one kind of feature set is not adequate to generate superior results; thus, selection and combination of complementary features are crucial steps. In this paper, the fusion of two useful descriptors, i.e., the Zernike moments (ZMs) and the local binary pattern (LBP)/local ternary pattern (LTP), has been proposed. The ZM descriptor consists of good global image representation capabilities besides being invariant to image rotation and noise, while the LBP/LTP descriptors capture the innate details within some local parts of face image and are insensitive to illumination variations. The fusion of these two is observed to incorporate the traits of both of these individual descriptors. Subsequently, in this work, the performance of diverse feature sets of ZMs (i.e., magnitude features, magnitude plus phase features, and the real plus imaginary component features) combined with the LBP/LTP descriptor is analyzed on FERET, Yale, and ORL face databases. The recognition results achieved by the proposed method are approximately 10 to 30% higher than those obtained with these descriptors separately. Recognition rates of the proposed method are also found to be significantly better (i.e., by 8 to 24%) in case of single example image per person in the training.

Keywords: Face recognition; Zernike moments (ZMs); Local binary pattern (LBP); Local ternary pattern (LTP); Invariant image features

1 Introduction

In recent times, face recognition has become one of the widely used biometric techniques having a number of real-world applications like human-computer interaction, surveillance, authentication, computer vision applications, computer user interfaces, etc. An automatic face recognition system consists of some methods to ascertain a person's identity on the basis of his/her physiological characteristics. The sensitivity of available classifiers to different kinds of disparities such as illumination variation, facial expression, facial occlusion, pose variation, aging, etc. is among the most challenging problems that the researchers face [1].

In order to improve the existing face recognition techniques, discriminative competence of the invariant features selected to represent the face images should be high because, thereafter, classification is performed on the basis of these invariant features only. In literature,

the approaches used to represent the face images are classified broadly into two categories, namely, the global feature extraction approaches and the local feature extraction approaches [2]. The global feature extraction approaches are based on the statistical methods, wherein features are extracted from the whole face image. In this category, the subspace-based methods, namely, principal component analysis (PCA), Fisher linear discriminant (FLD), two-dimensional PCA (2DPCA), and two-directional two-dimensional PCA ($2D^2PCA$) [3-7], are some of the popular and most frequently employed techniques. Moment invariants, such as Hu's seven moment invariants and orthogonal rotation invariant moments such as Zernike moments (ZMs), pseudo-Zernike moments (PZMs), and orthogonal Fourier-Mellin moments (OFMMs), are observed to be very effective in global image description and recognition [8], and MPEG-7 uses some of them as region-based shape descriptors for image retrieval [9]. The magnitude of these moments is invariant to image rotation, and after applying some geometric transformations, it becomes invariant to translation and scale [10,11].

* Correspondence: neerjamittal_2k1@yahoo.com
[2]Central Scientific Instruments Organisation, Sector 30-C, Chandigarh 160030, India
Full list of author information is available at the end of the article

Springer

Local feature extraction approaches deal with fine information within the specific parts of face images such as eyes, nose, mouth, etc. Recently, a lot of work has been done on these methods because the local features are known to be robust against illumination, occlusion, expression, and noise variations. The local feature extraction approaches have been classified into two categories, i.e., the sparse descriptors and the dense descriptors. The sparse descriptors initially divide a face image into patches and then determine its invariant features. A prominent descriptor in this category is the scale-invariant feature transform (SIFT) introduced by Lowe [12], which possesses useful characteristics of being invariant to scale and rotation. Soyel et al. used the discriminative SIFT (D-SIFT) approach for optimal facial expression recognition, but this method is somewhat susceptible to the illumination variation [13]. In face recognition technology, Gabor wavelet is one of the most frequently used and successful local image descriptors. It incorporates the characteristics of space and frequency domains. The local features extracted by Gabor filters are invariant to scale and orientation and are able to detect edges and lines in the face images [14]. The main difficulty with Gabor filters is their high computational complexity. In case of the dense descriptors, local binary pattern (LBP) is one of the most widely used approaches due to its invariance to monotonic gray-level changes and ease in extraction of the local features. Apart from texture analysis, it has provided excellent results in many areas of image processing and computer vision including its wide use in face recognition [15,16]. Several variants of LBP are available in literature to represent face images with compact feature sets. Such variants also improve the classification performance of the basic LBP approach [17,18].

In complex applications like face recognition, it is observed that one kind of feature set is not rich enough to capture the entire face information. Thus, finding and combining the complementary feature sets have become an active research topic in recent years. Specifically, global features are related to the holistic characteristics of face, whereas local features describe the finer details within face images, so it seems logical to combine both of these feature sets since the information conveyed by them belongs to different attributes of the face images. In recent times, many researchers are developing the face recognition algorithms by combining the multiple feature sets. Kim et al. [19] have proposed a combined subspace-based approach using both global and local features obtained by applying linear discriminant analysis (LDA)-based method for face recognition. Zhou and Yang [20] have proposed fusing feature Fisher classifier (F^3C) approach where the face image is first divided into smaller subimages and then the discrete cosine transform (DCT)

technique is applied to the whole image and some subimages to extract the holistic and local facial features. After concatenating these DCT-based holistic and local facial features, the enhanced Fisher linear discriminant model (EFM) has been employed to generate a low-dimensional feature vector. Similarly, local and global information extracted by using DCT coefficients along with the Fisher classifier developed for high-dimensional multiclass problem have been proposed in [21]. Singh et al. proposed a robust two-stage face recognition approach by the fusion of global ZMs and Weber law descriptor (WLD)-based local features [22]. The usefulness of combining the global and local facial features is presented in [23] where a hierarchical ensemble of global and local features is performed. In this technique, 2D Fourier transform is used to extract the global features and the Gabor wavelet is opted to extract local features. Subsequently, equal weights are assigned to both the global and local features for combining the outputs of two classifiers (although it is established by the authors that the contribution of both global and local features is different). Wong et al. have proposed dual optimal multiband feature (DOMF) method for face recognition in which wavelet packet transform (WPT) decomposes the image into frequency subbands and the multiband feature fusion technique is incorporated to select optimal multiband feature sets that are invariant to illumination and facial expression. In this method, parallel radial basis function (RBF) neural networks are used to classify the two sets of features. The decision scores are then combined and processed by an adaptive fusion mechanism [24]. The use of steerable pyramid decomposition (S-P transform) both in global and local appearance and feature/score fusion has been analyzed in [25]. In this work, each face image is described by a subset of band-filtered images containing steerable pyramid coefficient. These S-P subbands are divided into small subblocks to extract the compact and meaningful feature vectors that provide a better representation of the class information. Recently, Liu and Liu [26] have proposed an approach for face recognition that fuses color and local spatial and global frequency information. This method is composed of multiple features of face images derived from LBP, DCT, hybrid color space, and the Gabor image representation. The combination of Gabor and LBP enhances the power of spatial histogram that is impressively insensitive to appearance variations. This method has proven to be robust against illumination, pose, and expression variations [27,28]. However, the combination of Gabor and multiresolution LBP descriptors requires significantly greater computation time.

Although a lot of research is going on combining the multiple feature sets, still the selection of complementary feature sets for fusion and the techniques for combining these divergent feature sets are a challenge. In

view of that, in this paper, a fusion of two complementary feature sets is proposed, where the global information of the face images is extracted by the ZM descriptor employing its rotation invariance characteristic, while the LBP/LTP descriptor captures the significant local information. Among various global shape descriptors, ZMs are observed to be one of the best shape descriptors because of their many attractive characteristics [8]. They possess minimum information redundancy, rotation invariance of their magnitude, robustness to noise, etc. The estimation of head movement by using the phase coefficients of ZMs of the original image and that of the rotated image has been employed to generate a set of features that is significantly tolerant to pose variation as well [29]. The magnitude features of ZMs obtained at some higher order of moments are observed to be invariant to expression variation [30]. On the other hand, the LBP descriptor is observed to be relatively more insensitive to illumination changes. It is computationally efficient as well as quite simple to implement. Recently, a useful extension to this approach is introduced, namely, the local ternary patterns (LTP), that is observed to be more discriminative and invariant to image noise in near-uniform regions as compared to LBP [18]. Particularly, combining the feature sets that are invariant to global variations as well as to local changes of face images would be an effective approach to achieve an optimal face recognition system. As discussed earlier, the information conveyed by the ZM and LBP/LTP descriptors are distinct and belong to different aspects of a facial image. Fusion of these descriptors is expected to be enriched with the useful characteristics of both of them. One of the critical issues involved in the fusion process is the time spent in the computation of the combined features. It is shown through time analysis of the proposed approach that the total time required for the recognition process is very small and can be afforded by PCs and other low computation devices.

The ZM descriptor provides three different sets of features, namely, magnitude features, combined magnitude and phase features [31,32], and the modified real and imaginary component features [29]. In this study, these diverse feature sets of ZMs are referred to as ZM_mag, $\text{ZM}_\text{magPhase}$, and $\text{ZM}_\text{component}$, respectively. The performance of the feature sets of ZMs combined with LBP descriptor, in comparison to that of the ZMs coupled with the LTP descriptor, is also analyzed. Consequently, the proposed fusion of the diverse feature sets of ZMs and the LBP/LTP descriptors provides various combined approaches such as $\text{ZM}_\text{mag}\text{LBP}$, $\text{ZM}_\text{magPhase}\text{LBP}$, $\text{ZM}_\text{component}\text{LBP}$, $\text{ZM}_\text{mag}\text{LTP}$, $\text{ZM}_\text{magPhase}\text{LTP}$, and $\text{ZM}_\text{component}\text{LTP}$. In order to compare the performance of these combined approaches to that of the individual ZM and LBP/LTP approaches, exhaustive experiments are performed on three prominent face databases, namely, FERET, Yale, and ORL,

against pose, illumination, expression, and noise variations. The results obtained show that the recognition rate of the combined approaches, in comparison to that of their individual counterparts, is significantly better varying between 10 and 30%. Experimental results also prove the efficacy of the proposed methods over other existing works. A significant improvement in recognition rate is achieved for the case of single training image per person.

The rest of the paper is organized as follows: Section 2 presents a brief overview of the ZM approach and the diverse feature sets obtained from it and includes a brief introduction to the LBP/LTP approaches, Section 3 describes the similarity measures used to evaluate the matching score of these methods, the procedure involved in the proposed fusion of the ZM and the LBP/LTP descriptors is described in Section 4, the experiments and results obtained are presented in Section 5, and the conclusions and future directions are presented in Section 6.

2 Baseline image descriptors
2.1 Global image descriptor
2.1.1 Zernike moments
The Zernike functions constitute a set of orthogonal basis functions mapped over the unit circle. Zernike moments of a function $f(x, y)$ are constructed by projecting it onto those functions. The ZMs of order n and repetition m are defined by

$$Z_{nm} = \frac{n+1}{\pi} \iint\limits_{x^2+y^2\leq1} f(x,y) V_{nm}^*(x,y) dx dy \qquad (1)$$

where $n \geq 0$, $|m| \leq n$, and $V_{nm}^*(x, y)$ are the complex conjugates of the Zernike function $V_{nm}(x, y)$, where

$$V_{nm}(x,y) = R_{nm}(x,y) e^{jm\theta} \qquad (2)$$

with $= \sqrt{-1}$, $\theta = \tan^{-1}(y/x)$, $\theta \in [0, 2\pi]$, and

$$R_{nm}(x,y) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)! (x^2+y^2)^{\frac{n-2s}{2}}}{s! \left(\frac{(n+|m|)}{2}-s\right)! \left(\frac{(n-|m|)}{2}-s\right)!} \qquad (3)$$

The ZMs are derived for a discrete image function using zeroth-order approximation of Equation 1 given by [10]

$$Z_{nm} = \frac{n+1}{\pi} \sum_{\substack{i=0 \\ x_i^2+y_k^2\leq1}}^{N-1} \sum_{k=0}^{N-1} f(i,k) V_{nm}^*(x_i,y_k) \Delta^2 \qquad (4)$$

where

$$x_i = \frac{2i + 1 - N}{N\sqrt{2}}, y_k = \frac{2k + 1 - N}{N\sqrt{2}}, \quad (5)$$

$$i, k = 0, 1, 2, ..., N-1, \quad \text{and} \quad \Delta = \frac{2}{N\sqrt{2}}$$

2.1.2 Diverse feature sets of ZMs and related work

Since the magnitude of ZMs is invariant to rotation, usually it is used as invariant image descriptor in many image analysis and pattern recognition applications. The phase component of ZMs is, however, ignored. It is observed that the phase component also carries equally significant information as the magnitude component does [31]. Therefore, in recent years, significant research work has been carried out to incorporate ZM phase coefficients along with their magnitudes as invariant feature descriptors. At present, there are two approaches to realize this objective. In the first approach, developed by Revaud et al. [33], a similarity measure incorporating both the magnitude and phase coefficients of the query and database image is used. The method provides excellent pattern matching performance but at the cost of enhanced computation time. In the second approach, the rotation angle between a query image and the database image is estimated. It is assumed that the query image is the rotated version of the original database image. The estimated rotation angle is used to cancel the effect of rotation in order to compare the phases of the query and database images. Recently, we devised a novel way to correct phase coefficients without estimating rotation angle. The method was applied successfully in face recognition [29]. The method works as follows: Suppose Z_{nm} and Z'_{nm} are the ZMs of database and query images, respectively, and ϕ_{nm} and ϕ'_{nm} are their respective phase angles. We compute $m\theta = \phi_{nm} - \phi'_{nm}$ and correct the ZMs of the query image by evaluating $Z'^c_{nm} = Z'_{nm} e^{jm\theta}$. If the two images are same, then $Z'^c_{nm} = Z_{nm}$, otherwise $Z'^c_{nm} \neq Z_{nm}$; therefore, the real and imaginary components of ZMs of the query and database images can be compared separately, instead of comparing only their magnitude. An attractive advantage of this approach is that by using two-component feature vectors, the number of features is almost doubled as compared to the ZM magnitude only features for the same order of moments. This approach has additional advantages of having low computation cost, less susceptibility to image noise, and numerical stability, in addition to providing better recognition rate [29]. Throughout the paper, these features of ZMs based on magnitude, magnitude together with corrected phase [31], and the corrected real and imaginary parts of ZMs [29] are

referred to as ZM_{mag}, $ZM_{magPhase}$, and $ZM_{component}$, respectively.

2.2 Local image descriptor

2.2.1 Local binary pattern

Ojala et al. have introduced the local binary patterns (LBP) for effective texture description that has been used in many image processing and computer vision applications [15]. The most important property of this approach is its tolerance against illumination variation. Being computationally simple, it provides significant advantage over other approaches. The LBP operator takes some specific neighborhood around each pixel, and it then thresholds the values of these neighborhood pixels with respect to the central pixel's value. The resulting binary pattern is used as an element of the local image descriptor. Thus, it assigns a label to every pixel p_i of an image by thresholding its respective 3×3 neighborhood values with the value of the central pixel p_c and producing the result in the form of an 8-bit binary code. The LBP operator is computed as

$$\text{LBP} = \sum_{i=0}^{7} 2^i b(p_i - p_c) \quad (6)$$

$$b(p_i - p_c) = \begin{cases} 1, & \text{if } p_i \geq p_c \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where the values of i move along the eight neighbors of the central pixel. In case of 8-bit patterns, Ojala et al. [15] have observed that out of 2^8 patterns, only 58 uniform patterns provide approximately 90% information of the image neighborhoods while the remaining patterns consist of mostly noise. This attribute significantly reduces the number of LBP histogram bins from 256 to 59 where all the non-uniform patterns are stored in a single bin, the 59th bin.

2.2.2 Local ternary patterns

The local histogram features obtained from LBP have proven to be highly discriminative in face recognition [16,17]. However, they are found to be sensitive to noise because of the fact that they are thresholded exactly at the value of the central pixel especially in near-uniform and smooth regions of face images like cheeks or forehead. Recently, an important extension to the original LBP is provided by Tan et al. [18]. It generates three-valued codes corresponding to each image pixel. In their method, the binary LBP code is replaced with the ternary LTP code and the gray values in a zone of width $\pm w$ around the central pixel p_c are quantized to 0 and the values above this zone are quantized to +1 while those below it are quantized to −1. Specifically, the value of

$b(x)$ given in Equation 7 is replaced with the following three-valued function:

$$b'(p_i, p_c, w) = \begin{cases} 1, & p_i \geq (p_c + w) \\ 0, & |p_i - p_c| < w \\ -1, & p_i \leq (p_c - w) \end{cases} \qquad (8)$$

where w is a user-defined threshold. The LTP code is assumed to be invariant to image noise but may not be strictly invariant to gray-level transformations. The concept of uniform patterns to obtain histogram features is also applicable to LTP. For simplicity, the three-valued LTP codes are split into their positive and negative bisects which generate two sets of histogram features out of which one corresponds to the positive patterns and the other represents the negative patterns [18].

3 Similarity measures used

In this section, the similarity measures used for finding the matching scores of ZM and LBP/LTP descriptors are discussed briefly. In this work, it is observed that the fusion of matching scores obtained by applying L_2 – Norm/L_1 – Norm on the ZM descriptor and histogram intersection on the LBP/LTP descriptor generates superior performance. Hence, these different similarity measurement techniques are used on the feature sets generated by these descriptors. Since the matching scores obtained from these different approaches are heterogeneous, normalization is required to transform these matching scores to a common range before combining them.

3.1 Similarity measure for ZM descriptor

The magnitude features of ZMs, i.e., ZM_{mag}, of two images are compared by evaluating the normalized Euclidean distance (L_2 – Norm) between them. The normalized L_2 – Norm between the two sets of feature vectors of ZMs is given by

$$d_{mag} = \frac{1}{L} \sqrt{\frac{\left(|Z_i'| - |Z_i| \right)^2}{\max \left(|Z_i'|^2, |Z_i|^2 \right)}} \qquad (9)$$

where Z_i' and Z_i are the feature vectors of the query and the database images, respectively, and L represents the size of the feature vector consisting of the magnitude of ZMs. The normalized Euclidean distance d_{Phase} defined by [31], between ZM phases of the query and the database images, is computed as

$$d_{Phase} = \frac{1}{L} \sqrt{\sum_{i=1}^{L} \frac{1}{2\pi} \left(|\varphi_i'| - |\varphi_i| \right)^2} \qquad (10)$$

where φ is the phase angle of the database image and φ' is the phase angle after estimating the rotation angle between the query and database images and correcting the

phase [31]. The total distance $d_{magPhase}$ between the feature vectors consisting of $ZM_{magPhase}$ coefficients has been evaluated by using the distances d_{mag} and d_{Phase}, computed as per Equations 9 and 10, respectively. The formula used to compute the $d_{magPhase}$ is given as

$$d_{magPhase} = (w_1 d_{mag} + w_2 d_{Phase})/(w_1 + w_2) \qquad (11)$$

Normally, equal weights are assigned to simplify this process, i.e., $w_1 = w_2 = 0.5$.

The $ZM_{component}$ descriptor includes the modified real and imaginary parts of ZMs to formulate a two-component feature vector for each ZM. The normalized L_1 – Norm-based distance measure for the evaluation of similarity between component features of the database and the query images is given as under

$$d_{comp} = \frac{1}{L} \sum_{i=1}^{L} \frac{\left(|Re(Z_i) - Re(Z_i'^c)| \right)}{\max \left(|Re(Z_i)|, |Re(Z_i'^c)| \right)} + \frac{\left(|Im(Z_i) - Im(Z_i'^c)| \right)}{\max \left(|Im(Z_i)|, |Im(Z_i'^c)| \right)} \qquad (12)$$

The above mentioned distance metric d_{comp} has proven to be a better similarity measure between two sets of component feature vectors [29].

3.2 Similarity measure for LBP/LTP descriptor

The histogram intersection distance has been used to compare the feature vectors of query and database images for both the LBP and the LTP descriptors. The histogram intersection distance evaluated for every bin n between the database image and the query image is given as

$$D_h(Hd, Hq) = \frac{\sum_{n=1}^{B} \min(Hd_n, Hq_n)}{\min \left(\sum_{n=1}^{B} Hd_n, \sum_{n=1}^{B} Hq_n \right)}, \quad D_h(Hd, Hq) \in [0, 1] \qquad (13)$$

where Hd and Hq are the histograms consisting of LBP/LTP features of the database and the query images, respectively. B is the total number of bins in the histograms. If either of the two images, i.e., the database and the query images is identical, then the value of $D_h(Hd, Hq)$ is 1.

4 Fusion of ZM and LBP/LTP descriptors

The ZM descriptor and the LBP/LTP descriptors are observed to be complementary to each other, and their fusion is expected to be able to discriminate the face images even in the presence of diverse variations. The ZM descriptor is observed to extract the global information of the images more effectively as compared to that of any other global descriptor [9]. On the other hand, the LBP and the LTP descriptors have been established

to be successful methods for representing the finer interior details within the face images. The feature set established by the fusion of these autonomous approaches, i.e., ZMs and the LBP/LTP, is supposed to be enriched with the invariant characteristics of both of them. Exhaustive experiments performed against pose, illumination, expression, and noise variations on the suitable databases prove that the said hypothesis is correct.

The procedure followed to recognize the face images by the proposed combined approaches, i.e., $ZM_{mag}LBP$, $ZM_{magPhase}LBP$, $ZM_{component}LBP$, $ZM_{mag}LTP$, $ZM_{magPhase}LTP$, and $ZM_{component}LTP$, is briefly described in Figure 1. The recognition of face images through the proposed fusion of feature sets includes three stages - feature extraction, fusion of similarity score, and classification. The first stage of this procedure creates the invariant feature sets extracted by using ZM and LBP/LTP descriptors. The second stage involves fusion of the matching scores obtained from these feature sets after applying the similarity measures as described in the previous section. A number of feasible techniques such as fusion at the feature extraction level, matching score level, or decision level exist for combining the multiple feature sets. It is not easy to combine the information at the feature level when the feature sets obtained by different techniques are either inaccessible or incompatible. Fusion at the decision level is too rigid as only a limited amount of information is available at this level. Therefore, integration at the matching score level is generally preferred due to the

ease of accessing and combining matching scores [34]. In the proposed work, feature vectors are obtained by applying ZM and LBP/LTP descriptors which provide complementary information. Further, we observed that for the LBP/LTP descriptor, the matching score evaluated by the histogram intersection measure gives better results than using $L_2 - Norm$. Hence, in this work, the fusion at the matching score level is employed wherein the histogram intersection (using Equation 13) and $L_2 - Norm$ (using Equations 9 and 11 for ZM_{mag} and $ZM_{magPhase}$)/$L_1 - Norm$ (using Equation 12 for $ZM_{component}$) is used for evaluating the matching scores of the feature vectors obtained from the LBP/LTP and ZM descriptors, respectively. Thereafter, these individual matching scores are combined by using the sum rule to generate a single scalar score which is then used to make the final decision. The sum rule to compute the fusion of individual matching scores is given as below

$$\text{sum rule } (F_{sr}) = \frac{\{S_Z + (1 - S_L)\}}{2} \tag{14}$$

where S_Z and S_L represent the matching scores of the ZM descriptor and the LBP/LTP descriptors, respectively. The matching scores of these approaches are normalized before fusion. Normalization is required to map the matching scores obtained from multiple frameworks to a common range so that they can be easily combined. In order to combine these matching scores, S_L is subtracted
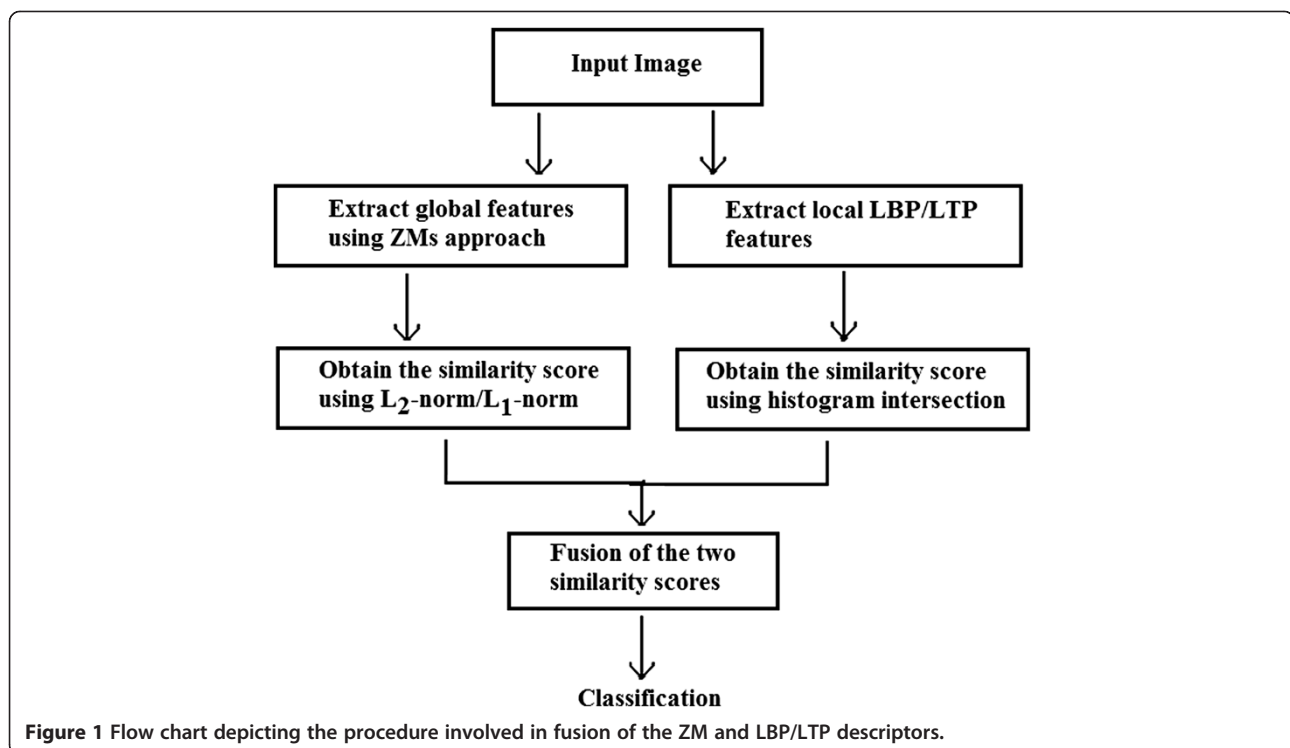

Figure 1 Flow chart depicting the procedure involved in fusion of the ZM and LBP/LTP descriptors.

from 1, so that the histogram intersection would now signify higher similarity with lower values. Finally, in the third stage, we use the nearest neighbor rule to perform classification. This method always gives us only one recognized image which is labeled as either correct or incorrect in order to evaluate the recognition performance. The recognition rate (in percentage) is measured by using the following formula:

$$\text{RecRate} = \frac{(N\text{test}-Nf)}{N\text{test}} \times 100 \qquad (15)$$

where Ntest is the total number of images in the test set and Nf is the number of images recognized incorrectly.

5 Experiments and results

In order to evaluate the performance of the considered autonomous approaches in comparison to that of the proposed combined approaches, experiments are performed on three well-known and calibrated face databases, namely, FERET face database [35] consisting of images in diverse variations, Yale face database [36] consisting of illumination and expression variations, and ORL face database [37] having small pose (tilt/yaw) changes. It is well known that the accuracy of the face recognition system is significantly affected by the kind of variations present in images of the face database as well as by the number of images of each subject (i.e., person) in the training set. Thus, exhaustive experiments in a comprehensive and deterministic manner are performed with respect to different types of variations present in these databases. The number of training images per person is also varied to observe its effect on recognition accuracy. The best results are highlighted in italics. All the experiments are performed in Visual C++6.0 under Microsoft Windows environment on a PC with a 3.0-GHz CPU and a 3-GB RAM.

5.1 Performance on FERET database

FERET grayscale face database has become the most popular and standard database in the field of face recognition. We have performed experiments on two subsets of this database, consisting of the frontal to profile pose variation. The first subset is formed by randomly selecting 100 persons with seven different poses (yaw) 0°, ±22.5°, ±67.5°, and ±90°. The second subset consisted of FERET '*b*' category images of 200 persons in different illumination, expression, and pose angles of 0°, ±15°, ±25°, ±40°, and ±60°. In this work, the first subset is called FERET_A. It consists of 700 images. The second subset is named as FERET_B and contains 2,200 images. FERET evaluation protocol partitions the database into gallery (1,196 images of 1,196 persons) and four probe sets, namely, *fafb*, *fc*, *dup I*, and *dup II*. The images in the *fafb* set are with facial expression

variation, the *fc* set contains images with illumination variations, and the images in the *dupI* and *dupII* sets represent aging effects. For detailed experimentation against pose, expression, and illumination variations, various data partitions are generated for these subsets which are described in Table 1. The original images of this database are of size 256 × 384 pixels. We transformed these images into 128 × 128-pixel size in order to reduce the time taken for conducting the experiments, while the face images from FERET_Gallery/Probe subset are cropped and resized to 64 × 64 pixels. Some sample images, for one person, from this database are shown in Figure 2. The face images of FERET_A, FERET_B, and FERET_Gallery/Probe subsets are partitioned into 64 patches of 16 × 16 and 8 × 8 pixels, respectively, to extract the local LBP/LTP features, while the global ZM features are extracted from whole face images.

In order to analyze the performance of the proposed combined approaches, the first set of comprehensive experiments is performed on FERET_A1 category as described in Table 1. The different possible trials for this setup, containing various combinations of the training and the test sets, are shown in Table 2. The average recognition performance of the individual and the combined approaches over these different trials (seven) is presented in Figure 3 for different values of maximum order of ZMs, denoted by $n_{\max}$. Further, it is pertinent to mention here that the values of $n_{\max}$ have no effect on the performance of the LBP and LTP descriptors. So, the results presented for the LBP/LTP descriptors remain the same for each value of $n_{\max}$. From the results presented, it is observed that among the autonomous approaches, the performance of the $ZM_{component}$ approach is better than that of others. However, the proposed combined approaches exhibit significantly high recognition rates compared to their individual counterparts. In this experiment, the $ZM_{magPhase}LBP$ approach provides the highest recognition rate at 71.24%.

Next, experiments are performed on FERET_A2 category, and the recognition results are presented in Figure 4 for both the individual and the combined approaches over different order of moments $n_{\max}$ used for the ZM descriptors. The basic LBP/LTP descriptors are not invariant to rotation; however, in this category, they perform better than the ZM descriptor. This is due to the fact that the higher pose angles occlude a significant portion of the face and on this kind of distortion, the local feature sets are observed to be more successful than the global features. On the other hand, the proposed fusion of the ZM descriptors and the LBP/LTP descriptors achieves approximately 20% improvement in the recognition results in comparison to that of these independent approaches. It is also noticed that the ZM descriptors coupled with the LBP descriptor generate superior results and the highest recognition rate

Table 1 Data partition on FERET database for performing various experiments

Set of experiment	Category	Training	Testing	Remarks
1	FERET_A1	One image of each person resulting in a total of 100 images	Remaining six images of each person in different pose variations, i.e., a total of 600 images	Seven different trials of this setup (as shown in Table 2) have been taken, and the recognition result given in Figure 3 is the average of all these trials
2	FERET_A2	One image of each person in frontal, i.e., 0° pose	Six remaining images having pose variations of ±22.5°, ±67.5°, and ±90°	Against the frontal pose image, all the images in different poses (≤ ±90°) are taken for testing
3	FERET_A3	One image in frontal pose	Four images having pose variations of ±22.5° and ±67.5°	Testing the performance for small and large pose variation up to ±67.5°
4	FERET_A4	One image in frontal pose	Two images in ±67.5° pose angle	Testing the performance for large pose variation in the left and right directions
5	FERET_A5	One image in frontal pose	Two images in ±22.5° pose angle	Testing the performance for small pose variation in the left and right directions
6	FERET_B1	One image in frontal pose, i.e., the image labeled 'ba', is in the training set	Ten remaining images having illumination, expression, and pose (up to ±60°) variations, i.e., the images labeled bk, bj, bb, bc, bd, be, bf, bg, bh, and bi, are kept in the test set	Images consisting of all the three variations, i.e., illumination, expression, and pose (up to ±60°), are taken for testing against the frontal pose image. A total of 200 images are in the training set, and 2,000 images are in the test set
7	FERET_B2	Three random images of each person resulting in a total of 600 images	Remaining seven images of each person in different variations, i.e., 1,600 images	Ten different trials of this setup have been taken, and the recognition result (Table 3) is the average of all these trials
8	FERET_B3	One image of each person in frontal pose, i.e., the image labeled 'ba,' is in the training set and contains a maximum of 200 face images in it	Eight images of each person having pose variations (up to ±60°), i.e., the images labeled bb, bc, bd, be, bf, bg, bh, and bi, are kept in the test set resulting in a maximum of 1,600 images in it	This category consists of the images of 200 persons in 9 pose variations, from frontal to profile pose, resulting in a total of 1,800 images in it. In this category, the different experiments are carried out by varying the number of subjects (persons) in the database
9	FERET_Gallery/ Probe	Standard gallery set fa contains 1,196 images of 1,196 subjects in frontal view	fafb, fc, dupl, and dupll	The images in the fafb (1,195 images) set are with facial expression variation, the fc (194 images) set contains images with illumination variation, and the images in the dupl (722 images) and dupll (234 images) sets represent aging effects

of 77.33% is accomplished by the $ZM_{magPhase}LBP$ approach. In case of the LTP descriptor, the LTP combined with the $ZM_{component}$ descriptor, i.e., $ZM_{component}LTP$, provides better results than those of other combinations.

Further experiments are performed on FERET_A3, FERET_A4, and FERET_A5 categories. The recognition results for these setups are shown in Table 3. On this database, fusion of the ZM features obtained for $n_{max} = 9$ provides better results. Accordingly, here, all other experiments have been conducted only for this order of moments. It is observed from Table 3 that there is an improvement in recognition rates by approximately 10 to 20% due to the fusion of the ZM and LBP/LTP descriptors. Further, it is observed that the recognition rates decline significantly with increase in the pose angle of test images (e.g., the highest recognition rate of only 74.5% is noticed for FERET_A4 which contains pose variations of ±67.5° in test images). This outcome is

obvious because of the fact that the presence of the higher pose angle will occlude a significant part of the face image. The highest recognition rate of 86.5% is achieved by the $ZM_{magPhase}LBP$ approach on FERET_A3 category. On FERET_A5 category, a superior recognition

Table 2 Different trials for one image in training and remaining six in test set for FERET_A1 category

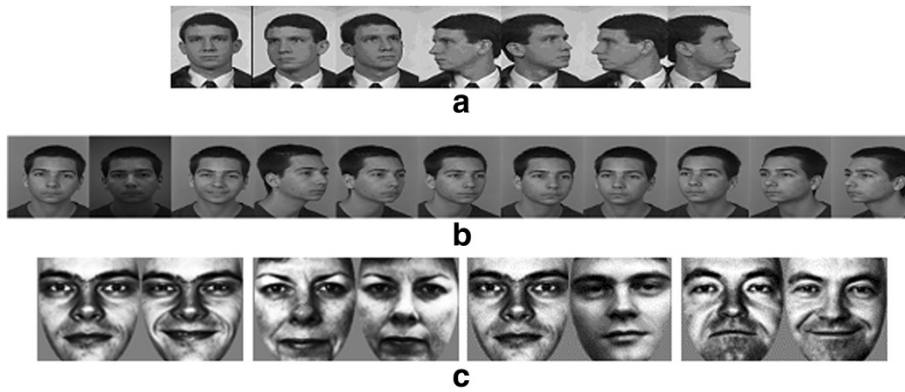| Trial	Image in training set in pose angle	Images in test set in pose angle
1	0°	±22.5°, ±67.5°, ±90°
2	+22.5°	0°, −22.5°, ±67.5°, ±90°
3	−22.5°	0°, +22.5°, ±67.5°, ±90°
4	+67.5°	0°, ±22.5°, −67.5°, ±90°
5	−67.5°	0°, ±22.5°, +67.5°, ±90°
6	+90°	0°, ±22.5°, ±67.5°, −90°
7	−90°	0°, ±22.5°, ±67.5°, +90°

Figure 2 Sample face images. Sample images from **(a)** FERET_A, **(b)** FERET_B, and **(c)** FERET_Gallery/Probe subsets.
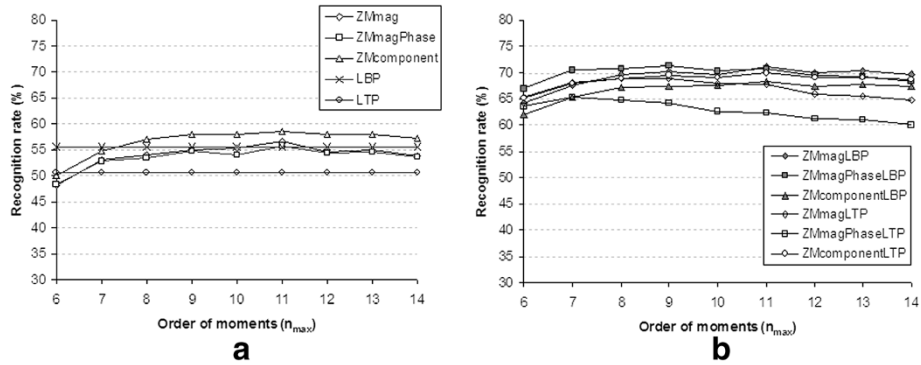

Figure 3 Performance (average) on FERET_A1 category. Average recognition performance of **(a)** individual and **(b)** combined approaches.
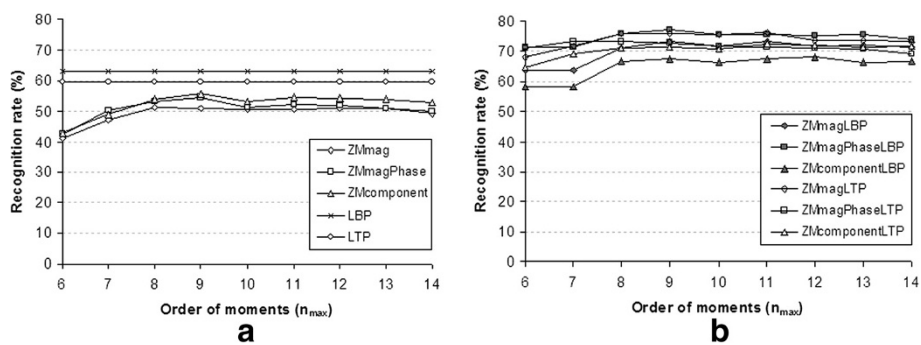

Figure 4 Performance on FERET_A2 category. Recognition results of **(a)** individual and **(b)** combined approaches.

Table 3 Performance of the considered approaches on FERET database

Descriptor	FERET_A3	FERET_A4	FERET_A5	FERET_B1	FERET_B2	FERET_Gallery/Probe			
						fb	*fc*	*dupl*	*dupll*
ZM_{mag}	63.75	34.0	93.5	62.0	76.3	67.53	31.8	35.0	38.0
$ZM_{magPhase}$	67.25	40.0	94.5	60.5	75.87	68.02	32.96	35.8	39.0
$ZM_{component}$	68.75	41.5	96.0	63.45	76.93	65.0	30.36	33.0	37.0
LBP	74.75	56.0	93.5	66.7	80.71	96.87	79.0	66.0	63.7
LTP	70.5	50.0	91.0	64.75	80.08	94.87	72.74	62.0	56.0
$ZM_{mag}LBP$	83.0	67.0	99.0	77.5	86.43	97.58	89.18	69.25	66.0
$ZM_{magPhase}LBP$	*86.5*	*74.5*	98.5	*78.5*	*88.2*	*98.04*	*91.5*	*69.39*	*66.5*
$ZM_{component}LBP$	79.5	59.5	*99.5*	74.05	84.69	96.0	88.0	68.5	64.0
$ZM_{mag}LTP$	84.5	70.0	99.0	76.3	85.48	93.0	85.0	65.0	59.5
$ZM_{magPhase}LTP$	81.75	65.5	98.0	75.55	85.97	93.5	85.8	65.7	58.5
$ZM_{component}LTP$	82.25	65.0	*99.5*	74.6	84.74	91.2	84.0	64.0	58.0

rate of 99.5% is achieved by both the $ZM_{component}LBP$ and the $ZM_{component}LTP$ approaches. Similarly, the $ZM_{magPhase}LBP$ approach provides the highest recognition rate on FERET_A4 category which reveals the contribution of ZM phase coefficients towards the improvement in recognition results.

For FERET_B, the recognition results obtained by performing the experiments on FERET_B1 and FERET_B2 categories are also shown in Table 3. It is clear from the experimental results that the combined approaches exhibit approximately 15 to 20% hike in the recognition rate than that obtained by the individual approaches. The highest recognition rate of 78.5 and 88.2% is achieved by the $ZM_{magPhase}LBP$ descriptor for FERET_B1 and FERET_B2 categories, respectively. The result obtained on FERET_Gallery/Probe set ascertains the robustness of the proposed system against changes in expression and lighting; however, further research in aging is required. In general, on FERET_B and FERET_Gallery/Probe subsets, the combination of $ZM_{magPhase}$ with LBP descriptor generates higher results than others.

5.2 Performance on Yale database

The Yale face database contains 11 images per person for 15 individuals resulting in a total of 165 images. The images in this database have major variations in illumination and facial expressions. They also have images demonstrating occlusion of eyes with eyeglasses. The original size of the images in this database is 243 × 320 pixels with 256 gray levels. For the experiments, these are cropped down to 64 × 64 pixels. Sample cropped images from this database, for one person, are shown in Figure 5. Here also, the face images are partitioned into 64 patches of 8 × 8 pixels to extract the local LBP/LTP features.

In order to examine the improvement in performance by the proposed combined approaches across the expression and illumination variations, exhaustive experiments are performed on this database by taking different number of images in training and test sets. Accordingly, for this purpose, various data partitions have been generated which are presented in Table 4. The first set of comprehensive experiments is performed on YALE 1 category where out of total 11 images of each person, one image is taken in the training set and all the remaining are placed in the test set. This process is repeated 11 times by taking different face images of each person in the training set. Average recognition results over 11 different runs of training and test sets are presented in Figure 6 for different n_{max} of ZMs.

From the results obtained, it is observed that an average improvement of approximately 12% is achieved by the proposed combined descriptors as compared to the individual approaches. In case of individual approaches, the performance of both the LBP and LTP approaches is better than that of the three descriptors of ZMs. The result obtained depicts that the local approaches are able to capture the interior details of a face image more efficiently than the global ones. This certainly enhances the

Figure 5 Sample cropped face images (a-k) for one person from Yale database.

Table 4 Data partition on Yale database for performing various experiments

Set of experiment	Category	Training	Testing	Remarks
1	YALE 1	Random one	All remaining (i.e., remaining ten images except the one selected for the training)	Eleven different trials of this setup have been taken, and the recognition result (Figure 6) is the average of all these trials
2	YALE 2.1	Random two	Remaining nine	The recognition result (Table 5) is the average of ten random trials on each setup, i.e., YALE 2.1, YALE 2.2, YALE 2.3, YALE 2.4, and YALE 2.5
	YALE 2.2	Random three	Remaining eight	
	YALE 2.3	Random four	Remaining seven	
	YALE 2.4	Random five	Remaining six	
	YALE 2.5	Random six	Remaining five	
3	YALE 3.1	a	f, g, h, i	Testing consists of experiments against illumination variation
	YALE 3.2	b, c, d, e, j, k	a, f, g, h, i	
	YALE 4.1	a	b, c, d, e, j, k	Testing consists of experiments against expression variation
	YALE 4.2	f, g, h, i	a, b, c, d, e, j, k	

suitability of these methods to outperform even in the presence of only a single exemplar image per person. From the results depicted in Figure 6, it is observed that among all the combined approaches, the highest recognition results are achieved by the $ZM_{magPhase}LBP$ descriptor. It is also observed that for the proposed combined methods, fusion of ZM features obtained for $n_{max} = 11$ provides better results. Hence, on this database, all other experiments have been carried out for this order of moments.

Next, the experiments are performed on YALE 2.1, YALE 2.2, YALE 2.3, YALE 2.4, and YALE 2.5 categories. The average recognition results over ten trials of each group of the training and the test sets are presented in Table 5. It is well known that the LBP and LTP descriptors are invariant to changes in intensities of the images, so the results obtained by these two approaches are quite higher than those obtained by the ZM descriptors. Hence, on this database, it has been realized that the LBP/LTP feature sets contribute much more towards the improvement in the recognition rate of the proposed

combined approaches which are significantly higher than the individual approaches. In most of the cases, the LBP and LTP descriptors combined with ZM_{mag} features generate superior results. For example, on YALE 2.4 category, the highest recognition rate of 97.56% is achieved by the $ZM_{mag}LBP$ approach. However, on YALE 2.3 category, the average highest recognition rate is 97.14% with the $ZM_{mag}LBP$ approach. Thus, in general, the proposed $ZM_{mag}LBP$ approach outperforms the others.

Thereafter, experiments are performed on YALE 3.1 and YALE 3.2 categories against illumination variation. Similarly, in order to examine the performance of the proposed approaches particularly over expression variation, experiments are carried out on YALE 4.1 and YALE 4.2 categories. The results obtained from these experiments are presented in Table 6. From the results shown, it is clearly noticed that the proposed combined approaches show an improvement in performance by approximately 30% over the ZM descriptors alone, whereas in comparison to the performance of individual LBP/LTP descriptors, an improvement of approximately 10% is achieved. As
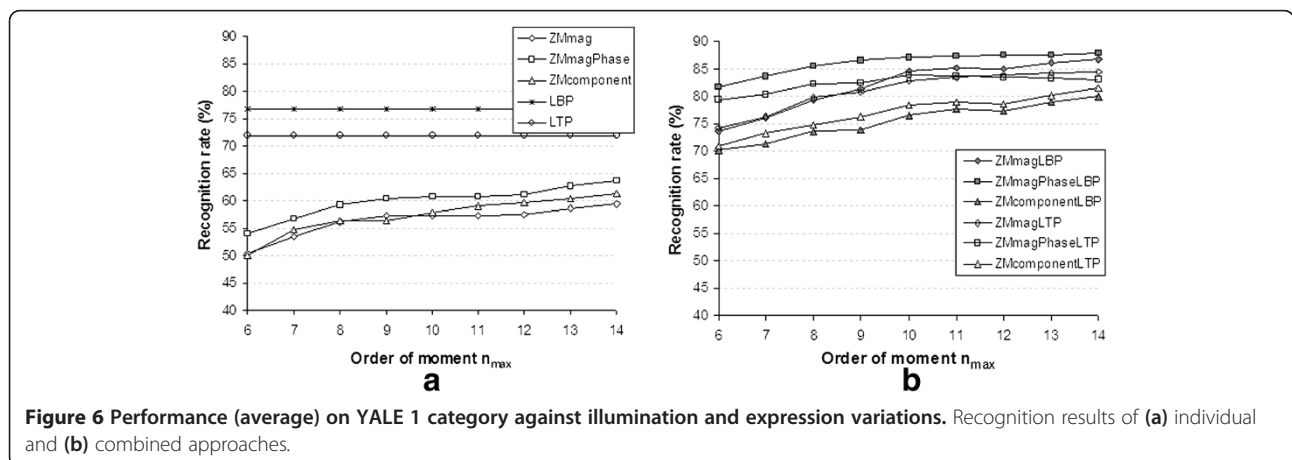

Figure 6 Performance (average) on YALE 1 category against illumination and expression variations. Recognition results of **(a)** individual and **(b)** combined approaches.

Table 5 Performance (average) of the individual/combined approaches on Yale database

Descriptor	YALE 2.1	YALE 2.2	YALE 2.3	YALE 2.4	YALE 2.5
ZM_{mag}	70.96	74.42	79.43	80.56	84.27
$ZM_{magPhase}$	74.37	79.08	81.62	81.33	84.53
$ZM_{component}$	74.29	77.75	81.05	81.22	83.6
LBP	89.11	89.67	93.14	92.89	93.33
LTP	85.85	89.75	92.57	94.22	93.73
ZM_{mag}LBP	93.19	94.83	*97.14*	*97.56*	*98.27*
$ZM_{magPhase}$LBP	*94.59*	94.83	96.67	96.78	96.67
$ZM_{component}$LBP	90.15	92.67	93.91	95.22	96.67
ZM_{mag}LTP	92.96	94.75	96.38	97.44	98.13
$ZM_{magPhase}$LTP	93.11	*95.5*	96.38	97.0	97.6
$ZM_{component}$LTP	91.04	93.92	94.57	95.89	96.53

described earlier, among the individual approaches, the performance of the LBP/LTP descriptors is better than that of the ZM descriptor. Between these two descriptors, the LBP descriptor generates higher recognition rate against the illumination and expression variations on YALE 3.1, YALE 3.2, and YALE 4.2 categories while the LTP descriptor gives higher results on only YALE 4.1 category for expression variation.

For illumination variation, i.e., on YALE 3.1 category, the highest recognition rate of 91.67% is achieved by the proposed ZM_{mag}LBP approach, whereas against expression variation, the $ZM_{magPhase}$LBP descriptor gives the highest recognition rate at 85.56% on YALE 4.1 category. Experiments are also conducted on YALE 3.2 category wherein all of the face images consisting of expression variation are taken in the training set and the remaining ones, i.e., one neutral and four images with illumination changes, are placed in the test set. The results for this setup are also shown in Table 6, from which it is observed that the performance of the ZM_{mag}LBP as well as that of $ZM_{component}$LBP is better. Particularly, on YALE 3.2

category, a superior recognition rate of 97.33% is achieved by both approaches. Similarly, in case of YALE 4.2 category, four images of each person consisting of illumination variation are used to create the training set while all of the remaining ones (i.e., one neutral and six images in varying expressions) are placed in the test set. As shown in Table 6, $ZM_{magPhase}$LBP achieves a high recognition rate of 98.89% for this category. Thus, from the results shown in Table 6, it can be concluded that ZM_{mag}LBP is illumination invariant and $ZM_{magPhase}$LBP is expression invariant. If we look at the overall performance of the proposed approaches on Yale database, ZM_{mag}LBP and $ZM_{magPhase}$LBP outperform the other combinations.

5.3 Performance on ORL database
The ORL face database consists of a total of 400 images of size 112×92 pixels of 40 persons with ten images per person in different states of variation. All the face images in this database are taken against a dark homogenous background. These images contain slight pose variation (tilt and yaw) up to ±20° with some basic facial

Table 6 Performance of the individual/combined approaches over illumination and expression variation on Yale database

Descriptor	YALE 3.1	YALE 3.2	YALE 4.1	YALE 4.2
ZM_{mag}	46.67	80.0	47.78	82.22
$ZM_{magPhase}$	61.67	84.0	54.44	81.11
$ZM_{component}$	53.33	81.33	50.0	81.11
LBP	83.33	85.33	66.67	96.67
LTP	76.67	76.0	72.22	96.67
ZM_{mag}LBP	*91.67*	*97.33*	84.44	97.78
$ZM_{magPhase}$LBP	88.33	96.0	*85.56*	*98.89*
$ZM_{component}$LBP	88.33	*97.33*	77.78	94.44
ZM_{mag}LTP	90.0	96.0	77.78	97.78
$ZM_{magPhase}$LTP	90.0	96.0	82.22	97.78
$ZM_{component}$LTP	88.33	94.67	72.22	96.67

Figure 7 Sample cropped face images (a-j) for one person from ORL database.

expressions (smiling/not smiling, open/closed eyes). For performing experiments, the images of this database are cropped to 64×64 pixels. Sample cropped images for one person are shown in Figure 7. The face images of these databases are partitioned into 64 patches of 8×8 pixels to extract the local LBP/LTP features. Detailed experiments are performed on this database in order to analyze robustness of the proposed combined approaches against the pose variation. Various data partitions generated for this purpose are presented in Table 7.

Firstly, experiments are performed on ORL 1 category by taking one image of each person in the training set, and all of the remaining ones are used to formulate the test set. Different trials are framed in this case. As there are nine different images in the test set, ten combinations of different training and test images are possible here. The average recognition results over these ten different trials are shown in Figure 8a,b. The results on different values of n_{max} are depicted in order to analyze the effect of maximum order of moments n_{max} of ZMs on the performance of the proposed combined approaches. As the basic LBP and the LTP descriptors used in this work are not invariant to image rotation whereas the ZM descriptor is an established rotation invariant scheme, it is observed from the results that the performance of the individual ZM descriptors is better than that of the LBP/LTP descriptors for this database. Among the ZM-based descriptors, $ZM_{component}$ and $ZM_{magPhase}$ descriptors give the highest recognition rates because of the inclusion of phase coefficients. However, an improvement of more
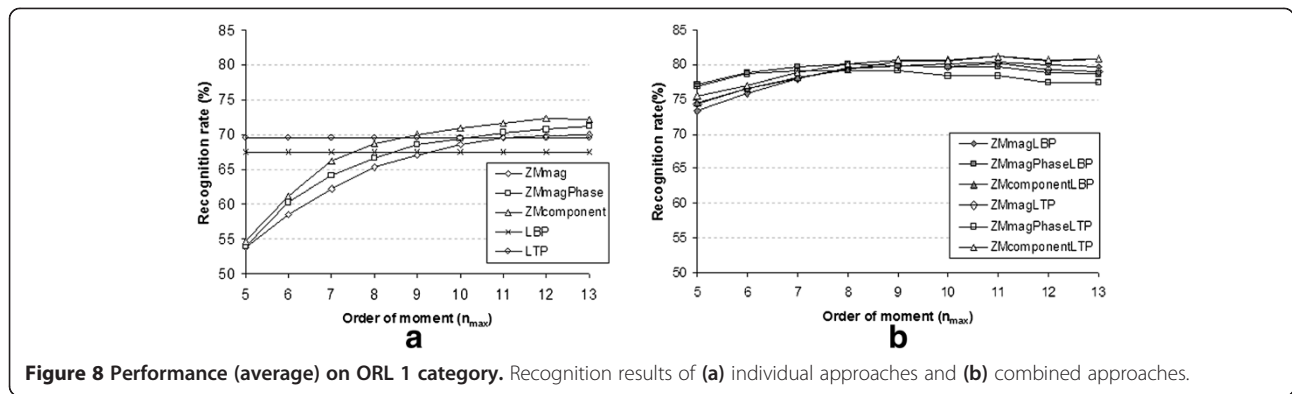
than 10% is achieved by fusion of the invariant feature sets of the ZM and LBP/LTP descriptors wherein the ZM descriptor plays a significant role in achieving rotation invariance. The highest recognition rate of 81.22% is achieved by the proposed $ZM_{mag}LTP$ approach. From the results presented in Figure 8a,b, it is observed that in the proposed combined methods, fusion of the ZM features obtained at $n_{max} = 9$ provides better results on this database. Accordingly, further experiments have been conducted only on this order of moments.

The average recognition results over ten different trials of each group (i.e., ORL 2.1, ORL 2.2, ORL 2.3, and ORL 2.4) of the training and the test sets are presented in Table 8. Excellent results are obtained by the proposed combined approaches, while $ZM_{component}LTP$ provides the best results. On taking five images in training and the remaining five in the test set (i.e., for ORL 2.4 over ten runs), the average recognition rate of 99.2% is achieved by both the $ZM_{mag}LBP$ and $ZM_{component}LTP$ approaches. Further, $ZM_{component}$ features have proven to be invariant to image rotation and tolerant to pose variations to some extent [29]. From this analysis, we can state that the $ZM_{component}$ combined with LTP as well as the ZM_{mag} coupled with LBP provides superior results against pose variations.

Next, experiments are performed on ORL 3 category by taking two neutral face images in the training set, while four images of each person consisting of scale and up/down head movement are taken in the test set. Similarly, in order to examine the performance of the

Table 7 Data partition on ORL database for performing various experiments

Set of experiment	Category	Training	Testing	Remarks
1	ORL 1	Random one	All remaining (i.e., remaining ten images except the one selected for the training)	Ten different trials of this setup have been taken, and the recognition result (Figure 8) is the average of all these trials
2	ORL 2.1	Random two	Remaining eight	The recognition result (shown in Table 8) is the average of ten random trials on each setup, i.e., ORL 2.1, ORL 2.2, ORL 2.3, and ORL 2.4
	ORL 2.2	Random three	Remaining seven	
	ORL 2.3	Random four	Remaining six	
	ORL 2.4	Random five	Remaining five	
3	ORL 3	a, b	c, g, i, j	Testing consists of experiments against scale and up/down (tilt) pose variation
4	ORL 4	a, b	b, d, e, f	Testing consists of experiments against left/right (yaw) pose variation

Figure 8 Performance (average) on ORL 1 category. Recognition results of **(a)** individual approaches and **(b)** combined approaches.

proposed approaches over yaw pose variation, two neutral images of each person are placed in the training set and four images with slight left/right head movement are placed in the test set, i.e., ORL 4 category. The results obtained from this experimental analysis are also presented in Table 8. On ORL 3 category, the performance of the $ZM_{component}$ coupled with that of the LBP/LTP descriptors is better, achieving a recognition rate of 90.0%. Similarly, on ORL 4 category, the highest recognition rate of 91.25% is achieved by both the ZM_{mag}LBP and the $ZM_{magPhase}$LBP approaches. Thus, in most of the cases of ORL database, the $ZM_{component}$ combined with LBP/LTP outperforms the other proposed combinations.

5.4 Performance analysis against noise variation

To examine the effect of noise on the recognition accuracy, we add impulsive noise, commonly named salt-and-pepper or spike noise, to the face images of the three databases. In the presence of impulsive noise, an image has dark pixels in bright regions and white pixels in dark regions [30]. In this analysis, a noise of 0.05 is added to the images of the test set whereas the training is done on original face images, i.e., on images with no noise.

The procedure of experimental setup to examine the performance of these approaches against additive noise is the same as before. That is, in order to analyze the performance on FERET database, experiments are performed on FERET_A3 data partition in which one frontal image (0° pose) is selected in the training set and the four images in different poses (±22.5°, ±67.5° and with additive noise) for each person are used in the test set. On YALE 2.4 data partition, robustness of the proposed approaches is analyzed against noise variation by selecting five images of each person in the training set and the remaining six images (with additive noise) in the test set. The results presented are the average recognition rates over ten different runs of training and test sets. In a similar manner, the experiments on the images of ORL 2.4 data partition are performed by taking random five images of each person in the training set and the remaining five images (with additive noise) in the test set, and the recognition results for the same are also the averaged recognition rates over ten different runs of training and test sets. The experimental results on the said databases are shown in Table 9.

From the results presented, it is observed that among the individual approaches, the LTP descriptor is more

Table 8 Performance of the considered approaches against different pose variations on ORL database

Descriptor	ORL 2.1	ORL 2.2	ORL 2.3	ORL 2.4	ORL 3	ORL 4
ZM_{mag}	81.87	88.71	93.08	94.55	81.88	80.63
$ZM_{magPhase}$	83.5	89.18	94.21	95.3	85.63	81.88
$ZM_{component}$	84.81	90.82	94.96	95.95	84.38	83.75
LBP	82.23	88.75	92.92	95.7	80.0	84.38
LTP	83.28	89.14	92.21	95.4	82.5	86.25
ZM_{mag}LBP	91.24	94.96	97.5	*99.2*	88.75	*91.25*
$ZM_{magPhase}$LBP	91.34	94.57	96.71	98.25	88.75	*91.25*
$ZM_{component}$LBP	*92.47*	95.07	97.5	98.9	*90.0*	88.75
ZM_{mag}LTP	90.69	94.39	97.08	98.85	88.13	88.13
$ZM_{magPhase}$LTP	90.0	93.64	96.5	98.35	86.88	90.0
$ZM_{component}$LTP	92.17	*95.21*	*97.79*	*99.2*	*90.0*	88.13

Table 9 Performance of the considered approaches against noise variation

Descriptor	Face datasets comprising additive noise in testing		
	FERET_A3	YALE 2.4	ORL 2.4
ZM_{mag}	62.75	80.22	94.65
$ZM_{magPhase}$	65.75	75.33	94.45
$ZM_{component}$	64.75	80.11	95.55
LBP	64.75	91.0	95.45
LTP	67.75	94.0	95.85
$ZM_{mag}LBP$	80.0	96.22	98.55
$ZM_{magPhase}LBP$	*81.5*	94.56	98.0
$ZM_{component}LBP$	77.0	93.33	98.85
$ZM_{mag}LTP$	*81.0*	*96.44*	98.85
$ZM_{magPhase}LTP$	79.25	96.11	98.45
$ZM_{component}LTP$	79.0	94.33	*99.0*

robust to noise variation than the LBP. On Yale and ORL databases (with noise variation), the performance of the proposed $ZM_{mag}LTP$ and $ZM_{component}LTP$ descriptors, respectively, is better as compared to all other combined approaches. On the other hand, if FERET images with noise variations are assessed, then the recognition rate of $ZM_{magPhase}LBP$ is 81.5%, whereas the recognition rate of $ZM_{mag}LTP$ is 81.0%. The percentage difference between the actual results obtained (without adding noise and with noise) for both approaches is 5 and 3.5%, respectively. Hence, from this observation, we can say that on FERET database, the performance of the proposed $ZM_{mag}LTP$ descriptor is better against noise variation. For the case of Yale and ORL databases, the degradation due to noise in recognition rates is very less.

5.5 Time complexity

One of the important issues involved in using combined approaches similar to the ones proposed here is the time complexity of these approaches. It is a common perception

Table 10 Dimensionality of the feature vectors of the ZM and LBP/LTP descriptors

Approach	Size of the feature vector	Remarks
ZM_{mag}	40	For moments up to order $n_{max} = 11$
$ZM_{magPhase}$	2×40	Size of the feature vector is double the size of magnitude features
$ZM_{component}$	2×40	Size of the feature vector is double the size of magnitude features
LBP	3,776	Taking image patch size of 8×8 pixels for uniform local binary patterns
LTP	$2 \times 3,776$	Number of features is double the number of LBP features because two feature vectors consisting of the positive and the negative uniform binary patterns are taken

that the moment-based descriptors are computation intensive which is true to some extent especially in case of the ZM calculation. The time complexity of the ZMs is of order $O(N^2 n_{max}^3)$ if all moments up to a maximum order n_{max} are computed for an image of $N \times N$ pixels. However, with the use of fast algorithms [10,11], the time complexity is reduced to $O(N^2 n_{max}^2)$. Further significant reduction in computation time is achieved by using symmetry/antisymmetry properties of kernel function of ZMs. The ZMs of the database images are computed offline and indexed with the images themselves. The ZMs of the test image are computed online. Although the time complexity of ZM calculation is still high, in this work, better recognition results have been obtained with $n_{max} = 9$ for FERET and ORL databases whereas Yale database exhibits good results by taking $n_{max} = 11$; therefore, we consider moments only up to these orders. As $Z_{0,0}$ and $Z_{1,1}$ have no discriminative capabilities, they do not affect the recognition rate. Hence, with $n_{max} = 11$, we have 40 features after discarding the coefficients $Z_{0,0}$ and $Z_{1,1}$. In contrast, although the number of features in the feature vector containing local histogram features of the LBP/LTP descriptors is high, the computation time of these descriptors is very low. Thus, the proposed fusion of the ZM and LBP/LTP descriptors maintains a good balance between speed and dimensionality. The size of feature vectors of the ZM and LBP/LTP descriptors is shown in Table 10.

We observe that for an image of 256×256 pixels, the CPU elapse time for calculating ZMs is only 0.032 s for $n_{max} = 12$ on a PC with a 3.0-GHz CPU and a 1-GB RAM under Microsoft Windows environment. The time

Table 11 Performance comparison (%) of some recent approaches with proposed methods on Yale and ORL databases

Method	ORL	Yale
PCA	55.0	55.0
2DPCA	54.0	56.0
$(PC)^2A$	57.0	55.0
$E(PC)^2A$	58.0	57.0
$2D(PC)^2A$	56.0	55.0
SVD perturbation	61.0	62.0
Hybrid Fourier-AFMT transform	76.0	52.0
RES [40]	0.77	0.78
$ZM_{mag}LBP$	*84.44*	*86.67*
$ZM_{magPhase}LBP$	84.17	*86.67*
$ZM_{component}LBP$	84.17	80.67
$ZM_{mag}LTP$	80.0	82.67
$ZM_{magPhase}LTP$	82.78	85.33
$ZM_{component}LTP$	81.22	81.52

Table 12 Performance comparison (%) of the proposed approaches with recent methods on Yale and ORL databases

Method	Yale database	ORL database
Two-dimensional LDA (2D-LDA) [42]	86.57	92.50
2D-WLDA [42]	88.00	93.50
2D-DWLDA [42]	89.33	94.00
Direct LDA (DLDA) [21]	93.20	92.50
Enhanced Fisher linear discriminant model (EFM) [21]	93.90	92.50
Intrinsicfaces [43]	74.00	97.00
Combined feature Fisher classifier (CF^{2}C) [21]	96.90	96.80
Feature Fisher classifier (F^{3}C) [20]	96.4	94.9
Block based S-P[a] [25]	100.0[a]	99.0
Algorithm A (WMs) [41]	–	93.5
Algorithm B (CWMs) [41]	–	96.0
ZM$_{mag}$LBP	*97.56*	*99.20*
ZM$_{magPhase}$LBP	96.78	98.25
ZM$_{component}$LBP	95.67	98.95
ZM$_{mag}$LTP	*97.56*	98.85
ZM$_{magPhase}$LTP	97.00	98.35
ZM$_{component}$LTP	97.00	*99.20*

Comparison of recognition performance for five random training images. [a]The results are presented on only one random set of five images in the training set and all the remaining in the test set.

taken for computing LBP and LTP features is 0.015 and 0.016 s, respectively. Thus, the total time elapsed for the extraction of the local and global features of a test image does not exceed 0.048 s. The time taken for classification is much less than the feature extraction time. Thus, in comparison to the gain in the recognition performance, the time taken by the combined features is much less and can be afforded by the low computation power devices in online mode. Since the time complexity does not depend on the contents of the image, these experiments are carried out for one image only.

5.6 Performance comparison

We have compared the performance of the proposed combined descriptors with other popular methods such as PCA, 2DPCA, (PC)^{2}A, E(PC)^{2}A, 2D(PC)^{2}A, SVD perturbation [38], and hybrid Fourier-AFMT transform [39] for face recognition with single (first) example image per person. As shown in Table 11, the proposed combined descriptors give the best recognition rate when compared with other well-established methods. On the other hand, the time complexity of PCA-based methods is very high as compared to the proposed approaches.

Comparison of performance of the proposed combined descriptors with other popular methods for face recognition with single (first) example image per person.

Dual optimal multiband features (DOMF) [24] give a recognition rate of 92.6 and 88.4% on Yale and ORL

databases, respectively, when two images of each person are taken in the training set and all the remaining are kept in the test set. On this similar setup for training and test images, the highest recognition rate achieved by the proposed ZM$_{magPhase}$LBP descriptor for YALE 2.1 is 94.59% while the ZM$_{component}$LBP descriptor achieves a recognition rate of 92.47% for ORL 2.1 category.

The performance of the proposed combined approaches is also compared with that of some recent face recognition methods when five images of each person are used for training. The recognition results of the proposed combined approaches and those of these recent methods on Yale and ORL databases for this case are shown in Table 12. The best results are highlighted in italics. All these methods use multidimensional features or combined approaches to represent the face images. As can be seen from the results presented, the recognition rate of the proposed approaches is higher as compared to that of the recent methods. In case of block-based S-P approach [25], one random set of five images per person is taken in the training set while all the remaining are kept in the test set for both the Yale and ORL databases, whereas the results presented for our proposed approaches are the average of ten random trials of training and test sets. It is worth mentioning here that on some of the random trials, our proposed descriptors also provide 100% recognition rate. Recently introduced wavelet moment (WM) and complex WM (CWM) approaches

[41] have achieved a recognition rate of 51.5 and 54.3%, respectively, on FERET_A2 subset, while the proposed $ZM_{magPhase}LBP$ descriptor has attained a recognition rate of 77.33%. On the *fafb* subset of FERET database, the recognition rate obtained by the RES [40], WM, and CWM [41] approaches is 95.0, 88.0, and 91.0% whereas the highest recognition rate achieved by the proposed $ZM_{magPhase}LBP$ approach is 98.04%. Thus, on the basis of superior results obtained by the proposed fusion technique, it can be concluded that combining the feature sets of the ZM and LBP/LTP descriptors is an efficient and practical approach for robust face recognition.

6 Conclusions

This paper proposes the fusion of two useful feature sets, i.e., the global ZMs and the local LBP/LTP descriptor. Face images capture extensive variation under varying pose and lighting conditions accompanied by the presence of expression and noise. Individually, the ZM and LBP/LTP descriptors are observed to be very effective in providing good recognition performance on the face images containing certain variations. In particular, the ZM descriptor extracts rotationally invariant shape features from the whole face images, whereas the LBP/LTP descriptors are able to capture the fine details and illumination-invariant characteristics within some local regions of the face images. However, the fusion of these two complementary approaches incorporates the benefits of both of these descriptors and as such proves to be invariant against various distortions present in the face images. Herein this work, diverse feature sets of ZMs are combined with LBP/LTP descriptors to generate various combined approaches, namely, $ZM_{mag}LBP$, $ZM_{magPhase}LBP$, $ZM_{component}LBP$, $ZM_{mag}LTP$, $ZM_{magPhase}LTP$, and $ZM_{component}LTP$. From the detailed experiments performed on FERET, Yale, and ORL face databases, it has been observed that the proposed combined approaches are highly robust against pose, expression, illumination, and noise variations, as the recognition rate achieved by the proposed approaches is approximately 10 to 30% higher than that obtained by applying these approaches individually. Fusion of ZM and LBP descriptor performs better over the pose, expression, and illumination variations, while in the presence of noise, ZMs combined with LTP descriptor generate superior results. Experimental results also prove the efficacy of the proposed methods over other existing techniques. Also, significant improvement in the recognition rate is achieved by the proposed scheme when only single training image per person is available.

Future work is suggested towards discovering the optimal ways to utilize the information acquired by the phase coefficients of ZM descriptor in addition to using different methods of classification to further improve the performance of the proposed fusion approach.

Author details
[1]Department of Computer Science, Punjabi University, Patiala 147002, India. [2]Central Scientific Instruments Organisation, Sector 30-C, Chandigarh 160030, India. [3]Department of Computer Science, South Asian University, Akbar Bhawan, Chanakyapuri, Delhi 110021, India.

References
1. W Zhao, R Chellappa, P Phillips, A Rosenfeld, Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)
2. E Hjelmas, BK Low, Face detection: a survey. Comput. Vision. Image. Underst. **83**, 236–274 (2001)
3. M Turk, A random walk through Eigenspace. IEICE Trans. Inf. Syst. **E84-D**(12), 1586–1595 (2001)
4. N Mittal, E Walia, Face recognition using improved fast PCA algorithm, in *Proceedings of the IEEE International Congress on Image and Signal Processing (CISp '08)* (Sanya, Hainan, 2008), pp. 554–558
5. PN Belhumeur, JP Hespanha, DJ Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 711–720 (1997)
6. Y Xu, D Zhang, J Yang, Y J–Y, An approach for directly extracting features from matrix data and its application in face recognition. Neurocomputing **71**, 1857–1865 (2008)
7. Z Daoqiang, Z Zhi-Hua, (2D)^{2}PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing **69**, 224–231 (2005)
8. D Zhang, G Lu, Review of shape representation and description techniques. Pattern Recognit. **37**(1), 1–19 (2004)
9. D Zhang, G Lu, Evaluation of MPEG-7 shape descriptors against other shape descriptors. Multimed. Syst. **9**, 15–30 (2003)
10. C Singh, E Walia, Fast and numerically stable methods for the computation of Zernike moments. Pattern Recognit. **43**(7), 2497–2506 (2010)
11. W C–Y, R Paramesran, On the computational aspects of Zernike moments. Image Vis. Comput. **25**, 967–980 (2007)
12. DG Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
13. H Soyel, H Demirel, Facial expression recognition based on discriminative scale invariant feature transform. IET Electron. Lett. **46**(5), 343–345 (2010)
14. L Huang, A Shimizu, H Kobatake, Robust face detection using Gabor filter features. Pattern Recognit. Lett. **26**(11), 1641–1649 (2005)
15. T Ojala, M Pietikäinen, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
16. T Ahonen, A Hadid, M Pietikäinen, Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
17. B Jun, T Kim, D Kim, A compact local binary pattern using maximization of mutual information for face analysis. Pattern Recognit. **44**, 532–543 (2011)
18. X Tan, B Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Process. **19**(6), 1635–1648 (2010)
19. C Kim, J Oh, C Choi, *Combined Subspace Method Using Global and Local features For Face Recognition*, 4th edn. (Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '05), Montreal Canada, 2005), pp. 2030–2035

20. D Zhou, X Yang, Feature fusion based face recognition using EFM. Proceedings International Conference on Image Analysis and Recognition (ICIAR '04). Lect. Notes Comput. Sci. **3212**, 643–650 (2004)

21. D Zhou, X Yang, N Peng, Y Wang, Improved-LDA based face recognition using both facial global and local information. Pattern Recognit. Lett. **27**, 536–543 (2006)

22. C Singh, E Walia, N Mittal, Robust two-stage face recognition approach using global and local features. Vis. Comput. **28**(11), 1085–1098 (2012)

23. Y Su, S Shan, X Chen, W Gao, Hierarchical ensemble of global and local classifiers for face recognition. IEEE Trans. Image Process. **18**(8), 1885–1895 (2009)

24. Y-W Wong, KP Seng, A Li-M, Dual optimal multiband features for face recognition. Expert Syst, Appl **37**(4), 2957–2962 (2010)

25. ME Aroussi, ME Hassouni, S Ghouzali, M Rziza, D Aboutajdine, Local appearance based face recognition method using block based steerable pyramid transform. Signal Process **91**, 38–50 (2011)

26. Z Liu, C Liu, Fusion of color, local spatial and global frequency information for face recognition. Pattern Recognit. **43**, 2882–2890 (2010)

27. B Jun, J Lee, D Kim, A novel illumination-robust face recognition using statistical and non-statistical method. Pattern Recognit. Lett. **32**, 329–336 (2011)

28. S Moore, R Bowden, Local binary patterns for multi-view facial expression recognition. Comput. Vision Image Underst. **115**, 541–558 (2011)

29. C Singh, E Walia, N Mittal, Rotation invariant complex Zernike moments features and their application to human face and character recognition. IET Comput. Vision **5**(5), 255–265 (2011)

30. SM Lajevardi, ZM Hussain, Higher order orthogonal moments for invariant facial expression recognition. Digit Signal Process **20**, 1771–1779 (2010)

31. S Li, L M–C, P Chi-Man, Complex Zernike moments features for shape based image retrieval. IEEE Trans. Syst. Man. Cybern. C Appl. Rev. **39**, 227–237 (2009)

32. C Singh, N Mittal, E Walia, Face recognition using Zernike and complex Zernike moment features. Pattern Recognit. Image Anal. **21**(1), 71–81 (2011)

33. J Revaud, G Lavoue, A Baskurt, Improving Zernike moments comparison for optimal similarity and rotation angle retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 627–636 (2009)

34. A Jain, K Nandakumar, A Ross, Score normalization in multimodal biometric systems. Pattern Recognit. **38**, 2270–2285 (2005)

35. The Facial Recognition Technology (FERET) face database. http://www.nist.gov/itl/iad/ig/colorferet.cfm

36. Yale face database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

37. Olivetti Research Laboratory (ORL) face database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

38. J Li, J-S Pan, A novel pose and illumination robust face recognition with a single training image per person algorithm. Chin. Optic. Lett. **6**(4), 255–257 (2008)

39. YM Chen, J-H Chiang, Fusing multiple features for Fourier Mellin-based face recognition with single example image per person. Neurocomputing **73**(16–18), 3089–3096 (2010)

40. C-H Kuo, JD Lee, Face recognition based on a two-view projective transformation using one sample per subject. IET Comput. Vision **6**(5), 489–498 (2012)

41. C Singh, AM Sahan, Face recognition using complex wavelet moments. Opt. Laser Technol. **47**, 256–267 (2013)

42. R Zhi, Q Ruan, Two-dimensional direct and weighted linear discriminant analysis for face recognition. Neurocomputing **71**, 3607–3611 (2008)

43. Y Wang, Y Wu, Face recognition using Intrinsicfaces. Pattern Recognit. **43**, 3580–3590 (2010)