

RESEARCH

Open Access



Bayesian face recognition using 2D Gaussian-Hermite moments

S. M. Mahbubur Rahman^{1*}, Shahana Parvin Lata² and Tamanna Howlader²

Abstract

This paper presents a statistical face recognition algorithm by expressing face images in terms of orthogonal two-dimensional Gaussian-Hermite moments (2D-GHMs). Motivation for developing 2D-GHM-based recognition algorithm includes the ability of these moments to capture higher-order hidden nonlinear 2D structures within images and the invariance of certain linear combinations of moments to common geometric distortions in images. The key contribution of this paper is that features of 2D faces are represented in terms of a statistically selected set of 2D discriminative GHMs (DGHMs) as opposed to commonly chosen heuristic set of first few order moments only. In particular, the intraclass correlation coefficient for the entire set of moments of the training images are used to select only a desired set of moments that maximize the discrimination among available classes. The naive Bayes classifier that yields optimal performance in many statistical applications is used for identification due to the simplicity of its implementation for handling huge size face database. Experiments are conducted to evaluate the performance of the proposed recognition algorithm on exhaustive databases such as the AT&T, Face Recognition Grand Challenge (FRGC), Face Recognition Technology (FERET), Labeled Faces in the Wild (LFW), and YouTube that possess face images or videos with significant variations in terms of appearance, occlusion, expression, pose, resolution, and illumination both in the constrained and unconstrained environments. In the constrained condition, comparisons with the well-established 2D-principal component analysis, 2D-linear discriminant analysis, and 2D-canonical correlation analysis methods as well as orthogonal 2D-Krawtchouk moment-based method reveal the superior performance of the proposed method in terms of recognition accuracy for varying numbers of training and probe images. The proposed DGHM features also show superior recognition or verification performance on the standard protocols of the unconstrained face databases when comparing with the commonly referred descriptors such as the local binary pattern or scale-invariant feature transform.

Keywords: Intraclass correlation coefficient, Gaussian-Hermite image moments, Naive-Bayes classifier, Statistical pattern recognition

1 Introduction

Over the past two decades, biometric security systems that allow identification of individuals using their physiological or behavioral traits, have become an integral part of many security-aware applications due to the increasing demand for “identity verification”. By overcoming the spoofing vulnerability of the traditional methods of authenticity verification such as passwords, PIN, or ID cards, biometric security systems have become indispensable tools in financial transactions, access control, and

surveillance. Among various existing biometric methods, automated face recognition is often preferred due to its nonintrusive nature, high level of machine compatibility, and positive attitude of the public [1]. In most of the existing face recognition algorithms, the identification of a person is performed using the pixels of two-dimensional (2D) face images captured by ubiquitous CCD-based visible light sensors. The video-based face recognition focuses on the identification of a person in the noncooperative environment by using the chronological variations in the appearances of 2D face images over the frames [2]. Instances may also be found in the literature with regard to the 3D face recognition algorithms, wherein identification is performed considering the information of light

*Correspondence: mahbubur@eee.buet.ac.bd

¹ Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, 1205 Dhaka, Bangladesh
Full list of author information is available at the end of the article

intensities captured by pixels of 2D face images along with the depth information of each pixel. Notably, the 3D face recognition techniques have two major difficulties with regard to the acquisition and processing of images [3]. In the acquisition process, the depth information of an image is captured using multiple sensors such as the CCD and laser sensors in which case the latter provide much lower resolution than the former, and thus, registration to obtain 3D image becomes complex. A number of CCD sensors may provide depth information, but the focus adjustment of camera lenses and maintaining spatial distances among cameras requires extreme precision. In addition to complexity of the acquisition process, 3D face recognition involves significant memory requirements, and computational load while performing a one-to-many searching task as compared to the 2D counterparts. It is for these very reasons that commercial face recognition systems, which deal with large numbers of people in busy places, still use the popular 2D face recognition techniques and research centers on ways to improve the identification performance of such methods. In the research studies, statistical methods are being increasingly used to analyze the random variations of 2D spatial distribution of pixel intensities among the face images of individuals or subjects and to develop efficient face recognition algorithms.

Face recognition consists of two major steps that rely on statistical methods: low-dimensional feature representation for face images and identification of the class of the face image in question based on the chosen set of features. There are two principal reasons why low-dimensional feature representation needs to be performed for face recognition. First, in a constrained face recognition algorithm, the face image is a high-dimensional data set that contains extreme information redundancy. For example, a 2D image having pixels of size 100×100 can be seen as a point in a 10^4 -dimensional feature space. In the unconstrained face recognition, the deep learning-based matching methods that use the densely sampled descriptors such as the local binary patterns (LBPs) or manifolds may even require the dimension of features of size as high as 100 K. In general, the classification accuracy obtained from such a high-dimensional feature space is restricted due to the fact that the number of training samples available is usually much smaller. Further, the face recognition task carried out in the high-dimensional feature space is also constrained due to the storage requirements and computational power. This problem, known as the “curse of dimensionality”, is very often tackled by the use of low-dimensional feature extraction of the entire set of pixels or sparse regression of high-dimensional local features [4]. Secondly, the pixels of different images of the same person may vary widely due to variations in expression, pose, viewing angle, illumination, and age, while the pixels of different persons may not vary significantly [5–7]. In such

a case, the within-person variability very often becomes larger than the between-person variability making the recognition task difficult in pixel domain. Thus, the design of an efficient face recognition algorithm requires judicious construction of a feature vector that maximizes the between-person variability, while minimizing the within-person variability across 2D spatial coordinates of an image as well as changing conditions of image capturing. In the literature, there are two approaches that have been used to tackle these problems. One approach is to use a suitable set of landmarks consisting of noticeable facial parts such as the eyes, nose, and mouth as well as the relative locations and the statistics of local neighboring pixels of these landmarks to construct feature vectors containing relevant and nonredundant information [8, 9]. The other approach, termed the holistic/global/appearance-based approach, treats the entire face image as a point in high-dimensional space [7]. In this practice, statistical dimension reduction techniques are used to represent the entire face images in terms of their projections into a low-dimensional space so that the constructed features capture the important characteristics of faces to be recognized [10, 11]. In general, the holistic approach is preferred to the landmark approach due to the fact that the former preserves the interrelations among the entire facial regions, whereas the latter considers just a few regions for recognizing the identity of an individual.

Traditional holistic face recognition algorithms are developed using the classical statistical techniques such as the principal component analysis (PCA), linear discriminant analysis (LDA), and canonical correlation analysis (CCA) [12]. For instance, the popular Eigenfaces method uses the PCA to reduce the dimensionality by identifying a small number of directions that capture the majority of variations in the images [13]. The Fisherfaces method [14], which is related to the well-known multivariate analysis of variance (ANOVA) framework in statistics [12], reduces the dimensionality by using a linear projection of the images that maximizes the ratio between the inter-class image variation to the intraclass variation. Methods have also been developed using the CCA approach, which allows two projection spaces—one for the training images and the other for the test images, commonly referred to as the probe images. These methods use a suitable set of canonical variables obtained from the covariance matrices of the training and probe image pairs in the two spaces to identify the subject [12]. Since the inception of the Eigenfaces and Fisherfaces methods, numerous variations and extensions of these methods that include the Eigenface maximum likelihood [15], adaptively weighted subpattern PCA [16], kernel PCA [17], null-space LDA [18], dual-space LDA [19], regularized discriminant analysis [20], boosting LDA [21], Fourier-LDA [22], Gabor-LDA [23], and incremental LDA [24] have been suggested

for face recognition. Most of the existing PCA- and LDA-based classification methods while applied to the images consider the transformation of image matrices to image vectors. Since the linear or nonlinear dependencies within and between classes of the second- or higher-order local structures of the face images may only be taken care of by using the 2D image matrices, processing of image vectors is undesirable. Further, vectorizing the image matrices very often shows a singularity problem during the inversion of scatter matrices in case of LDA-based training. In this context, 2D-PCA [25], symmetrical 2D-PCA [26], 2D-LDA [27], and 2D-CCA [28, 29] have been proposed and shown to be superior in face recognition compared to their 1D counterparts. In order to obtain spatially localized features so that the face recognition methods may be robust to local distortion or occlusion, statistical techniques such as the independent component analysis [30, 31] and factor analysis [32] have also been introduced. Apart from variations in the methods for feature construction, face recognition algorithms also differ in the choice of the classifier. Statistical classification techniques used for recognition include those based on the minimum distance [22], similarity score [33], Kullback Leibler distance [34], maximum a posteriori (MAP) model [15, 32], maximum entropy [30], likelihood dependency test [21], support vector machine [17], and nearest neighbor classifier [10, 35]. A detailed survey on feature extraction techniques and classification methods for both the 2D and 3D face recognition methods may be found in [3, 36]. In spite of the plethora of face recognition algorithms being available in the literature, variations in illumination, expression, resolution, and geometric distortions pose challenges even for the most sophisticated algorithms. Hence, face recognition is an ongoing area of research, wherein better methods for image representation, feature extraction, and classification are still being sought.

In statistics, moments are widely used to describe the shape of probability distributions. However, moments have been used for years to characterize functions, which may or may not be probability density functions (PDFs). The use of moments to characterize image functions dates back to Hu [37], who first introduced the 2D geometric moment invariants for pattern recognition. Later, Teague [38] suggested the notion of moments based on the theory of orthogonal polynomials to develop efficient algorithms for shape analysis and recognition of patterns in images. The recognition efficiency arises due to the fact that these moments are believed to capture higher-order nonlinear structures of the image functions that are obtained from the projection of the higher-order orthogonal polynomials. Surprisingly, the potential of orthogonal moments for feature extraction in the face recognition problem has been investigated only by a few research

groups until recently. For instance, the orthogonal 2D Zernike moments and their variants have been used for face recognition in [39–42]. Rani and Devaraj [43] proposed a method based on the 2D Krawtchouk moments (KCMs) and showed that their method performs well in the presence of noise, tilt, and expression. Among the existing orthogonal moments, the Gaussian-Hermite moments (GHMs) are popular in many visual signal processing algorithms [44], since the width of the Gaussian weight function of the Hermite polynomial expansion provides flexibility in isolating the visual features just as the human visual system does. In face recognition, 1D GHMs have been used to develop 3D face recognition algorithms in which the GHMs of 1D vectors have been used to represent facial features. For example, zeroth- to fourth-order 1D GHMs have been used to describe the shape vector consisting of mesh nodes of face surfaces in [45]. Features are also obtained by using the 1D GHMs up to second-order for representing the depth vectors of selected facial points [46] and the bending invariants of face surfaces in terms of these moments [47]. In all these methods, the order of moments used for feature extraction are chosen heuristically without any statistical justification. A contribution of the approach adopted in this paper is that the orthogonal 2D GHMs are used instead of the 1D GHMs for feature representation in a 2D face recognition algorithm. The development of 2D GHM-based face recognition is motivated by the fact that, recently, some attractive mathematical properties of these moments are obtained. For instance, a linear combination of 2D GHMs may form geometric moments that are invariant to scale, shift, and rotation of a pattern [48, 49]. In addition to these properties, the Hermite polynomials like others may be obtained recursively, and thus, the computation of the 2D GHMs as well as the reconstruction of face images from these moments, if necessary, may be performed very efficiently. These issues motivate us to develop a fast face recognition algorithm that is robust to variations in illumination, resolution, appearance, expression, or pose using the 2D GHMs for feature representation. The major novelty of this paper lies in the fact that the order of the 2D moments used for classification are chosen using a suitable statistical measure instead of adopting a heuristic set of moments to construct the facial features. In particular, the intraclass correlation coefficient (ICC), which is widely used in reliability studies [50], is chosen as a statistical measure of class separability for selecting a subset of the moments called the 2D discriminative GHMs (DGHMs) to construct the feature vectors to be classified. Since the features in the proposed face recognition algorithm are obtained from the orthogonal moments, it is expected that features are independent for a given class. In this regard, the classification is performed using the naive Bayes classifier, which

is a special case of the Bayes classifier that yields the minimum misclassification risk among all statistical classification rules [51]. Further, the implementation of this classifier requires lower memory and computational load and a small amount of training data to estimate the statistical parameters necessary for classification [52]. Hence, the naive Bayes classifier can deal with a large number of features and large data sets making it an ideal choice as the classifier for face recognition problems. To investigate the performance of the proposed DGHM-based recognition method, experiments are conducted considering varying numbers of training and probe images chosen randomly and comparisons made with well-established methods using the standard protocols of commonly referred face databases.

The paper has been organized as follows. In Section 2, image moments are defined, and then, the process of obtaining the 2D GHMs and reconstruction of face images from these moments are described. Section 3 presents the statistical approach for identification of a person using the stored orthogonal moments. In Section 4, the face databases used in the experiments are described, and results of the proposed and existing face recognition methods are presented. Finally, conclusions are drawn in Section 5.

2 Face representation by 2D Gaussian-Hermite moments

Let $I(x, y) \in L_2(\mathbb{R}^2)$ be a continuous square integrable 2D image signal. The set of 2D geometric image moments of order (p, q) ($p, q \in \mathbb{Z}^1$) denoted as M_{pq}^g may be obtained as [53]

$$M_{pq}^g = \int \int_{\mathbb{R}^2} I(x, y) x^p y^q dx dy \quad (1)$$

The analogy of statistical moments to the image moments may be drawn by considering x and y as random variables instead of spatial positions and $I(\cdot)$ as a joint PDF instead of intensity function of the image. Orthogonal image moments are obtained from $I(\cdot)$ in a similar fashion to the geometric moments by using two independent generalized set of polynomial functions $\Psi_p(\cdot)$ and $\Psi_q(\cdot)$ having orders p and q ($p, q \in \mathbb{Z}^1$), respectively, and are given by [53]

$$M_{pq}^\Psi = \int \int_{\mathbb{R}^2} I(x, y) \Psi_p(x) \Psi_q(y) dx dy \quad (2)$$

In this paper, the moments that will be used for face recognition are obtained from the orthogonal Gaussian-Hermite polynomials. Hence, a brief review of Hermite polynomials and their orthogonality relations with the Gaussian weighting function are given first. Next, the method of obtaining the orthogonal GHMs of the face

images from these polynomials and the reconstruction of images from these moments are presented.

2.1 Hermite polynomials: a brief review

The Hermite polynomial of order $p \in \mathbb{Z}^1$ on the real line $x \in \mathbb{R}^1$ is given by [54]

$$H_p(x) = (-1)^p \exp(x^2) \frac{d^p}{dx^p} \exp(-x^2) \quad (3)$$

These polynomials may be computed efficiently using the following recursive relations:

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_{p+1}(x) &= 2xH_p(x) - 2pH_{p-1}(x) \quad p \geq 1 \end{aligned} \quad (4)$$

The Hermite polynomials satisfy the orthogonality property with respect to the weight function $\nu(x) = \exp(-x^2)$ such that

$$\int_{-\infty}^{\infty} \exp(-x^2) H_p(x) H_q(x) dx = 2^p p! \sqrt{\pi} \delta_{pq} \quad (5)$$

where δ_{pq} is the Kronecker delta function. An orthonormal relation may be obtained by using a normalized version of the Hermite polynomials given by

$$\tilde{H}_p(x) = \sqrt{2^p p! \sqrt{\pi}} \exp(-x^2/2) H_p(x) \quad (6)$$

A generalized version of (6) may be obtained by using a spread factor s ($s > 0$) on the real line $x \in \mathbb{R}^1$. In such a case, the so-called generalized Gaussian-Hermite (GH) polynomials may be written as

$$\bar{H}_p(x; s) = \sqrt{2^p p! \sqrt{\pi} s} \exp(-x^2/2s^2) H_p(x/s) \quad (7)$$

for which the orthonormal relation is maintained as

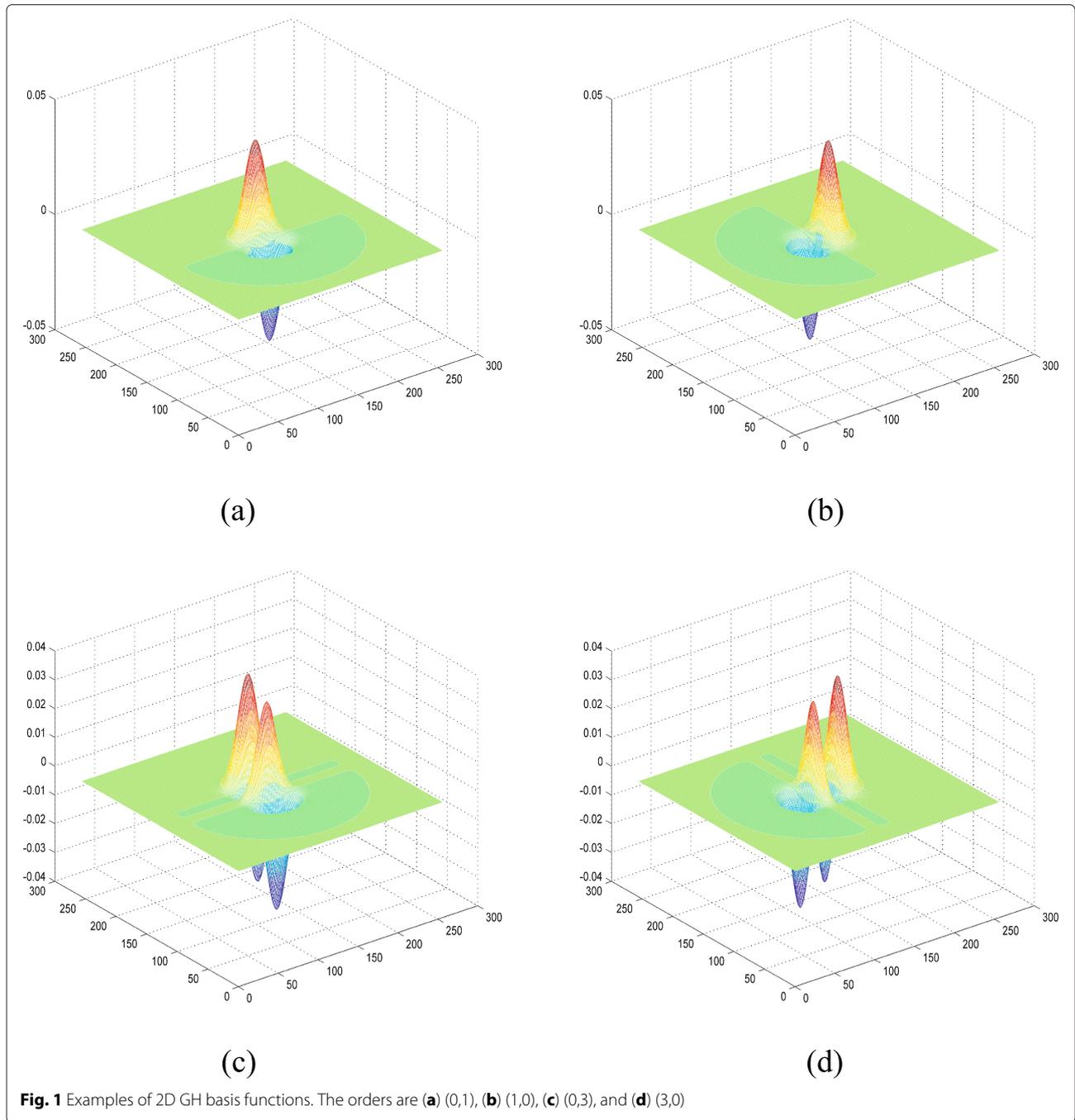
$$\int_{-\infty}^{\infty} \bar{H}_p(x; s) \bar{H}_q(x; s) dx = \delta_{pq} \quad (8)$$

2.2 Gaussian-Hermite moments

The set of 2D GHMs of order (p, q) ($p, q \in \mathbb{Z}^1$) denoted as M_{pq}^{GH} may be obtained from 2D GH basis functions expressed in terms of p -th and q -th order GH polynomials using the following relation [53]:

$$M_{pq}^{GH} = \int \int_{\mathbb{R}^2} I(x, y) \bar{H}_p(x; s) \bar{H}_q(y; s) dx dy \quad (9)$$

In order to improve readability, the superscript of M_{pq}^{GH} will be removed from the remainder of this paper and M_{pq} will be referred to as the 2D GHM of order (p, q) . Figure 1 shows a few number of 2D GH basis functions obtained from the tensor product of two independent 1D GH polynomials. The GHMs of the 2D image signal may be considered as the projections of the signal onto these 2D basis functions. Thus, these moments characterize the image signal at different spatial modes that are defined by certain combinations of the derivatives of the Gaussian



functions. Ideally, from all possible moments, the image $I(x, y)$ may be reconstructed without any error as

$$I(x, y) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} M_{pq} \bar{H}_p(x; s) \bar{H}_q(y; s) \quad (10)$$

It is to be noted that the GHMs are obtained from two real lines $x \in \mathbb{R}^1$ and $y \in \mathbb{R}^1$, and hence, a modification is required for obtaining moments from the discrete coordinates of practically available face images. Let $G^\ell(i, j)$ ($(i, j) \in \mathbb{Z}^2$) be a face image of size $U \times V$ having a

class label ℓ in a face database of K number of identities. Using a similar approach considered in [55], to obtain the 2D GHMs using (9), we have normalized the coordinates such that $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$ by choosing only the following discrete values

$$x = \frac{2i - U + 1}{U - 1} \quad i = 0, 1, 2, \dots, U - 1 \quad (11)$$

$$y = \frac{2j - V + 1}{V - 1} \quad j = 0, 1, 2, \dots, V - 1 \quad (12)$$

In terms of discrete implementation, the 2D moments for the face images of class label ℓ ($\ell = 1, 2, \dots, K$) may be obtained as

$$M_{pq}^\ell = \frac{4}{(U-1)(V-1)} \sum_{i=0}^{U-1} \sum_{j=0}^{V-1} \mathcal{G}^\ell(i, j) \bar{H}_p(i; s) \bar{H}_q(j; s) \tag{13}$$

A crucial choice for obtaining the GHMs is the value of the spread factor s ($s > 0$) in the GH polynomials. Since the support of the discrete image is defined as $(-1, 1)$, and at the same time the modes of the highest order GH polynomials are expected to remain within this support during the implementation, we have chosen the spread as

$$s = \frac{\gamma}{\sqrt{N}} \tag{14}$$

where N is the maximum order of polynomials and γ ($0 < \gamma < 1$) is the normalization factor due to support. For a face image, a value of γ may be chosen close to unity considering the fact that face boundary remains very often close to the image boundary. If a face boundary appears in the central area of the image, then, a smaller value of γ can be chosen. Figure 2 shows the distribution of the magnitudes of the 2D GHMs with respect to orders of GH polynomials for a typical face image wherein γ is chosen as 0.9. It may be seen from this figure that with the increase of the order of the moments from zero, the magnitude of the moments decreases exponentially. Hence, only first few order moments are required for a sufficient good approximation of the face image. In such a case,

the face image may be reconstructed from the moments obtained up to the N -th order GH polynomials as

$$\hat{\mathcal{G}}^\ell(i, j) = \sum_{p=0}^N \sum_{q=0}^N M_{pq}^\ell \bar{H}_p(i; s) \bar{H}_q(j; s) \tag{15}$$

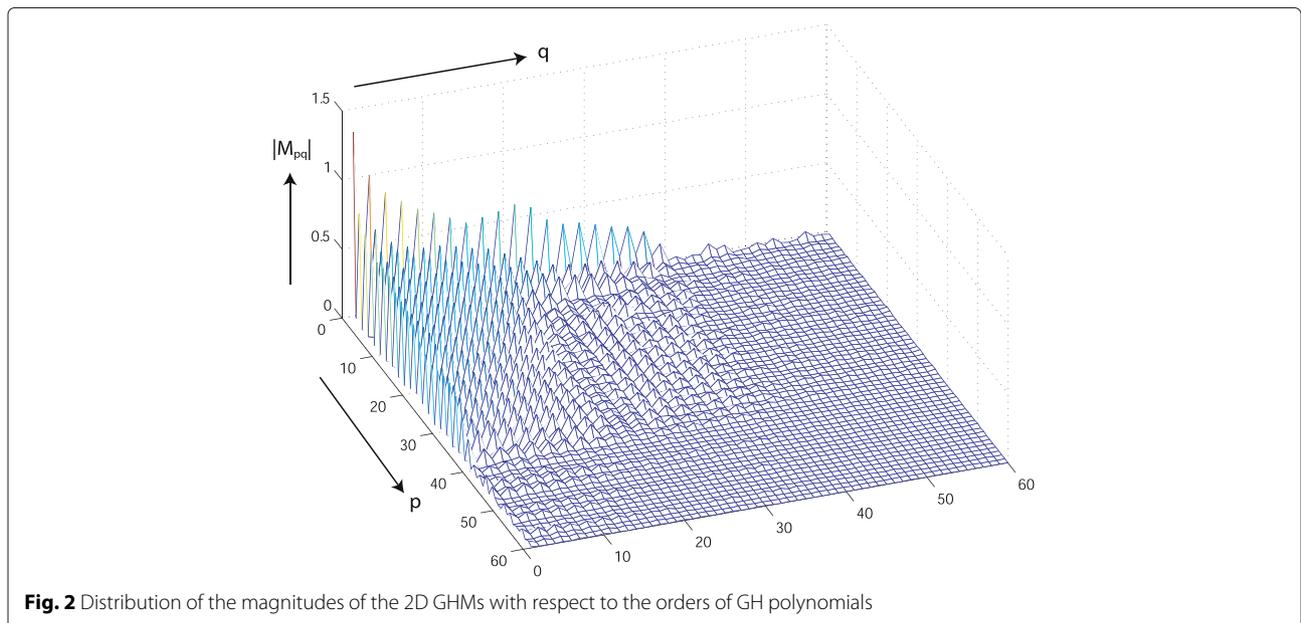
It may be noted that the maximum order of GH polynomials N may be chosen in such a way that the number of moments are only a fraction of the face image data. Let α ($0 < \alpha < 1$) be the compression factor through which a face is stored in terms of the GHMs. In such a case, the maximum order of moments would be

$$N = \lfloor \sqrt{\alpha UV} \rfloor - 1 \tag{16}$$

where $\lfloor z \rfloor$ denotes the largest integer contained in z . Considering the fact that $N \gg 1$, the computational complexity of face image reconstruction from 2D GHMs may be shown to be at most $\mathcal{O}(N^2/\alpha)$ ($0 < \alpha < 1$). The complexity of computation of the GHM can be even reduced further by considering the symmetric property of the 2D GH polynomials [56].

3 Face recognition by 2D Gaussian-Hermite moments

In this section, a face recognition algorithm is developed with the consideration that a face image is stored in terms of the moments M_{pq} ($p, q = 0, 1, 2, \dots, N$). The scale, shift, and rotation invariants of GHMs may be obtained by using certain linear combinations of these moments [49]. In the proposed method, however, the entire set of stored GHMs is considered to construct the feature vector instead of the moment invariants. This is mainly



due to the fact that the invariants are considered indirectly even though the orthogonal moments are treated independently. In addition, no specific type of geometric distortions or occlusions are ensured to exist in the face images of a database. To select the GHMs most significant for face recognition, a statistical learning approach is adopted. In this regard, the most relevant GHMs that give rise to maximum discrimination between the classes are identified. A simple but effective and reliable statistical measure of separability among classes is the ICC [50], which in this scenario is defined as the proportion of total variance accounted for by the between-subject variations of the face classes. ShROUT and Fleiss [50] described six different forms of the ICC which differ according to the study design and the underlying mathematical models. Next, we describe the experimental design and associated ANOVA model of a specific form of ICC for GHMs that is suitable for the face recognition study.

3.1 The intraclass correlation coefficient for moments

In the proposed method, the K distinct subjects or classes within the face database constitute a random sample from a large population of face classes. Let λ_{tr} ($\lambda_{tr} > 1$) be the equal number of training images belonging to class ℓ ($\ell = 1, 2, \dots, K$) that form a random sample from a population of images for that class. Let $M_{pq}^\ell(k)$ ($k = 1, 2, \dots, \lambda_{tr}$) denote the moment of order (p, q) for the k -th image belonging to class ℓ . Then, a suitable linear model for $M_{pq}^\ell(k)$ is the one-way random effects model given by

$$M_{pq}^\ell(k) = \mu_{pq} + b_{pq}^\ell + w_{pq}^\ell(k) \quad (17)$$

where μ_{pq} is the overall mean of the moments considering the entire training database, $b_{pq}^\ell = \mu_{pq}^\ell - \mu_{pq}$, the notation μ_{pq}^ℓ being the mean of the moments within the class label ℓ and $w_{pq}^\ell(k)$ is a residual component. It is considered that $b_{pq}^\ell \sim \mathcal{N}(0, \sigma_{b_{pq}}^2)$ and $w_{pq}^\ell(k) \sim \mathcal{N}(0, \sigma_{w_{pq}}^2)$ as well as they are independent of all components in the model. Normality of the GHMs may be assessed by constructing the commonly used quantile-quantile (Q-Q) plot [57]. Figure 3 shows the Q-Q plots obtained for the GHMs of order (0, 1) and (10, 14) considering seven face images of a typical subject. From this figure, it may be seen that the sample quantiles of the GHMs are approximately linearly related to the quantiles of the normal PDF. It may be mentioned that the Q-Q plots of the moments of other orders for various subjects also follow a similar pattern and are not shown to avoid repetitive results. Thus, the GHMs of the face images may be treated as random variables that follow the Gaussian PDF. It can be shown that under this

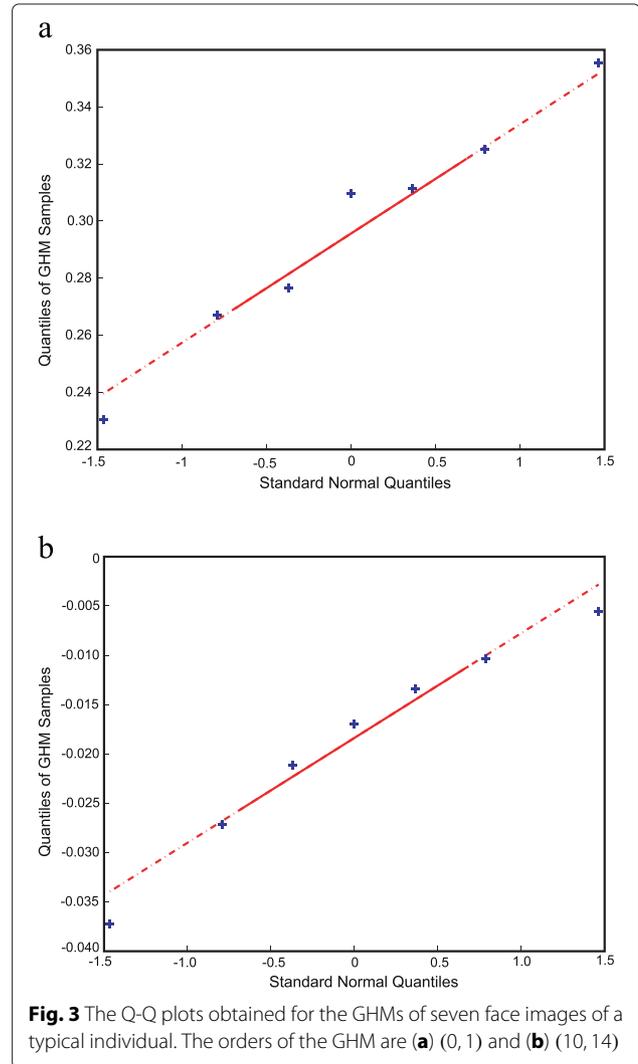


Fig. 3 The Q-Q plots obtained for the GHMs of seven face images of a typical individual. The orders of the GHM are (a) (0, 1) and (b) (10, 14)

setting, the population ICC denoted by ρ_{pq} ($0 \leq \rho_{pq} \leq 1$) for M_{pq} is given by [50]

$$\rho_{pq} = \frac{\sigma_{b_{pq}}^2}{\sigma_{b_{pq}}^2 + \sigma_{w_{pq}}^2} \quad p, q = 0, 1, 2, \dots, N \quad (18)$$

The population ICC can be estimated from the training samples using [50]

$$\hat{\rho}_{pq} = \frac{\Phi_b - \Phi_w}{\Phi_b + (\lambda_{tr} - 1)\Phi_w} \quad p, q = 0, 1, 2, \dots, N \quad (19)$$

where Φ_b and Φ_w represent the “between-class mean-square” and “within-class mean-square”, respectively, and are given by

$$\Phi_b = \frac{\lambda_{tr} \sum_{\ell=1}^K (\mu_{pq}^\ell - \mu_{pq})^2}{K - 1} \quad (20)$$

$$\Phi_w = \frac{\sum_{\ell=1}^K \sum_{k=1}^{\lambda_{tr}} (M_{pq}^\ell(k) - \mu_{pq}^\ell)^2}{K(\lambda_{tr} - 1)} \quad (21)$$

The computational complexity of estimating the Φ_b or Φ_w is $\mathcal{O}(\lambda_{tr}K(N + 1)^2)$. Noting that $N \gg 1$, the computational cost of calculating the ICC becomes $\mathcal{O}(\lambda_{tr}KN^2)$, which is directly proportional to the size of the face database. Since the ICC describes how strongly the GHMs in the same class resemble each other, the moments of the training face images that have high ICC values are consistent within the classes and possess small within-class variability relative to total variability in the training database. These sets of moments having high values of ICC are capable of discriminating between the classes and may therefore be selected for representing the feature vector of each face image.

3.2 The feature vector for face images

The values of the moments M_{pq} are computed using (13) for each of the images in the face database, whereas the ICCs for each of the moments are estimated using (19–21) only for the images of the training set. Thus, a total of $(N + 1)^2$ number of ICCs are computed corresponding to the $(N + 1)^2$ number of GHMs representing each face image. It is to be noted that not all these moments are useful for face classification. Further, the estimated ICCs quantify the discrimination capability of the moments towards identification of classes. Hence, we select as features only those GHMs that correspond to the T ($T \ll N$) largest ICCs, T being the number of moments used for classification. In such a case, the features referred to as the DGHMs for the k -th image of class ℓ may be denoted as $\mathbf{F}_k^\ell = [f_{1k}^\ell, f_{2k}^\ell, f_{3k}^\ell, \dots, f_{Tk}^\ell]'$, where

$f_{rk}^\ell \in \{M_{pq}^\ell(k) : p, q = 0, 1, 2, \dots, N\}$ is the moment corresponding to the r -th element of the vector $\hat{\rho} = [\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_T, \hat{\rho}_{T+1}, \dots, \hat{\rho}_{(N+1)^2}]'$ that comprises the estimated ICCs arranged in descending order of their magnitude. Figure 4 shows a 2D scatter plot depicting the clustering performance or sparse representation of the first two DGHMs selected from \mathbf{F}_k^ℓ for five randomly chosen classes, each having 10 samples obtained from the popular AT&T face database [58]. From the scatter plot, it may be seen that the discrimination capabilities of the selected moments in order to classify the subjects are considerable. It is to be noted that similar clustering performance are observed for any two of the moments that correspond to an ICC close to unity. In order to reduce the dimension of the feature vector, we prefer to use the number of moments for classification as a fraction of stored moments defined as $T = \beta(N + 1)^2$ where β ($0 < \beta \leq 1$) is a classification parameter. When there exists significant variations among the images in each class of a face database, a relatively small number of moments with high values of ICCs are preferred to construct a discriminative feature set for a subject, in which case a lower value of β may be chosen. On the other hand, a higher value of β is preferred for a decreasing level of within-class variability of the face images while constructing the feature set, since a larger number of moments possess discriminative nature in such a case. It is to be noted that within-class variabilities of the moments occur due to the variations of faces in terms of appearance, expressions, pose, age, resolution, illumination,

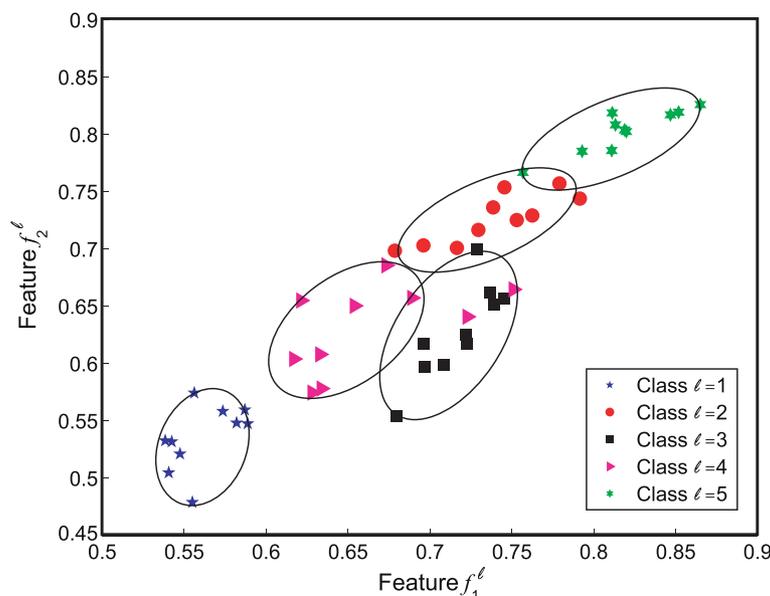


Fig. 4 Graphical representation depicting the sparse representation of two DGHMs used to construct the face features

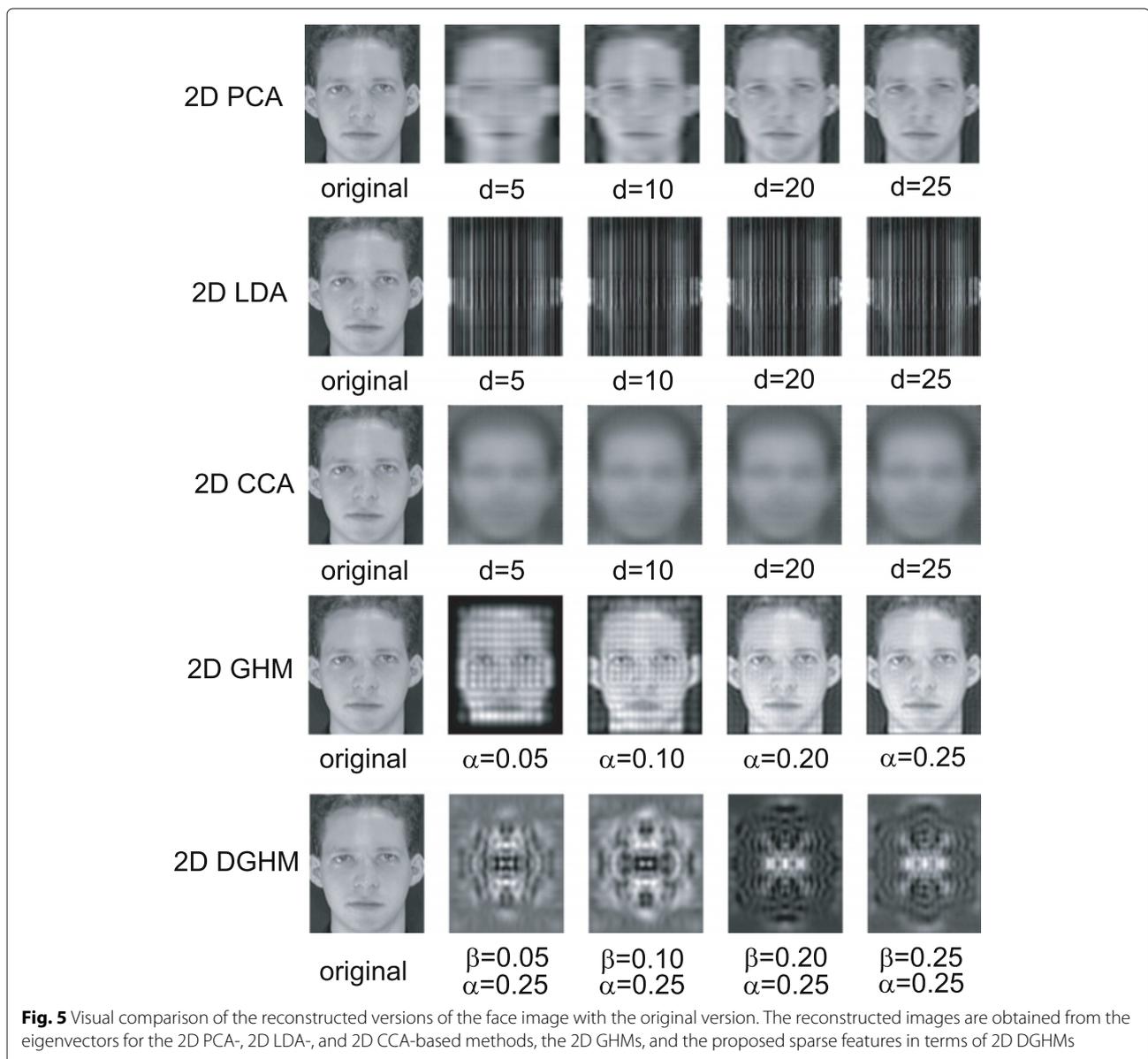
and occlusion on which the classification parameter β depends.

Figure 5 shows a comparison of the reconstructed versions of a typical face image projected from the d ($d \leq \min(U, V)$) number of eigenvectors of the 2D-PCA [25], 2D-LDA [27], and 2D-CCA [29] methods, αUV ($0 < \alpha < 1$) number of 2D GHMs, and $\alpha\beta UV$ ($\alpha = 0.25, 0 < \beta < 1$) number of 2D DGHMs. It is seen from this figure that a face image of very good quality can be reconstructed, if the number of GHMs that are stored equals only 25 % of the face data. However, with the exception of 2D-PCA, images reconstructed from increasing numbers of discriminative sparse features do not resemble the original face image, as can be seen in the case of the 2D-LDA, 2D-CCA, and 2D-DGHM-based methods. These results are consistent

with the fact that the discriminative sparse features do not possess good reconstruction ability of images in many cases [59].

3.3 The Naive-Bayes classifier to identify subjects

In the classification technique, a given face is assigned to one of the K subjects or classes, namely, $\omega_1, \omega_2, \dots, \omega_K$ on the basis of a feature vector $\mathbf{F} = [f_1, f_2, f_3, \dots, f_T]^T$ associated with the face image. Statistical pattern recognition considers that the feature vector \mathbf{F} is a T -dimensional observation drawn randomly from the class conditional PDF $p(\mathbf{F}|\omega_\ell)$, where ω_ℓ is the class to which the feature vector belongs [60]. There are several statistical decision rules for assigning a given pattern, face in this case, to a class. Among these rules, the Bayes decision is optimal in



the sense that it minimizes the Bayes risk, which is the expected value of the loss function [51]. The Bayes decision rule assigns the face in question described by \mathbf{F} to the subject ω_ℓ for which the conditional risk

$$R(\omega_\ell|\mathbf{F}) = \sum_{m=1}^K L(\omega_\ell, \omega_m) p(\omega_m|\mathbf{F}) \quad (22)$$

is minimum, where $L(\omega_\ell, \omega_m)$ is the loss incurred in deciding the subject ω_ℓ when the true subject is ω_m , and $p(\omega_m|\mathbf{F})$ is the posterior PDF [61]. In the case of the 0/1 loss function, the Bayes decision rule simplifies to the MAP decision rule [51], which assigns the input face represented by \mathbf{F} to the subject ω_ℓ if

$$p(\omega_\ell|\mathbf{F}) > p(\omega_m|\mathbf{F}) \quad \forall m \neq \ell \quad (23)$$

Using the Bayes theorem, the posterior PDF may be written as

$$p(\omega_m|\mathbf{F}) = \frac{p(\mathbf{F}|\omega_m)p(\omega_m)}{p(\mathbf{F})} \quad (24)$$

Since the moments that constitute \mathbf{F} are obtained from the orthogonal polynomial functions, the features may be treated as independent given the class. Thus, the PDF of features given the class may be written as

$$p(\mathbf{F}|\omega_m) = \prod_{r=1}^T p(f_r|\omega_m) \quad (25)$$

In designing the classifier, $p(\mathbf{F})$ may be ignored since it does not depend on m . Hence, an unknown face image having feature vector \mathbf{F} should be assigned to the class m that maximizes the following decision function [62]:

$$d_m(\mathbf{F}) = \prod_{r=1}^T p(f_r|\omega_m)p(\omega_m) \quad (26)$$

In Section 3.1, it was verified that the class conditional density of a GHM of any given order may be considered as approximately normal. Thus, the feature components given the class are assumed to follow the normal distribution, i.e., $f_r|\omega_m \sim \mathcal{N}(\mu_{rm}, \sigma_{rm}^2)$. Such a choice has the added advantage of yielding a naive Bayes classifier that is mathematically tractable. In this study, the number of subjects per class in the training set λ_{tr} is considered to be the same, and hence, the class prior may be obtained as $p(\omega_m) = K^{-1}$. A common problem in face recognition is that the number of training images available for estimating the parameters of $p(f_r|\omega_m)$ in each class is usually small and at the same time the number of classes are relatively large. Thus, a strategy often employed to improve the Bayes classifier performance is to assume that $\sigma_{rm}^2 = \sigma_r^2$ for all m and then replace the estimates of σ_{rm}^2 obtained from the training images in each class by the pooled estimate [63, 64]. In such a case, the test face image

having feature vector \mathbf{F} will be assigned to the class m that minimizes the decision function

$$\mathcal{D}_m(\mathbf{F}) = \sum_{r=1}^T (f_r^m - \mu_{rm})^2 \quad (27)$$

The class conditional mean of the feature components are estimated from the training face images as

$$\hat{\mu}_{rm} = \frac{1}{\lambda_{tr}} \sum_{k=1}^{\lambda_{tr}} f_{rk}^m \quad (28)$$

The computational complexity of finding the feature vector \mathbf{F} using the sorting operation of ICCs is $\mathcal{O}((N+1)^4)$ and the complexity of the proposed naive Bayes classifier is $\mathcal{O}(N+1)$. Since $N \gg 1$ in practice, the computational complexity of the proposed 2D GHM-based face recognition method is $\mathcal{O}(N^4)$. On the other hand, the computational complexity of the PCA-, LDA-, or CCA-based methods may be shown to be $\mathcal{O}(N^6/\alpha^3)$ ($0 < \alpha < 1$). Hence, the proposed DGHM-based face recognition method is computationally efficient as compared to the traditional PCA-, LDA-, and CCA-based methods.

4 Simulation results

Extensive experimentations have been carried out in order to evaluate the performance of the proposed 2D DGHM-based face recognition method as compared to the existing methods. This section describes the databases, the experimental conditions, and results of the comparisons.

4.1 Face databases

The proposed method was evaluated on a number of face databases; however, the results presented in this paper are those obtained using the popular AT&T face database [58], a generic face database obtained from the comprehensive Face Recognition Grand Challenge (FRGC) v2.0 database [65], and the standard protocols of the Face Recognition Technology (FERET) database [66], the Labeled Faces in the Wild (LFW) database [67], and the YouTube Faces (YTF) database [68]. The details of the databases considered in the experiments are discussed under separate subheadings as a convenience.

4.1.1 AT&T database

The AT&T face database contains a total of 400 images from 40 individuals, each individual having 10 different images. Images are captured at different times for some of the subjects. The facial images have variability in terms of expression such as smiling or nonsmiling, open or closed eye, and in terms of appearance such as with or without glasses. The tolerance of geometric distortions of the AT&T database include rotation up to 20° and scaling up to 10%. All the images of this face database have a size of 112×92 pixels with a bit depth of 8. In order to obtain

only the face part of the images, suitable free-hand elliptic masks are used. Further, the masked images are normalized to a mean level of 128 and standard deviation of 85, so that variations due to illuminations are reduced.

4.1.2 FRGC v2.0 database

The FRGC v2.0 database contains 4007 number of 8-bit face images having pixel resolution 480×640 captured from shoulder level up for 466 subjects in both controlled and uncontrolled environments [69]. The demographics of this database include partition by race (White-68 %, Asian-22 %, others-10 %), age (18 to 22 years-65 %, 23 to 27 years-18 %, 28+ years-17 %), and sex (male-57 %, female-43 %). The number of replica of face images captured from individual subjects varies from 1 to 22. The images captured have in-plane and out-of-plane rotations of the head up to about 15° . In this database, the faces display various facial expressions including neutral, happy, surprise, sad, and disgust. There are major illumination variations in the images of each subject. A few of the subjects have facial hair, but none of them wears glass. A generic face database is obtained from the FRGC v2.0 in such a way that each subject has at least 10 sample face images. In such a case, the generic face database used in the experiments includes 2774 face images obtained from 186 subjects. The nose coordinates of these images are identified first, and then, 161×161 pixel-size images are cropped using the nose coordinate as center. The processes of extracting facial parts and reducing illumination variation among the faces of the FRGC database are similar to that of the AT&T database.

Figure 6 shows a few examples of the pairs of face images of the AT&T and generic FRGC databases considered in the experiments. The variations of expression, appearance, pose, scaling, rotation, and illumination of face images in these databases are evident from this figure. For example, the second and fourth subjects from the left in the first database have face images with and without glasses, and the second subject in the second database has face images with and without mustache. The illumination variations are observed between the left and right cheeks in the face images of the second database, and even a horizontal texture pattern is seen in one of the images of the third subject.

4.1.3 FERET database

In the color FERET database, the standard testing subsets include the fa, fb, dup-I, and dup-II that consist of 994, 992, 736, and 228 frontal facial images, respectively. These frontal face images each of size 512×768 are considered in the experiments. The images of fa and fb are used to investigate the effect on recognition due to changes in expressions. The sets of images of dup-I and dup-II are used for considering the variations arising from aging as

well as appearances due to hairstyles or glasses. In our experiments, the color images are converted to grayscale representation and the face images of size 200×180 are cropped using the nose coordinates provided as the centers. The GHMs for face recognition are obtained from these cropped images without any further processing such as the scaling, rotation, intensity normalization, or any masking.

4.1.4 LFW database

The LFW-a database consists of 13,233 images of 5749 subjects collected from the web. The facial parts of the images are detected using the Viola-Jones detector [70], and the images are aligned by the organizer using a commercial software. The face images have significant variations in terms of pose, occlusion, expression, and even resolution. Organizers of this database recommend to report the performance of unconstrained face verification on this database as 10-fold cross validation using the splits of facial image pairs that are randomly generated and provided. In the experiments, we use the cropped images of size 150×130 with the center coordinate of the original images as the face center and without any further processing.

4.1.5 YTF database

The data set of YTF contains 3425 videos of 1595 subjects collected from the website of YouTube. There are huge variations of pose, expressions, appearances, occlusions, colors, illumination, and resolutions of the frames of these videos. The average length of each video clip is 181.3 frames. The standard protocol for evaluation of the performance of the unconstrained face verification includes 5000 video pairs of this database. These pairs are equally divided into 10-folds, and each fold has almost equal numbers of intra-personal pairs and inter-personal pairs. The facial parts of the frames are detected using the Viola-Jones detector, and the aligned frames are provided by the authors of [68]. In the experiments, we use the grayscale version of the cropped frames of size 150×120 by considering the center coordinate of the aligned frames as the face center.

4.2 Experimental setup

The proposed 2D DGHM-based method is first compared with the classical 2D-PCA [25], 2D-LDA [27], 2D-CCA [29], and 2D KCM [43] methods as these methods represent competitive algorithms in the area of constrained face recognition using the PCA, LDA, CCA, and orthogonal moments, respectively. In the case of proposed method, it is assumed that face images are stored in terms of GHMs in a compressed form using a value of α less than unity. Each face image of the AT&T and generic FRGC databases is stored in terms of 2601 and 6480 GHMs, respectively,

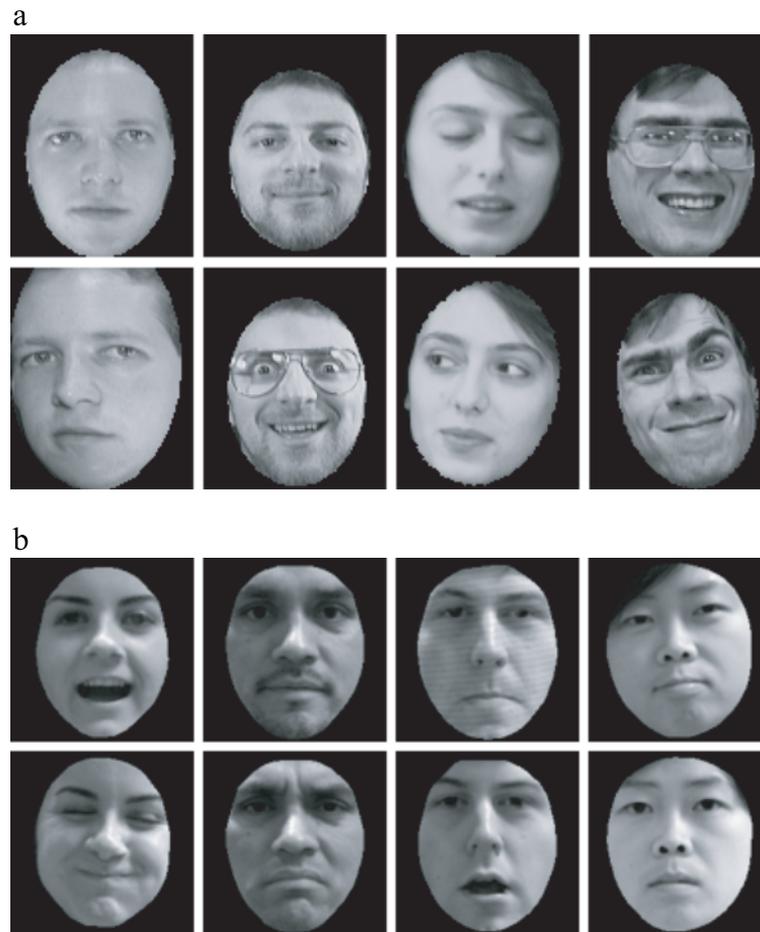


Fig. 6 Variations of the expressions, appearances, poses, and illuminations in the face images of the two databases, namely, (a) AT&T and (b) generic FRGC v2.0

by choosing $\alpha = 0.25$. In the case of 2D-PCA and 2D-LDA methods, the number of eigenvectors for generating feature matrices is chosen as $d = \lambda_{tr} - 1$, since such a choice gives an optimum performance for the recognition accuracy [25]. In the case of 2D-CCA method, the value of d is chosen to be 10 and 25 for the AT&T and FRGC databases, respectively, due to the fact that such choices provide the highest level of identification accuracy. The 2D KCM-based method stores each face image in terms of five sets of KCMs that are obtained from the full, left, right, upper, and lower parts of the face. Default values of the parameters of the binomial distribution are chosen to generate the 2D KCMs of each of the five parts, while maximum order of the moment is chosen as 19 due to the fact that such an order shows the highest performance [43]. In order for the results to be statistically robust in terms of accuracy, 25 subjects are chosen randomly from the generic FRGC database and then λ_{tr} number of training images per subject are selected randomly while the rest of the images of the selected subjects are treated as probe

images. Since AT&T has only a few number of subjects, we consider the entire database while generating a random set for testing the face recognition accuracy. Performance of a testing set is measured in terms of the recognition rate defined as

$$RR = \frac{\text{Number of probe images correctly classified}}{\text{Total number of probe images}} \times 100 \quad (29)$$

The results presented in this paper are obtained by averaging the accuracies obtained from such type of 1000 random testing sets. In order to investigate the performance of classification for the five methods considered in the experiments, the number of training images per subject λ_{tr} is also varied. The RR of the proposed 2D DGHM-based method depends on the classification parameter β which is chosen from the ICCs. Here, β defines the length of the moment-based feature vector used in the naive Bayes classification, and hence, the

accuracy of the proposed DGHM-based face recognition method is estimated for increasing values of this parameter.

We also compare the proposed DGHM-based features with some of the state-of-the-art descriptors that are used in the unconstrained face recognition problems. In this scenario, the standard protocols of the color FERET, the LFW-a, and the YTF databases are used for the evaluation of the performance of face recognition or verification. The fa dataset in FERET is used as the gallery images, while the fb, dup-I, and dup-II are used as the probe sets. The rank 1 recognition accuracies of the known classes are estimated from the minimum of proposed decision function and by using the features of the gallery and probe sets. In the experiments, "View-2" set of the LFW that consists of 10 folds of randomly chosen 300 positive and 300 negative image pairs is used for evaluation of face verification. The images of the FERET and LFW databases are represented in terms of the GHMs by choosing the values of α to be 0.25 and 0.15, respectively. In order to train the GHM-based features of the FERET datasets, the entire set of frontal images are considered. On the other hand, the features of the LFW-a are trained using the subjects that have sample sizes in between 10 and 50 by avoiding the subjects having too small or large sample sizes. The GHM features of the videos of YTF are obtained by taking the mean of each moment obtained from the frames of the clips and by choosing α to be 0.15. In order to train the GHM features of the YTF database for the experiment of verifications, the intra-personal and inter-personal pairs are chosen as two different classes while estimating the ICCs.

4.3 Recognition and verification results

In this subsection, first, we present the results of variation of recognition accuracy with respect to the classification parameter β . Next, the recognition or verification results of the five databases are presented separately.

4.3.1 Effect of β on recognition rates

Figure 7 shows the variations in face recognition accuracy of the proposed 2D DGHM-based method with respect to changes in the values of β and λ_{tr} when the databases AT&T and generic FRGC are used. Similar variations of recognition rates are also obtained from the FERET, LFW, and YTF databases but are not included to avoid presentation of repetitive results. From Fig. 7, it can be seen that in general, the recognition accuracy increases with the number of training images per subject for the databases considered. It is also seen from this figure that for a given λ_{tr} , the recognition accuracy increases initially with the value of β that defines the length of the feature vector. However, for β greater than a certain value, the recognition accuracy remains almost unchanged. In

the experiments, it is found that the recognition accuracy does not change significantly for the AT&T, generic FRGC, FERET, LFW, and YTF databases with variations of β within the range (0.20–0.40), (0.30–0.50), (0.05–0.10), (0.03–0.06), and (0.02–0.05), respectively. These results are consistent with the explanations given in Section 3.2 and by noting that in general, the variations of the faces for a given class are increasing in the order of five experimental databases, viz., generic FRGC, AT&T, FERET, LFW, and YTF.

4.3.2 Recognition rate on AT&T

Table 1 shows the comparisons of the face recognition accuracies of the proposed 2D DGHM-based method with that of the 2D-PCA [25], 2D-LDA [27], 2D-CCA [29], and 2D-KCM [43] methods for training samples of sizes 3 to 7. The average number of probe images for the databases decrease with the number of training images per subject. The recognition accuracies of the proposed method shown in Table 1 are obtained by using $\beta = 0.30$. From the accuracies in percentage given in this table, it may be found that the proposed 2D DGHM-based face recognition method outperforms the PCA, LDA, and CCA-based methods or even KCM-based method for any set of training images per subject.

4.3.3 Recognition rate on FRGC

Table 2 shows the comparisons of the face recognition accuracies of the proposed 2D DGHM-based method with that of the 2D-PCA, 2D-LDA, 2D-CCA, and 2D-KCM methods, when the training sample size varies in between 3 and 7. Average number of probe images for the databases are given in the table with a consideration that each of the randomly chosen 1000 test sets may contain variable number of face images per subject. The classification parameter β was chosen as 0.40 to evaluate the recognition accuracies of the proposed method. The superiority of the proposed DGHM-based features in terms of the recognition rates given in Table 2 are very consistent with that of the AT&T face database.

4.3.4 Recognition rate on FERET

Table 3 shows the rank 1 recognition accuracies that compare the performance of the proposed DGHM-based method with that of the contemporary methods including the LBP [71, 72], local visual primitive (LVP), local Gabor textons (LGT) [73], and hierarchical ensemble global classifier (HEGC) [74] employed on the FERET database. From this table, it can be seen that the proposed DGHM-based features show the best performance in the case of fb and dup-II datasets and a competitive performance on the dup-I dataset when compared to the Gabor-based features.

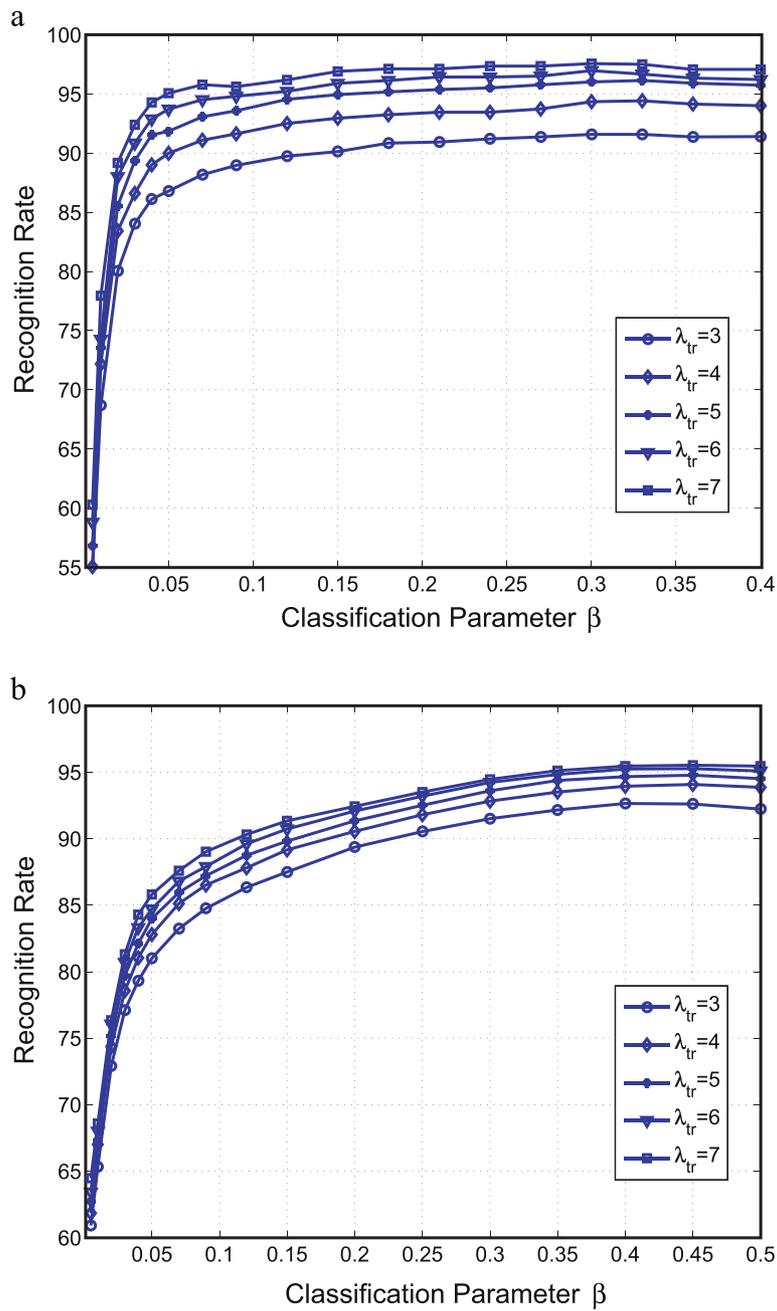


Fig. 7 Variations on identification accuracies of the proposed 2D DGHM-based face recognition method with respect to the training sample size λ_{tr} and classification parameter β . The databases are **(a)** AT&T and **(b)** generic FRGC

Table 1 Comparisons of recognition rates on AT&T face database

λ_{tr}	Average number of probe images	Recognition rate in %				
		2D-PCA [25]	2D-LDA [27]	2D-CCA [29]	2D-KCM [43]	2D-DGHM
3	280	89.95	91.17	90.37	86.19	91.57
4	240	92.75	93.81	93.75	90.36	94.42
5	200	94.67	95.47	95.67	93.02	96.13
6	160	95.58	95.54	96.59	94.93	96.67
7	120	96.78	96.83	97.24	96.19	97.50

Table 2 Comparisons of recognition rates on generic FRGC face database

λ_{tr}	Average number of probe images	Recognition rate in %				
		2D-PCA [25]	2D-LDA [27]	2D-CCA [29]	2D-KCM [43]	2D-DGHM
3	298	73.06	81.42	89.47	88.49	92.65
4	273	82.96	88.73	90.63	91.39	93.94
5	248	87.43	92.33	92.43	93.01	94.66
6	223	91.15	93.77	93.02	94.37	95.24
7	198	93.05	95.00	94.63	94.99	95.46

4.3.5 Verification rate on LFW

To compare the DGHM features with other descriptors previously reported on the LFW database, the verification results acquired from the restricted image set are provided. Figure 8 shows the receiver operating characteristic (ROC) curves that compare verification performance of the features obtained from the Eigenfaces [13], scale-invariant feature transform (SIFT) [75], LBP [76], multiple kernel learning (MKL) [77], and the DGHMs. It is seen from this figure that the proposed DGHM features provide a true positive rate better than the SIFT or PCA-based features and a competitive true rate when comparing with the LBP or MKL-based features. Figure 9 shows the face verification results obtained from the proposed DGHM method using 16 pairs of face images considered in the View 2 experiments of the LFW database. In this figure, the results are presented using four types of verifications, namely, the correct or incorrect verification when the subjects are identical and the same when the identities are different. The results of the positive pairs show that the proposed method is capable of verifying the subject when there remains scaling, rotation of faces both in- and out-of-planes, or even illumination variation. However, if pose variation, occlusion, or makeup is significant, then the positive pairs may not be verified accurately. It is also found from this figure that if the appearances of faces are very similar, then the negative pairs can be wrongly identified.

Table 3 Comparisons of recognition rates on standard partitions of color FERET database

Methods	fb	dup-I	dup-II
FERET Baseline [66] ^a	96.00	59.00	52.00
LBP [71, 72] ^a	97.00	66.00	64.00
LVP [80] ^a	97.00	66.00	50.00
LGT [73] ^a	97.00	71.00	67.00
HEGC [74] ^a	96.00	53.00	23.00
DGHM	97.15	70.61	67.04

^aNote that the results are cited from original papers

4.3.6 Verification rate on YTF

The proposed DGHM features are compared with the existing features using the restricted video pairs; the results of which are reported in the website maintained by the organizers of this database. Figure 10 shows the ROC curves obtained from the minimum distance-based similarity and the matched background similarity (MBGS) of LBP features [68], the adaptive probabilistic elastic matching (APEM) [78], and the large margin multimetric learning (LM³L) [79] methods both of which use a fusion of SIFT and LBP features. It can be seen from this figure that the proposed DGHM feature performs better than the LBP feature independent of the similarity measures chosen. At the same time, the proposed feature shows a competitive rate of verifying true-positives for a given false-positive as compared to the fused features obtained from the SIFT and LBP such as those in the APEM and LM³L methods. Figure 11 shows selected frames of videos of typical 16 pairs of identities that are used in the verification experiments of the YTF database. Among these pairs, half represent the intra-personal pairs and the rest the inter-personal pairs. The correct and incorrect verifications of these identities are shown in separate groups. The sample images of correct verification reveal that the proposed DGHM features are capable of identifying subjects; the videos of which have significant variations in terms of resolutions, pose, and color. However, the proposed method fails when the frames of the video suffer from severe motion blurs or occlusions due to embedded texts and shadows. A few examples of close appearance of inter-personal facial frames that are verified correctly by the proposed DGHM features are shown in Fig. 11. In the experiments, it is observed that the correctness of inter-personal verification increases with the level of dissimilarity of the frames of the video.

5 Conclusions

Representation of images and formation of feature sets from such a representation play key roles in the success of a face recognition algorithm. Compact representation of images is desirable to reduce the storage requirement, a critical issue for face databases. The feature sets should

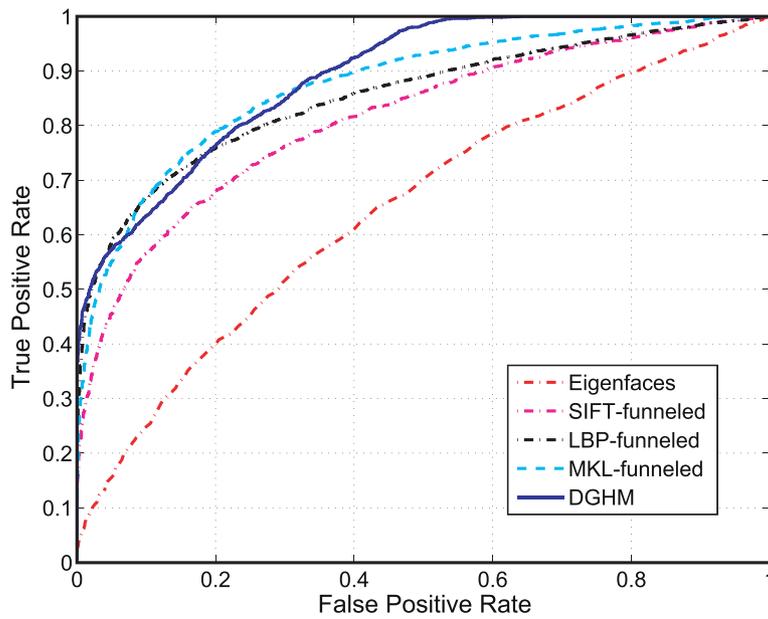


Fig. 8 Comparisons of the ROC curves obtained from five face recognition algorithms considered in the View 2 experiments of the LFW database. The results of Eigenfaces, SIFT, LBP, and MKL are cited from the website of database

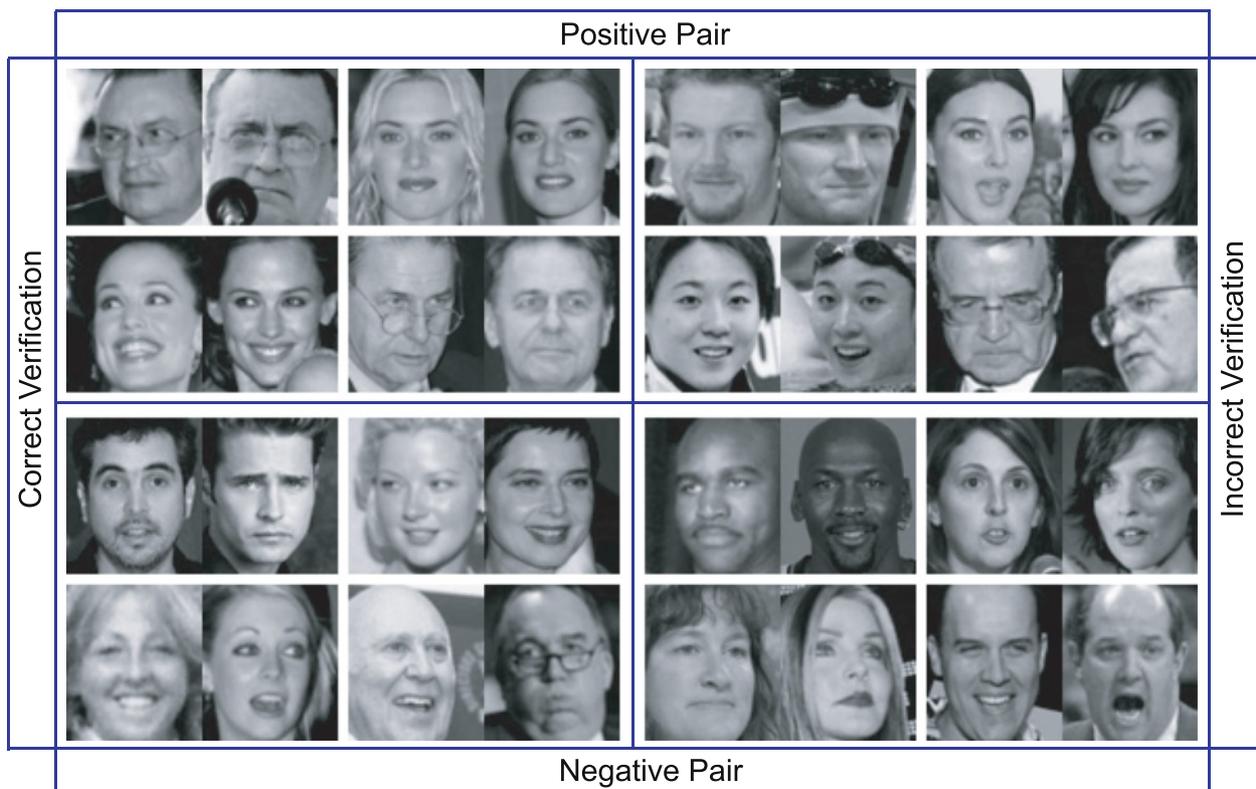


Fig. 9 Examples of correct and incorrect verifications of the positive and negative pairs considering the View 2 experiments of the LFW database

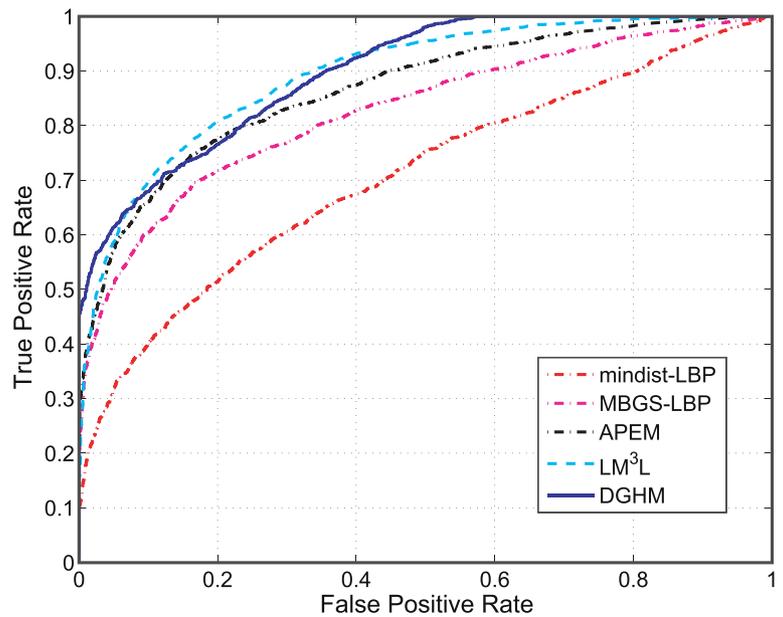


Fig. 10 Comparisons of the ROC curves obtained from five face recognition algorithms considered in the restricted experiments of the YTF database. The results of mindist-LBP, MBGS-LBP, APEM, and LM³L are cited from the website of database

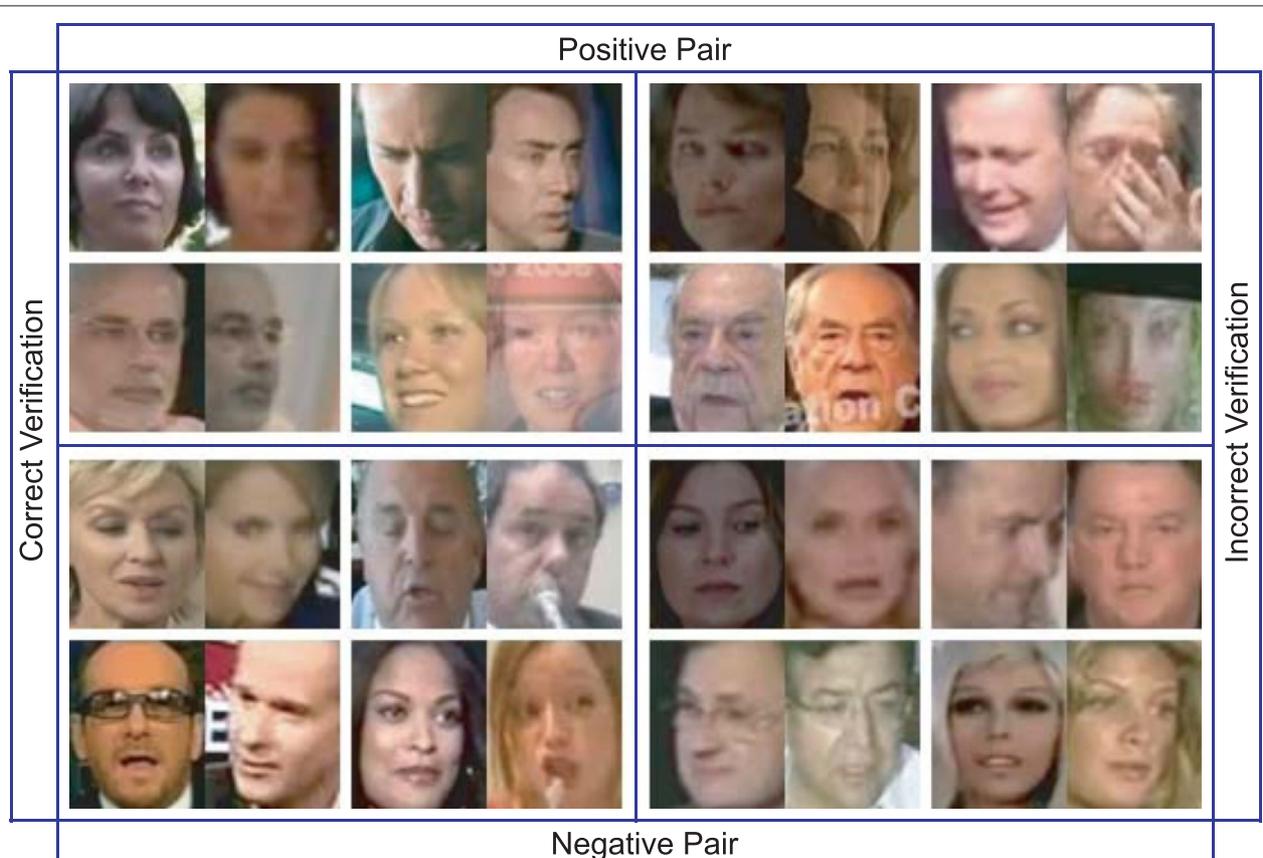


Fig. 11 Examples of correct and incorrect verifications of the positive and negative pairs considering the experiments of the YTF database

be capable of capturing the higher-order nonlinear structures hidden in a face image, and at the same time, the features must be sparse enough to discriminate the identities. In this paper, the 2D GHMs have been used to represent face images so that a face database may be compactly represented. Use of these orthogonal moments for face recognition has also been motivated from the fact that certain linear combinations of these moments form geometric moments that are invariants to shift, scale, and rotation of a pattern. The key contribution of the paper is that features of facial images have been obtained by selecting only those moments having greater discriminatory power, as measured by their large values for the ICC, instead of choosing a heuristic set of GHMs. This is an effective means of filtering out those moments that contribute little to distinguishing among different classes of the face images. Classification of test images has been performed using the naive Bayes classifier, which is simple, but fast and known to perform remarkably well in many practical applications with huge size database. In order to compare the recognition performance of the proposed method, extensive experiments have been carried out on a number of commonly used image and video-based face databases, such as the AT&T, FRGC, color FERET, LFW, and YTF that have facial images or frames with variations in terms of appearance, occlusion, expression, pose, color, resolution, and illumination both in the constrained and unconstrained environments. The results have shown that the proposed 2D-DGHM-based method provides higher accuracy in face recognition than those provided by the popular 2D-PCA, 2D-LDA, and 2D-CCA methods, as well as the 2D KCM method even when the training data set has small number of samples per subject in the constrained environments. The face recognition and verification results on the standard protocols of the databases of unconstrained environments also reveal that the proposed DGHM features can perform better than the commonly used descriptors such as the LBP or SIFT.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to give thanks to the anonymous reviewers for their valuable comments that were useful to improve the quality of the paper.

Author details

¹Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, 1205 Dhaka, Bangladesh. ²Institute of Statistical Research and Training, Dhaka University, 1000 Dhaka, Bangladesh.

Received: 18 August 2014 Accepted: 29 October 2015

Published online: 12 November 2015

References

1. R Heitmeyer, Biometric identification promises fast and secure processing of airline passengers. *Int. Civ. Aviat. Org. J.* **55**(9), 10–11 (2000)

2. S Chen, S Mau, MT Harandi, C Sanderson, A Bigdeli, BC Lovell, Face recognition from still images to video sequences: a local-feature-based framework. *EURASIP J. Image and Video Processing.* **2011**(790598), 1–14 (2011)
3. AF Abate, M Nappi, D Riccio, G Sabatino, 2D and 3D face recognition: a survey. *Pattern Recogn. Lett.* **28**(14), 1885–1906 (2007)
4. D Chen, X Cao, F Wen, J Sun, in *Proc. Computer Vision and Pattern Recognition*. Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification (IEEE Computer Society Portland, OR, 2013), pp. 3025–3032. doi:10.1109/CVPR.2013.389
5. H Mohammadzade, D Hatzinakos, in *Proceedings of the Fourth IEEE International Conference on Biometrics: Theory Applications and Systems*. An expression transformation for improving the recognition of expression-variant faces from one sample image per person (IEEE Systems, Man, & Cybernetics Society Washington DC, USA, 2010), pp. 1–6
6. J Daugman, The importance of being random: statistical principles of iris recognition. *Pattern Recognit.* **36**, 279–291 (2003)
7. AP James, One-sample face recognition with local similarity decisions. *Int. J. Appl. Pattern Recogn.* **1**(1), 61–80 (2013)
8. Y Lu, J Zhou, S Yu, A survey of face detection, extraction and recognition. *Comput. Inform.* **22**(2), 163–195 (2003)
9. KV Mardia, A Coombes, J Kirkbride, A Linney, JL Bowie, On statistical problems with face identification from photographs. *J. Appl. Stat.* **23**(6), 655–675 (1996)
10. Y Jiang, P Guo, in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. Comparative studies of feature extraction methods with application to face recognition (IEEE Systems, Man, & Cybernetics Society Montreal, QC, Canada, 2007), pp. 3627–3632
11. S Rani, JD Devaraj, R Sukanesh, in *Proceedings of International Conference on Computational Intelligence and Multimedia Applications*. A novel feature extraction technique for face recognition system, vol. 2 (IEEE Computer Society, Tamil Nadu, India, 2007), pp. 431–435
12. GH Givens, JR Beveridge, YM Lui, DS Bolme, BA Draper, PJ Phillips, Biometric face recognition: from classical statistics to future challenges. *Wiley Interdiscip. Rev. Comput. Stat.* **5**(4), 288–308 (2013)
13. M Turk, A Pentland, Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
14. PN Belhumeur, JP Hespanha, DJ Kreigman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern. Anal. Mach. Intell.* **19**(7), 711–720 (1997)
15. B Moghaddam, A Pentland, Probabilistic visual learning for object recognition. *IEEE Trans. Pattern. Anal. Mach. Intell.* **19**(7), 696–710 (1997)
16. K Tan, S Chen, Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing.* **64**, 505–511 (2005)
17. KI Kim, K Jung, HJ Kim, Face recognition using kernel principal component analysis. *IEEE Signal Proc. Lett.* **9**(2), 40–42 (2002)
18. W Liu, Y Wang, SZ Li, T Tan, in *Lecture Notes in Computer Science in Biometric Authentication*. Null space approach of Fisher discriminant analysis for face recognition, vol. 3087 (Springer-Verlag Prague, Czech Republic, 2004), pp. 32–44. doi:10.1007/978-3-540-25976-3_4
19. X Wang, X Tang, in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Dual-space linear discriminant analysis for face recognition, vol. 2 (IEEE Computer Society Washington, DC, USA, 2004), pp. 564–569
20. JH Friedman, Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84**(405), 165–175 (1989)
21. JW Lu, KN Plataniotis, AN Venetsanopoulos, in *Proceedings of IEEE International Conference on Image Processing*. Boosting linear discriminant analysis for face recognition, vol. 1 (IEEE Signal Processing Society Barcelona, Spain, 2003), pp. 657–660
22. XY Jing, YY Tang, D Zhang, A Fourier-LDA approach for image recognition. *Pattern Recogn.* **38**, 453–457 (2005)
23. YW Pang, L Zhang, MJ Li, ZK Liu, WY Ma, in *Lecture Notes in Computer Science in Advances In Multimedia Information Processing*. A novel Gabor-LDA based face recognition method, vol. 3331 (Springer-Verlag Tokyo, Japan, 2004), pp. 352–358. doi:10.1007/978-3-540-30541-5_44
24. H Zhao, PC Yuen, Incremental linear discriminant analysis for face recognition. *IEEE Trans. Syst. Man Cybern. B.* **38**(1), 210–221 (2008)
25. J Yang, D Zhang, AF Frangi, J-Y Yang, Two-dimensional PCA: a new approach of appearance-based face representation and recognition. *IEEE Trans. Pattern. Anal. Mach. Intell.* **26**(1), 131–137 (2004)

26. J Meng, Y Yang, Symmetrical two-dimensional PCA with image measures in face recognition. *Int. J. Adv. Robot. Syst.* **9**, 1–10 (2012)
27. M Li, B Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix. *Pattern Recogn. Lett.* **26**, 527–532 (2005)
28. SH Lee, S Choi, Two-dimensional canonical correlation analysis. *IEEE Signal Proc. Lett.* **14**(10), 1–4 (2007)
29. G Kukharev, E Kamenskaya, Application of two-dimensional canonical correlation analysis for face image processing and recognition. *Pattern Recog. Image Anal.* **20**(2), 210–219 (2010)
30. BA Draper, K Baek, MS Bartlett, JR Beveridge, Recognizing faces with PCA and ICA. *Comp. Vision Image Underst.* **91**, 115–137 (2003)
31. J Yang, X Gao, D Zhang, J-Y Yang, Kernel ICA: an alternative formulation and its application to face recognition. *Pattern Recognit.* **38**, 1784–1787 (2005)
32. SJD Prince, JH Elder, J Warrell, FM Felisberti, Tied factor analysis for face recognition across large pose differences. *IEEE Trans. Pattern. Anal. Mach. Intell.* **30**(6), 1–15 (2008)
33. S Xie, S Shan, X Chen, J Chen, Fusing local patterns of Gabor magnitude and phase for face recognition. *IEEE Trans. Image Process.* **19**(5), 1349–1361 (2010)
34. G Anbarjafari, Face recognition using color local binary pattern from mutually independent color channels. *EURASIP J. Image and Video Processing.* **2013**(6), 1–11 (2013)
35. T-K Kim, J Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Trans. Pattern. Anal. Mach. Intell.* **27**(3), 318–327 (2005)
36. R Jafri, HR Arabia, A survey of face recognition techniques. *J. Inf. Process. Syst.* **5**(2), 41–68 (2009)
37. MK Hu, Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory.* **8**(2), 179–187 (1962)
38. MR Teague, Image analysis via a general theory of moments. *J. Opt. Soc. Am.* **70**(8), 920–930 (1980)
39. J Haddadnia, K Faez, M Ahmadi, An efficient human face recognition system using pseudo-Zernike moment invariant and radial basis function neural networks. *Int. J. Pattern Recognit. Artif. Intell.* **17**(1), 41–62 (2003)
40. Y-H Pang, ABJ Teoh, DCL Ngo, A discriminant pseudo-Zernike moments in face recognition. *J. Res. Pract. Inf. Technol.* **38**(2), 197–211 (2006)
41. NH Foon, Y-H Pang, ATB Jin, DNC Ling, in *Proceedings of the International Conference on Computer Graphics, Imaging and Visualization*. An efficient method for human face recognition using wavelet transform and Zernike moments (IEEE Computer Society Penang, Malaysia, 2004), pp. 65–69
42. W Arnold, VK Madasu, WW Boles, PK Yarlagadda, in *Proceedings of the RNISA Security Technology Conference*. A feature based face recognition technique using Zernike moments (Queensland University of Technology Melbourne, Australia, 2007), pp. 341–345
43. JS Rani, D Devaraj, Face recognition using Krawtchouk moment. *Shadhana.* **37**(4), 441–460 (2012)
44. SMM Rahman, MM Reza, QM Hasani, in *Proc. 2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*. Low-complexity iris recognition method using 2D Gauss-Hermite moments (IEEE Signal Processing Society Trieste, Italy, 2013), pp. 142–146
45. C Xu, Y Wang, T Tan, L Quan, in *Lecture Notes in Computer Science: Advances in Biometric Person Authentication*. 3D face recognition based on G-H shape variation, vol. 3338 (Springer-Verlag, 2004), pp. 233–244
46. N Belghini, A Zarghili, J Kharroubi, 3D face recognition using Gaussian Hermite moments. *Spec. Issue Int. J. Comput. Appl. Softw. Eng. Databases Expert Syst.* **1**, 1–4 (2012)
47. Y Ming, Q Ruan, R Ni, in *Proceedings of IEEE International Conference on Image Processing*. Learning effective features for 3D face recognition, vol. 1 (IEEE Signal Processing Society Hong Kong, 2010), pp. 2421–2424
48. B Yang, M Dai, Image analysis by Gaussian-Hermite moments. *Signal Process.* **91**, 2290–2303 (2011)
49. B Yang, G Li, H Zhang, M Dai, Rotation and translation invariants of Gaussian-Hermite moments. *Pattern Recogn. Lett.* **32**, 1283–1298 (2011)
50. PE Shrouf, JL Fleiss, Intra-class correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**(2), 420–428 (1979)
51. AK Jain, RPW Duin, J Mao, Statistical pattern recognition: a review. *IEEE Trans. Pattern. Anal. Mach. Intell.* **22**(1), 4–37 (2000)
52. D Soria, JM Garibaldi, F Ambrogi, EM Biganzoli, IO Ellis, A 'non-parametric' version of the naive Bayes classifier. *Knowl.-Based Syst.* **24**(6), 775–784 (2011)
53. J Shen, W Shen, D Shen, On geometric and orthogonal moments. *Int. J. Pattern Recognit. Artif. Intell.* **14**(7), 875–894 (2000)
54. M Abramowitz, I Stegun, *Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables*, 10th edn. (Dover, New York, 1965)
55. B Yang, M Dai, Image reconstruction from continuous Gaussian-Hermite moments implemented by discrete algorithm. *Pattern Recogn.* **45**, 1602–1616 (2012)
56. KM Hosny, Fast computation of accurate Gaussian-Hermite moments for image processing applications. *Digit. Signal Proc.* **22**, 476–485 (2012)
57. RA Johnson, DW Wichern, *Applied Multivariate Statistical Analysis*, 1st edn. (Prentice-Hall, NJ, 1982)
58. AT&T Database. <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>. Accessed date August 18, 2014
59. T Guha, R Ward, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. A sparse reconstruction based algorithm for image and video classification (IEEE Signal Processing Society Kyoto, Japan, 2012), pp. 3601–3604. doi:10.1109/CASSP.2012.6288695
60. J Kittler, Statistical pattern recognition in image analysis. *J. Appl. Stat.* **21**(1–2), 61–75 (1994)
61. RO Duda, PE Hart, *Pattern Classification and Scene Analysis*. (Wiley & Sons, New York, 1973)
62. I Rish, J Hellerstein, J Thathachar, *An analysis of data characteristics that affect naive Bayes performance*. (Technical Report RC21993, IBM T.J. Watson Research Center, New York, USA, 2001), pp. 103–130
63. CE Thomaz, DF Gillies, RQ Feitosa, Using mixture covariance matrices to improve face and facial expression recognitions. *Pattern Recogn. Lett.* **24**, 2159–2165 (2003)
64. CE Thomaz, DF Gillies, RQ Feitosa, A new covariance estimate for Bayesian classifiers in biometric recognition. *IEEE Trans. Circ. Syst. Video Tech.* **14**(2), 214–223 (2004)
65. FRGC Database. <http://www.nist.gov/itl/iad/ig/frgc.cfm>. Accessed date August 18, 2014
66. PJ Phillips, H Moon, SA Rizvi, PJ Rauss, The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
67. GB Huang, M Ramesh, T Berg, EL Miller, *Labeled faces in the wild: A Database for Studying Face Recognition in Unconstrained Environments*. (University of Massachusetts, Amherst, Massachusetts, USA, 2007), pp. 7–49
68. L Wolf, T Hassner, I Maoz, in *Proc. Computer Vision and Pattern Recognition*. Face recognition in unconstrained videos with matched background similarity (IEEE Computer Society Providence, RI, 2011), pp. 529–534. doi:10.1109/CVPR.2011.5995566
69. PJ Phillips, PJ Flynn, T Scruggs, KW Bowyer, J Chang, K Hoffman, J Marques, J Min, W Worek, in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. Overview of the face recognition grand challenge (IEEE Computer Society San Diego, CA, USA, 2005), pp. 947–954
70. P Viola, MJ Jones, Robust real-time face detection. *Int. J. Comp. Vision.* **57**(2), 137–154 (2004)
71. T Ahonen, A Hadid, M Pietikainen, in *Lecture Notes in Computer Science: Computer Vision - Euro. Conf. Computer Vision*. Face recognition with local binary patterns, vol. 3021 (Springer-Verlag Prague, Czech Republic, 2004), pp. 769–481
72. T Ahonen, A Hadid, M Pietikainen, Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
73. Z Lei, SZ Li, R Chu, X Zhu, in *Lecture Notes in Computer Science: Advances in Biometrics - Int. Conf. ICB*. Face recognition with local Gabor textures, vol. 4642 (Springer-Verlag Seoul, Korea, 2007), pp. 49–57
74. Y Su, S Shan, X Chen, W Gao, Hierarchical ensemble of global and local classifiers for face recognition. *IEEE Trans. Image Process.* **18**(8), 1885–1896 (2009)
75. GB Huang, V Jain, EL Miller, in *Proc. Int. Conf. Computer Vision*. Unsupervised joint alignment of complex images (IEEE Janeiro, Brazil, 2007), pp. 1–8. doi:10.1109/ICCV.2007.4408858
76. L Wolf, T Hassner, Y Taigman, in *Proc. Eur. Conf. Computer Vision*. Descriptor based methods in the wild (Marseille, France, 2008), pp. 1–14. <http://eccv2008.inrialpes.fr/>
77. N Pinto, JJD Carlo, DD Cox, in *Proc. Computer Vision and Pattern Recognition*. How far can you get with a modern face recognition test set

- using only simple features? (IEEE Computer Society Miami Beach, FL, 2009), pp. 1–8. doi:10.1109/CVPR.2009.5206605
78. H Li, G Hua, Z Lin, J Brandt, J Yang, in *Proc. Computer Vision and Pattern Recognition*. Probabilistic elastic matching for pose variant face verification (IEEE Computer Society Portland, OR, 2013), pp. 1–8. doi:10.1109/CVPR.2013.449
79. J Hu, J Lu, J Yuan, Y-P Tan, in *Lecture Notes in Computer Science - Asian Conference on Computer Vision*. Large margin multi-metric learning for face and kinship verification in the wild, vol. 9005 (Springer-Verlag Singapore, Singapore, 2015), pp. 252–267. doi:10.1007/978-3-319-16811-1_17
80. X Meng, S Shan, X Chen, W Gao, in *Proceedings of IEEE International Conference on Pattern Recognition*. Local visual primitives (LVP) for face modelling and recognition, vol. 2 (IEEE Hong Kong, 2006), pp. 536–539. doi:10.1109/ICPR.2006.773

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
