Neural Comput & Applic (2006) 15: 170–182
DOI 10.1007/s00521-005-0017-7

ORIGINAL ARTICLE

Siu-Yeung Cho · Tommy W. S. Chow

Robust face recognition using generalized neural reflectance model

Received: 12 September 2004 / Accepted: 1 November 2005 / Published online: 2 December 2005

Abstract A generalized neural reflectance (GNR) model for enhancing face recognition under variations in illumination and posture is presented in this paper. Our work is based on training a number of synthesis images of each face taken at single lighting direction with frontal/posture view. This way of synthesizing images can be used to build training cases for each face under different known illumination conditions from which face recognition can be significantly improved. However, reconstructing face shape may not easily be achieved and the human face images usually form by highly complex structure which suffers from strong specular and unknown reflective conditions. In this paper, these limitations are addressed by Cho and Chow (IEEE Trans Neural Netw 12(5):1204–1214, 2002). Face surfaces are recovered by this GNR model and face images in different poses are synthesized to create a database for training. Our training algorithm assigns to recognize the face identity by similarity measure on face features extracting first by the principle component analysis (PCA) method and then further processing by the Fisher's discrimination analysis (FDA) to acquire lower dimensional patterns. Experimental results conducted on the Yale Face Database B show that lower error rates of classification and recognition are achieved under different variations in lighting and pose and the performance significantly outperforms the recognition without using the proposed GNR model.

S.-Y. Cho (✉)
Division of Computing Systems, School of Computer Engineering,
Nanyang Technological University, Nanyang Avenue,
Singapore, Singapore 639798
E-mail: assycho@ntu.edu.sg
Tel.: +65-6790-5491
Fax: +65-6792-6559

T. W. S. Chow
Department of Electronic Engineering,
City University of Hong Kong, Tat Chee Avenue,
Kowloon Tong, Hong Kong, China
E-mail: eetchow@cityu.edu.hk

Keywords Face recognition · Neural networks · Shape from shading

1 Introduction

Face recognition, the art of matching a given face to a database of faces, is a non-intrusive biometric method since 1960s. Demands on face recognition have rapidly increased in industrial applications such as identification for law enforcement and authentication for banking and security system access. Many techniques and useful results have been reported [1–3]. In efforts going back to far early times, current systems (see surveys [2–4]) of face recognition involve extracting geometric or statistical features derived from face images. Despite human being able to recognize and identify faces in a scene with little features, building an efficient and accuracy system is still very challenging. The challenges are that a recognition system has to be invariant both to external changes (such as environmental light, person's position and distance between camera and person) and to internal deformations (such as facial expression, aging and makeup [5]). The performance of existing techniques is still not very consistent and the system may produce results which are very different from those obtained from human visual perception. There are some important issues that need to be addressed. Many researchers have explored geometrical features based methods for face recognition. Kanade [6] presented an automatic feature extraction method based on ratios of distances and reported a recognition rate of between 45% and 75% with a database of 20 people. Brunelli and Poggio [7] extracted a set of geometrical features such as nose width and length, mouth position, and chin shape in which the method could achieve 90% recognition rate on a database of 47 people. Gao and Leung [8] have recently proposed a line-edge map (LEM) features for face coding and used the line segment Hausdorff distance measurement for human face recognition. The results are encouraging with a single public database with over

90% recognition rate under lighting variations but below 75% under pose variations. However, such geometrical featuring methods would be dependent on the accuracy of the feature-locating algorithm. Current algorithms for automatic location of feature points do not provide a high degree of accuracy and require considerable computational capacity [9].

Not only the geometrical features based methods can be used for face identification, but statistical approaches are also used for processing the face image under low-level dimension. Eigenfaces are usually used in [10] and every face in the database can be represented as a vector of weights. They are obtained by projecting the image into eigenface space by a simple inner product operation. The resulting eigenfaces are classified by comparison with known individuals. However, the performance of this method is limited by a number of factors because optimal performance requires a high degree of correlation between the pixel intensities of the training and test images. These limitations are caused by lighting, facial expression and other factors. Most recently, in accordance with the fact that eigenface features are not promisingly good for discrimination, using other discriminant feature paradigms has become more popular. One method is called Fisher's/linear discrimination analysis (FDA) which aims at overcoming the drawback of the eigenface by integrating FDA criteria [11]. FDA training is carried out via scatter matrix analysis in projecting faces from a high-dimensional space to a significantly lower dimensional feature space. Apart from using statistical approaches, neural network based feature extraction has been proposed to develop a compact internal representation of faces [12–14]. Good results were obtained from a database with up to 20 individuals and there is no lighting variation.

Although many face recognition techniques have been able to deliver promising results, the task of robust face recognition remains very difficult [15]. Indeed, there are at least two major problems in the current approaches: the illumination variation problem and the pose variation problem. Either of these two problems may cause significant degradation in the performance of face recognition systems. For instance, changes in illumination conditions can change the 2D appearance of a 3D face object dramatically and hence affect the system performance. These two problems have been reported in many evaluations, e.g., [16, 17]. However, their performances are unacceptable when face images are acquired in an uncontrolled environment such as in surveillance video. The illumination problem is basically illustrated in Fig. 1 where the same face appears differently due to a variation in lighting. The differences introduced by varied illuminations often cause systems to mis-classify input images that have been theoretically proved by Adini et al. [18] for systems based on eigenface projection. Another problem is pose variation. The performance of face recognition may also drop significantly when pose variations are presented. Some works have been proposed to handle the pose problem. They are included into two approaches: first, multiple database images of each person were used by a template-based correlation matching scheme. Second, hybrid methods were used when multiple images are available during training but only one database image per person is available during recognition. The second approach seems to be the most popular one in the literature [19].

In this paper, a generalized neural reflectance (GNR) model is proposed to enhance face recognition in a way of delivering more robust output. Our approach is model based which differs from the other methods in that a single frontal view face image is required to synthesize other face images under different lighting conditions and a small number of face images in different poses are used to synthesize other face postures. This generalized model [20] is established using a hybrid structure of two neural activation functions, i.e., sigmoid and radial basis functions. Based on this model design, the diffuse model's parameters would be generalized by the sigmoid function, whereas the other parameters, such as specular reflectivity, could be approximated by the radial basis function. The radial-based function is selected because of its separable capability in ill-posed hypersurface structure. All components for real face images are generalized by this model, and a set of synthesized face images can then be rendered in different occasions of illuminations and postures. In our study, we only make use of one image in frontal view to estimate its face surface, and thus a set of synthetic face images under different illumination conditions (i.e., different light source directions) can be synthesized. Our method for handling lighting variability in face images differs from [21] in that our model is able to synthesize a large image database. In order to tackle the issue of posture variations affecting the recognition, we synthesize different posture images by image warping dictated by 3D rigid transformations of the reconstructed face surface. Our method differs from the others [22, 23] in that we first estimate the transformation matrices for each different posture and then warp synthetic face surfaces of each posture using its corresponding transformation matrices. Hence, a set of synthetic images under different postures can then be formed. In the recognition stage, a set of most expressive features is generated by the principal component analysis (PCA) to

Fig. 1 Face images from the same person under different illumination conditions

compress each face representation and then the FDA is further implemented to generate a set of the most discriminant features so that different classes of training data can be classified. The identity of a test image can then be measured by means of different kinds of similarity measure. Our recognition approach is performed using a 4,050 images subset of the publicly available Yale Face Database B [24]. This subset contains nine poses multiplied by 45 illumination conditions for 10 individuals.

The paper is organized as follows: Sect. 2 briefly describes a generalized reflectance model by use of a hybrid structure of neural models and shows how to synthesize face images for training under different illuminations and pose. Section 3 describes the face features extraction by means of the PCA and the FDA and also different similarity measures that can be expressed for the face recognition. Section 4 presents experimental results, and finally the conclusion is drawn in Sect. 5.

2 Generalized neural reflectance model

2.1 Face surface reconstruction

It was shown that a $m \times n$ face image can be formed as a convex object in the image space $\Re^{m \times n}$ under arbitrary combinations of points or extended light sources [24]. Assume the surface of a convex object contains Lambertian reflectance surface that reflects light in diffuse reflection. Suppose that the surface, represented by $z(x,y)$, depends on the systematic variation of image brightness with surface orientation, where z is the height field and x and y form the 2D pseudo-plane over the domain Ω of the image plane. The Lambertian reflectance model used to represent a surface illuminated by a single point light source is given as

$$R_{\text{Lambertian}} = \max\left(\eta \mathbf{n}\mathbf{s}^{\mathrm{T}}, \mathbf{0}\right), \tag{1}$$

where $\max\left(\eta \mathbf{n}\mathbf{s}^{\mathrm{T}}, \mathbf{0}\right)$ sets to zero for all negative components. η is the composite albedo, $\mathbf{s} = (\cos\tau \sin\sigma \quad \sin\tau \sin\sigma \quad \cos\sigma)$ is the illuminate source direction and τ and σ denote the tilt and slant angles, respectively. $\mathbf{N} \in \Re^{(m \times n) \times 3}$ is defined to be a matrix whose rows are given as the surface normal, $\mathbf{n}$, represented as

$$\mathbf{n}(x,y) = \left(\frac{-p(x,y)}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \quad \frac{-q(x,y)}{\sqrt{p^2(x,y)+q^2(x,y)+1}} \quad \frac{1}{\sqrt{p^2(x,y)+q^2(x,y)+1}}\right), \tag{2}$$

where $p = \partial z/\partial x$ and $q = \partial z/\partial y$ are the surface gradients. An ideal Lambertian surface requires a known and distant light source according to this model. But in most practical cases, the surface does not often contain the Lambertian surface because the light source is often located at a finite distance and an unknown position.

Specular component occurs when the incident angle of the light source is equal to the reflected angle. This component is formed by two terms: the specular spike and the lobe. Healy and Binford [25] derived the specular model by simplifying the Torrance–Sparrow model [26], in which the Gaussian distribution was used to model the facet orientation function. More sophisticated model based on the geometrical optics approach was also presented as the specular reflectance model [27], such that

$$R_{\text{Specular}} = \kappa_{\text{spec}} \frac{L_i dw_i}{\cos\theta_{\text{r}}} \exp\left(-\frac{\alpha^2}{2\beta^2}\right), \tag{3}$$

where κ_{spec} represents the fractions of incident energy determined by the Fresnel coefficients and the geometrical attenuation factor. The term $\cos\theta_{\text{r}}$ describes the emitting angle that the radiance of the surface in the viewing direction is determined. As most object surfaces in the real world are neither purely Lambertian reflectance models nor purely specular components, they are a linear combination of them. They are hybrid surfaces that include diffuse and specular components. Nayar et al. [28] formed a hybrid model to tackle the problem such that the model consists of three components: diffuse lobe, specular lobe and specular spike. In contrast, this paper describes a straightforward representation of the hybrid surface that the total intensity of the hybrid surface is the summation of the specular intensity and the Lambertian (diffuse) intensity as follows:

$$R_{\text{Hybrid}} = (1-\omega)R_{\text{Lambertian}} + \omega R_{\text{Specular}}, \tag{4}$$

where R_{Hybrid} is the total intensity for the hybrid surface, $R_{\text{Lambertian}}$ and R_{Specular} are the diffuse intensity and specular intensity, respectively, and ω the weight of the specular component.

Although these models are widely used for the approximation of the reflectance components, the critical parameters (i.e., the light source and the viewing direction) are required a priori. Incorporating more reflectance parameters and effects is inevitable for generating a GNR model. In this paper, a neural network self-learning scheme, based on the relationship between the surface orientation and the intensity, is exploited to model the unknown parameters for generalizing the reflectance model. Apparently the use of a sigmoid activation model and a radial basis function model can provide approximations of the Lambertian model and the specular model, respectively, under the theoretical view of the universal approximation capability of neural networks [29, 30]. These are clearly advantages of establishing a hybrid-type neural reflectance model [20],

which combines the sigmoid and radial basis functions. The GNR model is shown in Fig. 2 and expressed as

$$R_{\mathrm{GNR}} = \varphi_{\mathrm{sig}}\left(v_0 + \sum_{k=1}^{N} v_k\left(\varphi_{\mathrm{sig}}\left(\mathbf{w}_k\mathbf{a}_{i,j}\right) + \varphi_{\mathrm{rad}}\left(\left\|\mathbf{a}_{i,j} - \mathbf{c}_k\right\|\right)\right)\right), \quad (5)$$

where ϕ_{sig} and ϕ_{rad} are the sigmoid activation function and radial basis function, respectively, $\mathbf{w}_k$ is the synapse weights of the sigmoid activation and $\mathbf{c}_k$ is the centers of the radial basis function. The input vector $\mathbf{a}_{i,j} = (p_{i,j} \;\; q_{i,j})^{\mathrm{T}}$ acts as surface gradient vector in (i,j) coordinates of a face surface. The surface gradients would be optimized to form the optimal reflectance model $\hat{R}_{\mathrm{GNR}}$ such that this model is equivalent to the given intensity image. This approach would enable us to generalize either the purely Lambertian surface or the non-Lambertian surfaces, which are most existing in the convex surface of the face images. Using the above GNR model, the face surface orientations can be reconstructed from the intensity image in frontal view by solving a shape from shading (SFS) problem. In solving the SFS algorithm by the GNR model, the cost function we commonly used is as shown in (6).

$$E_{\mathrm{T}} = \int\int_{\Omega} (I - R_{\mathrm{GNR}})^2 + \lambda\left(\left(\frac{\partial p}{\partial x}\right)^2 + \left(\frac{\partial p}{\partial y}\right)^2 + \left(\frac{\partial q}{\partial x}\right)^2 + \left(\frac{\partial q}{\partial y}\right)^2\right)\mathrm{d}x\,\mathrm{d}y. \quad (6)$$

The first term is the intensity error term, and the second term is a smoothness constraint given by the spatial derivatives of p and q. λ is a scalar that assigns a positive smoothness parameter. Based on this objective function, the free parameters of the GNR model and the object surface gradients are determined by performing a unified learning mechanism. Through the learning process, the synapse weights, the radial basis centers as well as the surface depths are optimized by a specific learning framework which has been reported in [20, 31]. The general learning framework is shown in Fig. 3. Throughout this learning framework, given an intensity image I, an error between the given intensity and the neural reflectance model can be computed and used to optimize the R_{GNR} parameters as well as the surface gradients by the cost function (6). The R_{GNR} parameters are optimized by a specific learning algorithm for the corresponding neural network structure and the surface gradients are computed by a variational calculus approach on a discrete grid of points. Hence, the object surface depth can be estimated once the optimal surface gradients are being obtained. The details of this framework computation can be referred to in [20] or [31].

The whole learning framework can normally converge within 10–20 iterations. With the cropped face image of size 100 × 100 pixels in the frontal view, the convergence took at most 1 min running by a workstation with P-IV 2.4 GHz processor. Figure 4 demonstrates the reconstructed face surfaces by the proposed GNR model. Figure 4a shows the original single light source images with the frontal view of 10 individuals in the Yale Database B. The light source direction was chosen with 12° of the optical axis which the images do exhibit as little as shadowing, such as left and right of the nose. Figure 4b shows the reconstructed face surfaces by the GNR model for these corresponding individuals. These face surfaces can encode the corresponding surface normal fields which are synthesized face images under arbitrary illumination conditions. We used the surface normal field obtained by the reconstructed surface to project the reflected intensities to the image plane by the arbitrary lighting directions for getting face images under different illumination scenarios. Figure 5 demonstrates some samples of synthesized images of these 10 persons in different lighting

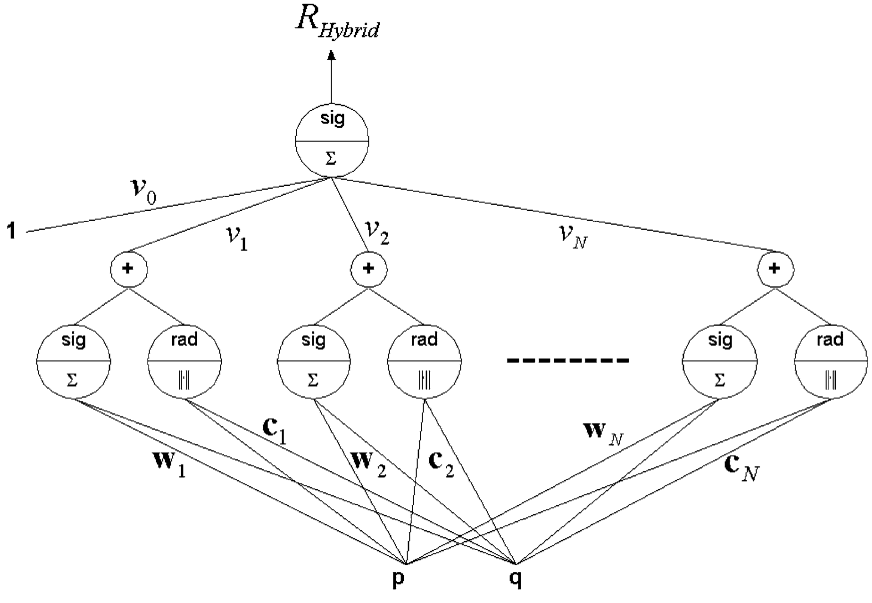

Fig. 2 Hybrid of sigmoid and radial basis activations for GNR model

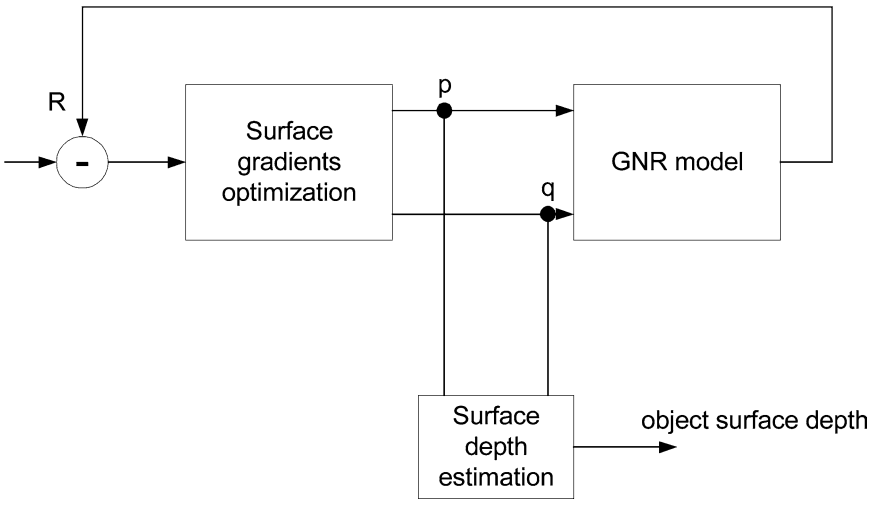

Fig. 3 The general learning framework for generating the object surface depth

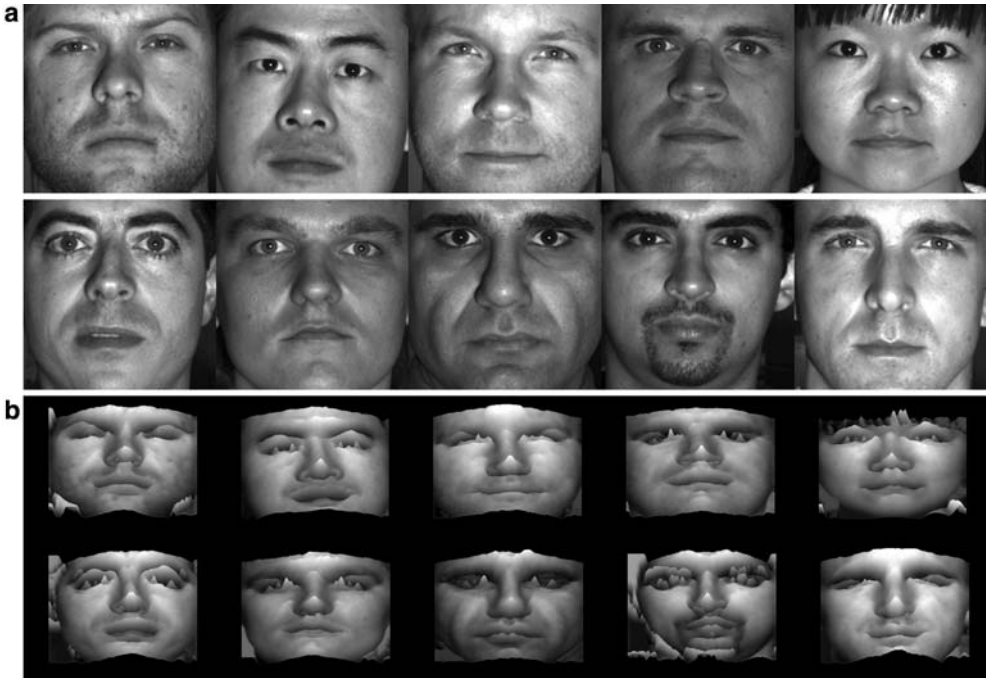

Fig. 4 The reconstruction results by the proposed GNR model. **a** The training images from 10 individuals under frontal pose and illumination in Yale Database B (note that all the training images are cropped in size of 100×100 pixels). **b** The reconstruction surfaces by the SFS in the proposed GNR model

directions. Note that we did not synthesize the most extreme illumination scenario for training, thus this extreme scenario was also ignored during testing for the recognition algorithm.

2.2 3D rigid transformation for pose variations

The posture variation is another issue to degrade the accuracy of the face recognition. In order to handle this problem, we warp the face images defining from given views using 3D rigid transformations of the surface geometry. This method for handling posture variation differs from other works such as [22] and [23], in which we warp a given pose image to another pose image of the same person's face using the estimated rigid transformation. In this method, estimations of a set of 3D rigid transformation are a key component for generating a collection of face images in pose variations. This 3D rigid transformation is basically a composition matrix of rotations and translations, which are computed by the frontal-to-pose and/or pose-to-pose surface geometries. Figure 6 shows the face surfaces of the same individual for estimating the parameters in the rigid transformation in that they have been reconstructed by the GNR model. Suppose a 3D dominant point defined by its homogeneous coordinates $(X^{\mathrm{f}}, Y^{\mathrm{f}}, Z^{\mathrm{f}}, 1)^{\mathrm{T}}$ in a face surface at a given posture and let $(kX^{\mathrm{p}}, kY^{\mathrm{p}}, kZ^{\mathrm{p}}, k)^{\mathrm{T}}$ be the corresponding point's coordinates at a pre-defined target posture after warping by the rigid transformation. The transformation from the frontal/posture to another posture is equivalent to saying that it is a composition

Fig. 5 Synthesized images of the 10 individuals under different illumination conditions

matrix of 3D rotations $\left(\mathbf{R}_{\theta_x}, \mathbf{R}_{\theta_y}\right)$ by x-axis and y-axis, respectively, and 3D translations ($\mathbf{T}$) such that

$$\begin{pmatrix} kX^{\mathrm{p}} \\ kY^{\mathrm{p}} \\ kZ^{\mathrm{p}} \\ k \end{pmatrix} = \mathbf{R}_{\theta_x}\mathbf{R}_{\theta_y}\mathbf{T}\begin{pmatrix} X^{\mathrm{f}} \\ Y^{\mathrm{f}} \\ Z^{\mathrm{f}} \\ 1 \end{pmatrix} = \mathbf{M}_{\mathrm{c}}\begin{pmatrix} X^{\mathrm{f}} \\ Y^{\mathrm{f}} \\ Z^{\mathrm{f}} \\ 1 \end{pmatrix}, \tag{7}$$

where

$$\mathbf{M}_{\mathrm{c}} = \begin{pmatrix} m_1 & 0 & m_2 & m_3 \\ m_4 & m_5 & 0 & m_6 \\ m_7 & m_8 & 0 & m_9 \\ m_{10} & m_{11} & m_{12} & 1 \end{pmatrix}$$

is the composition matrix in which the parameters $(m_1, \ldots, m_{12})$ are estimated for the transformation from the frontal-to-pose or pose-to-pose face surfaces regeneration. In each component, (7) can be written as

$$\begin{aligned} X^{\mathrm{p}} &= \frac{m_1 X^{\mathrm{f}} + m_2 Z^{\mathrm{f}} + m_3}{m_{10}X^{\mathrm{f}} + m_{11}Y^{\mathrm{f}} + m_{12}Z^{\mathrm{f}} + 1}, \\ Y^{\mathrm{p}} &= \frac{m_4 X^{\mathrm{f}} + m_5 Y^{\mathrm{f}} + m_6}{m_{10}X^{\mathrm{f}} + m_{11}Y^{\mathrm{f}} + m_{12}Z^{\mathrm{f}} + 1}, \\ Z^{\mathrm{p}} &= \frac{m_7 X^{\mathrm{f}} + m_8 Y^{\mathrm{f}} + m_9}{m_{10}X^{\mathrm{f}} + m_{11}Y^{\mathrm{f}} + m_{12}Z^{\mathrm{f}} + 1}. \end{aligned} \tag{8}$$

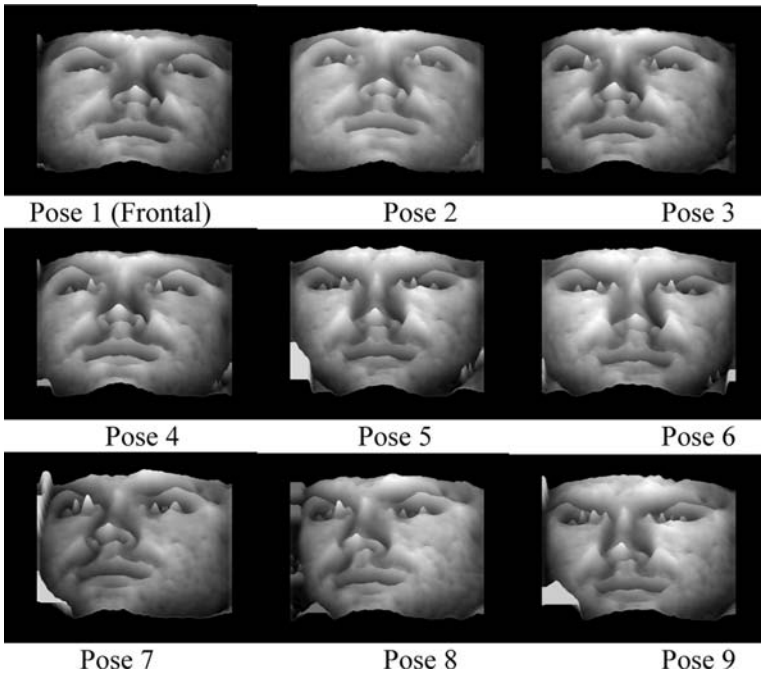

Fig. 6 The face surfaces of the same individual reconstructed by the proposed reflectance model under different postures. Note that Postures 1, 3, 5, 7 are used for original postures to be transformed and Postures 2, 4, 6, 8 are used as target postures from the transformation matrices. These sets of face surfaces are basically used for estimating the rigid transformations

Since (8) can be thought of as three linear equations for the 12 unknowns $\mathbf{m} = (m_1, m_2, \ldots, m_{12})^{\mathrm{T}}$, they can be written as

$$\begin{pmatrix} X^{\mathrm{p}} \\ Y^{\mathrm{p}} \\ Z^{\mathrm{p}} \end{pmatrix} = \begin{pmatrix} X^{\mathrm{f}} & Z^{\mathrm{f}} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -X^{\mathrm{f}}X^{\mathrm{p}} & -Y^{\mathrm{f}}X^{\mathrm{p}} & -Z^{\mathrm{f}}X^{\mathrm{p}} \\ 0 & 0 & 0 & X^{\mathrm{f}} & Y^{\mathrm{f}} & 1 & 0 & 0 & 0 & -X^{\mathrm{f}}Y^{\mathrm{p}} & -Y^{\mathrm{f}}Y^{\mathrm{p}} & -Z^{\mathrm{f}}Y^{\mathrm{p}} \\ 0 & 0 & 0 & 0 & 0 & 0 & X^{\mathrm{f}} & Y^{\mathrm{f}} & 1 & -X^{\mathrm{f}}Z^{\mathrm{p}} & -Y^{\mathrm{f}}Z^{\mathrm{p}} & -Z^{\mathrm{f}}Z^{\mathrm{p}} \end{pmatrix} \mathbf{m}. \tag{9}$$

A least squares equation can be formulated for the N points on the face surfaces and can be written as

$$\mathbf{P} - \mathbf{H}\mathbf{m} = 0, \tag{10}$$

where $\mathbf{P}$ is a 3N×1 column vector given by $\mathbf{P} = \left(X_1^{\mathrm{p}} \;\; Y_1^{\mathrm{p}} \;\; Z_1^{\mathrm{p}} \;\; \cdots \;\; X_N^{\mathrm{p}} \;\; Y_N^{\mathrm{p}} \;\; Z_N^{\mathrm{p}}\right)^{\mathrm{T}}$ and $\mathbf{H}$ is a $3N$×12 matrix given by

$$\mathbf{H} = \begin{pmatrix} X_1^{\mathrm{f}} & Z_1^{\mathrm{f}} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -X_1^{\mathrm{f}}X_1^{\mathrm{p}} & -Y_1^{\mathrm{f}}X_1^{\mathrm{p}} & -Z_1^{\mathrm{f}}X_1^{\mathrm{p}} \\ 0 & 0 & 0 & X_1^{\mathrm{f}} & Y_1^{\mathrm{f}} & 1 & 0 & 0 & 0 & -X_1^{\mathrm{f}}Y_1^{\mathrm{p}} & -Y_1^{\mathrm{f}}Y_1^{\mathrm{p}} & -Z_1^{\mathrm{f}}Y_1^{\mathrm{p}} \\ 0 & 0 & 0 & 0 & 0 & 0 & X_1^{\mathrm{f}} & Y_1^{\mathrm{f}} & 1 & -X_1^{\mathrm{f}}Z_1^{\mathrm{p}} & -Y_1^{\mathrm{f}}Z_1^{\mathrm{p}} & -Z_1^{\mathrm{f}}Z_1^{\mathrm{p}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N^{\mathrm{f}} & Z_N^{\mathrm{f}} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -X_N^{\mathrm{f}}X_N^{\mathrm{p}} & -Y_N^{\mathrm{f}}X_N^{\mathrm{p}} & -Z_N^{\mathrm{f}}X_N^{\mathrm{p}} \\ 0 & 0 & 0 & X_N^{\mathrm{f}} & Y_N^{\mathrm{f}} & 1 & 0 & 0 & 0 & -X_N^{\mathrm{f}}Y_N^{\mathrm{p}} & -Y_N^{\mathrm{f}}Y_N^{\mathrm{p}} & -Z_N^{\mathrm{f}}Y_N^{\mathrm{p}} \\ 0 & 0 & 0 & 0 & 0 & 0 & X_N^{\mathrm{f}} & Y_N^{\mathrm{f}} & 1 & -X_N^{\mathrm{f}}Z_N^{\mathrm{p}} & -Y_N^{\mathrm{f}}Z_N^{\mathrm{p}} & -Z_N^{\mathrm{f}}Z_N^{\mathrm{p}} \end{pmatrix}.$$

As $N >> 12$ and the N points are not coplanar, the solution would be non-trivial which can be determined by a simple least squares method, such that the estimated $\mathbf{m}$ is given by

$$\hat{\mathbf{m}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{P}. \tag{11}$$

By the above computations, we first identified a number of dominant points (around 100 points on the face surface) at the positions of eyes, nose and mouth for each posture of the given surfaces. All these points are substituted in (11) to compute the estimated frontal-to-pose and pose-to-pose transformation matrices. In our work, we selected the frontal (Posture 1) and the Postures 3, 5, 7 as the given postures to be transformed into the target Postures 2, 4, 6 and 8, respectively, such that four transformation matrices are computed correspondingly as transformed from Posture 1 into 2, Posture 3 into 4,

Fig. 7 Some examples of synthesized face images by different posture transformations

Posture 5 into 6 and Posture 7 into 8. Using these four transformation matrices, the surface geometrics of the postures can then be evaluated and hence the face images are synthesized in different posture variations. Figure 7 shows some examples of the face images synthesized by these four posture transformations. Note that these images were generated from the face surfaces of the original postures (1, 3, 5, 7) from the corresponding individuals together with the 3D rigid transformations estimated by the given face surfaces from one person shown in Fig. 6. All the postures are fixed for every individual and they have only small changes in eyes, nose and mouth. Nevertheless, in our method, we use prior knowledge about the shape variations of faces in different postures to resolve the transformation matrices. The disadvantage is that we cannot generate the faces in infinite degree of freedom because restricted prior knowledge can only be obtained in advance. Our method is different from the other 3D face reconstruction methods such as in [32] which can generate a whole 3D face such that the face images can then be warped under different postures arbitrarily. Furthermore, our method cannot attempt to deal with facial expression, aging or occlusion.

3 Extraction and recognition of face features

In our study, about 80–120 synthesized images in each posture were more or less sufficient to provide for the different illuminate conditions. The next step is to extract the face features from all these synthesized images to provide a representation for recognition. One of the simple ways is that the whole face representation is projected down to a moderate-dimensional linear subspace in order to reduce the complexity and speed up the recognition process. Basically, the basis vectors of this subspace, which is specific to faces, are commonly computed by performing PCA in which those basis vectors have been scaled by their corresponding eigenvalues. We then selected the eigenvectors corresponding to the largest eigenvalues to be the basis vectors of the faces. We intended this as an approximation to find the basis vectors by performing PCA directly on all the synthesized images of the face under different illuminate conditions. In the simulations described in Sect. 4, the subspace of each face had a dimension of 100 as this was good enough to specify all the variabilities in the different illuminate conditions and different postures.

3.1 Principal component analysis

Let a face image $\mathbf{X}_i$ be a two-dimensional $m \times n$ array of intensity values. An image may also be considered as a vector of dimension m^2. Suppose that there are n face images used for training $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n) \subset \Re^{m^2 \times n}$, and assume that each image belongs to one of classes c. The covariance matrix is defined as

$$\Psi = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{X}_i - \mathbf{X}\right)\left(\mathbf{X}_i - \mathbf{X}\right)^{\mathrm{T}} = \Phi\Phi^{\mathrm{T}}, \tag{12}$$

where

$$\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \quad \text{and } \Phi = (\Phi_1, \Phi_2, \ldots, \Phi_n) \subset \Re^{m^2 \times n}. \tag{13}$$

The eigenvalues and eigenvectors of the covariance matrix Ψ are calculated. Let $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_r) \subset \Re^{m^2 \times r}$ where $r < n$ be the r eigenvectors corresponding to the r largest eigenvalues. Thus, for a set of original face images $\mathbf{X}$, their corresponding eigenface feature $\mathbf{Y} \subset \Re^{r \times n}$ can be obtained by projecting $\mathbf{X}$ into the linear subspace (i.e., eigenface space) as

$$\mathbf{Y} = \mathbf{A}^{\mathrm{T}}\mathbf{X}. \tag{14}$$

In the recognition process, a test face image $\mathbf{X}_j$ is performed by first projecting $\mathbf{X}_j$ to $\mathbf{Y}_j$ at the eigenface space and then computing the metric to the eigenface representation of each face $\mathbf{X}$ in the database. This metric is defined as a similarity measure to the closest projected eigenface space $\mathbf{Y}_j$ within $\mathbf{Y} \subset \Re^{r \times n}$. The face $\mathbf{X}_j$ is then assigned the identity of the closest representation.

However, the PCA paradigm does not provide any information for class discrimination. It means that the scatter being maximized is due to not only the between-class scatter that is useful for classification, but also the within-class scatter that is unwanted information. Accordingly, the Fisher's discrimination analysis (FDA) is applied to the projection of the set of training samples in the eigenface space and then it finds an optimal subspace for classification in which the ratio of the between-class scatter and the within-class scatter is maximized [11, 33].

3.2 Fisher's discrimination analysis

Let the between-class scatter matrix be defined as

$$\mathbf{S}_{\mathrm{B}} = \sum_{i=1}^{c} n_i \left(\mathbf{X}_i - \mathbf{X}\right)\left(\mathbf{X}_i - \mathbf{X}\right)^{\mathrm{T}}, \tag{15}$$

and the within-class scatter matrix be defined as

$$\mathbf{S}_{\mathrm{W}} = \sum_{i=1}^{c} \sum_{\mathbf{X}_i \in n_i} \left(\mathbf{X}_i - \mathbf{X}_i\right)\left(\mathbf{X}_i - \mathbf{X}_i\right)^{\mathrm{T}}, \tag{16}$$

where $\mathbf{X}_i$ is the mean image of class $\mathbf{X}_i$ and n_i is the number of samples in class $\mathbf{X}_i$. The optimal subspace, $\mathbf{Z}_{\mathrm{opt}}$, by the FDA is determined as follows [11]:

$$\begin{aligned}\mathbf{Z}_{\mathrm{opt}} &= \operatorname*{argmax}_{\mathbf{E}} \frac{\left|\mathbf{E}^{\mathrm{T}}\mathbf{S}_{\mathrm{B}}\mathbf{E}\right|}{\left|\mathbf{E}^{\mathrm{T}}\mathbf{S}_{\mathrm{W}}\mathbf{E}\right|} \\ &= \left[\mathbf{z}_1 \quad \mathbf{z}_2 \quad \cdots \quad \mathbf{z}_r\right],\end{aligned} \tag{17}$$

where $\{\mathbf{z}_i | i = 1, 2, \ldots, r\}$ is the set of generalized eigenvectors of $\mathbf{S}_{\mathrm{B}}$ and $\mathbf{S}_{\mathrm{W}}$ corresponding to the r largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, r\}$, i.e.,

$$\mathbf{S}_{\mathrm{B}}\mathbf{z}_i = \lambda_i \mathbf{S}_{\mathrm{W}}\mathbf{z}_i, \quad i = 1, 2, \ldots, r. \tag{18}$$

Note that there are at most $c - 1$ non-zero generalized eigenvalues, and so an upper bound on r is $c - 1$, where c is the number of classes.

In the face recognition problem, it is difficult that the within-class scatter matrix is always singular. Thus, the rank of $\mathbf{S}_{\mathrm{W}} \leq \min\{r, c(n_i - 1)\}$, in general the value of r, should be smaller than $n_i - c$. On the other hand, in the rank of $\mathbf{S}_{\mathrm{B}} \leq \min\{r, c - 1\}$, there are at most $c - 1$ non-zero generalized eigenvectors. In other words, the FDA transforms the r-dimensional space into $(c-1)$-dimensional space to classify c classes of faces. In order to overcome the complication of a singular $\mathbf{S}_{\mathrm{W}}$, we propose an alternative feature extraction, which is achieved using PCA to reduce the dimension of the feature space $N - c$ and then applying the standard FDA to reduce the dimension to $c - 1$. Thus, the feature vectors $\mathbf{f}_{\mathrm{query}}$ for any query face images $\mathbf{X}_{\mathrm{query}}$ in the most discriminant sense can be calculated as follows:

$$\mathbf{f}_{\mathrm{query}} = \mathbf{Z}_{\mathrm{opt}}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{X}_{\mathrm{query}}. \tag{19}$$

Basically, it is noted that the FDA is a linear transformation which maximizes the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix of the projected samples. The results are globally optimal for linear separable data. Moreover, the separability criterion is not directly related to the classification accuracy in the output space.

3.3 Similarity measures

For recognition and classification purposes, we define the L_1, L_2 and cosine metrics for similarity measures and the nearest mean neighbor classification rule for face recognition after obtaining the feature vectors from the PCA/FDA paradigm. The nearest neighbor classification rule is defined as follows:

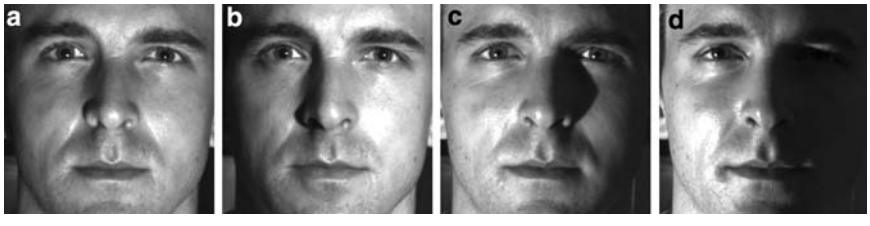

Fig. 8 Example images in four subsets for experiments in variations of lighting: **a** Subset 1's image (up to 12° of the optical axis), **b** Subset 2's image (up to 25° of the optical axis), **c** Subset 3's image (up to 35° of the optical axis), **d** Subset 4's image (up to 50° of the optical axis)

$$\sigma(\mathbf{f}, \mu_k) = \underset{j}{\operatorname{argmin}} \sigma(\mathbf{f}, \mu_j) \rightarrow \mathbf{f} \in \omega_k. \quad (20)$$

The feature vector $\mathbf{f}$ is classified to the class of the closest mean μ_k based on the similarity measure σ. Similarity measures used in our studies include σ_{L_1}, σ_{L_2} and $\sigma_{\cos}$ that, respectively, denote the L_1 distance measure, L_2 distance measure and cosine similarity measure, which are defined as follows:

$$\sigma_{L_1}(\alpha, \beta) = \sum_i |\alpha_i - \beta_i|, \quad (21)$$

$$\sigma_{L_2}(\alpha, \beta) = (\alpha - \beta)^{\mathrm{T}}(\alpha - \beta), \quad (22)$$

$$\sigma_{\cos}(\alpha, \beta) = \frac{-\alpha^{\mathrm{T}}\beta}{\|\alpha\|\|\beta\|}, \quad (23)$$

where $\|\cdot\|$ denotes the norm operator.

4 Experimental results

We assess the feasibility and performance of our proposed face recognition method performing on the Yale Face Database B [24]. We performed two experiments in this database. In the first one, tests were performed under variable illumination but fixed frontal view, and the goal was to compare the face representation after reconstructing by the GNR model and without reconstructing by the GNR model. The second experiment was performed under variations in both illumination and posture. It is used to determine the recognition capability while 3D rigid transformations were computed. As demonstrated, our proposed face recognition more effectively handles variations both in lighting and posture.

4.1 Face database

The face recognition simulations were performed on the Yale Face Database B [24]. In this database, face images were captured by a constructed geodesic lighting rig with 64 computer-controlled xenon strobes, whose positions are in spherical coordinates. The set-up was established in Yale University. With this rig, the illumination could be modified at frame rate and images captured under variable illumination and postures. Ten individuals' face images were acquired under 64 different lighting conditions in nine postures (a frontal posture and eight postures at different angles from the camera axis). Of these 64 images per person in each posture, 40 were selected in our experiments, for a total of 3,600 images. The original size of the images is 640×480 pixels. In order to extract the facial region, all images were manually cropped to include the face with as little hair as possible. All cropped images were scaled in dimension of 100×100 pixels using manually detected centers of the eyes.

4.2 Experimental results of variation in illumination conditions

In this experiment, four subsets of face images at frontal view were divided. Figure 8 shows the sample images in these four subsets. We performed the recognition using 80 images (8 images per face) in Subset 1 as training and the other three subsets as testing in the variations of lighting. These three subsets are sampled in 100 images (10 images per face) in Subset 2, 120 images (12 images per face) in Subset 3 and 100 images (10 images per face) in Subset 4 where the light source directions were chosen with up to 25°, 35° and 50° of the optical axis, respectively. This experimental study, where only illumination varies while posture remains fixed (i.e., frontal), was designed to compare the performance in between the same recognition method using the GNR model and without using the GNR model to reconstruct face surfaces from Subset 1. As the face images under different illumination scenarios were synthesized from the corresponding face surfaces reconstructed by the GNR model, the training size increased to around 3,200 synthesis images (320 images per face), which apparently

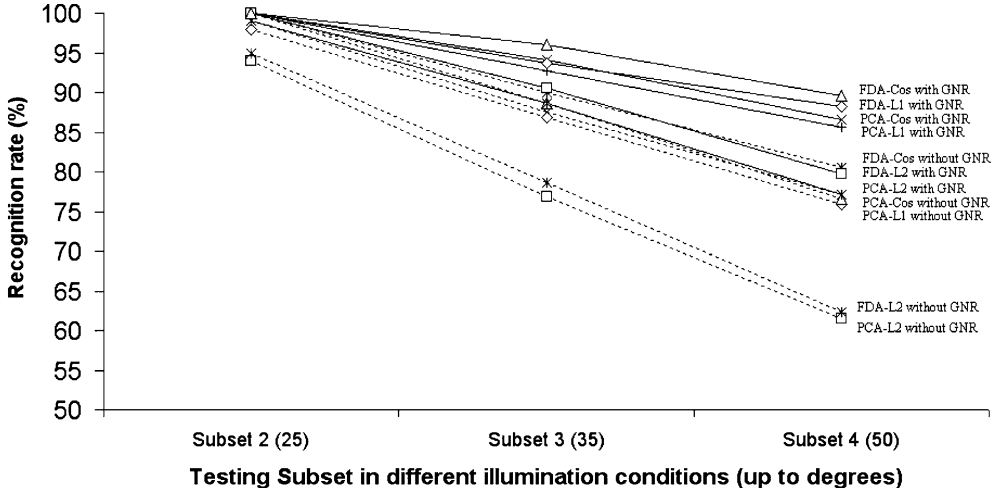

Fig. 9 Recognition rates in different illumination conditions. Each recognition method was trained with and without using GNR model to reconstruct face surfaces from Subset 1. The testing simulations were conducted under images in subsets under different lighting conditions, i.e., Subset 2, Subset 3 and Subset 4

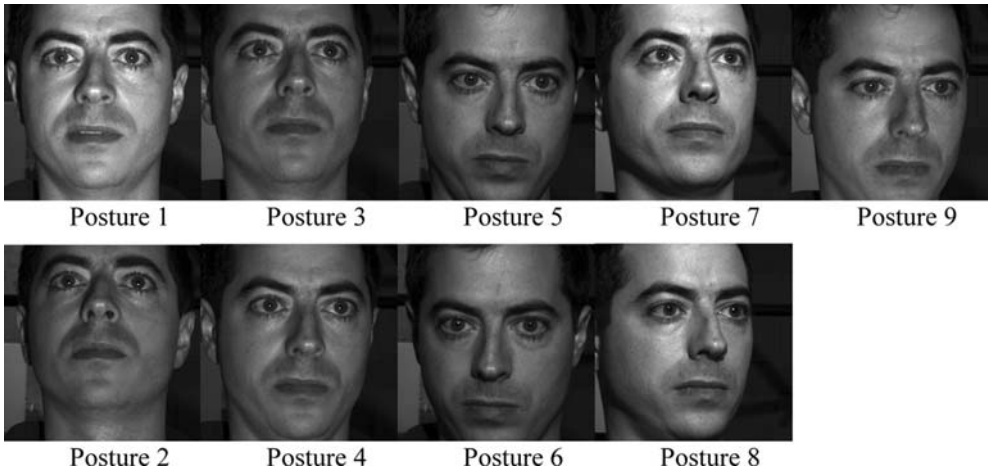

Fig. 10 Example images of the same person in the nine different postures. Note that images of Postures 1–5 in the first row are used for training and images of Postures 6–9 in the second row are used for testing

Table 1 Recognition rates in posture variations

Method	Posture 2	Posture 4	Posture 6	Posture 8	Average (overall) (%)	Variance (overall) (%)
PCA-L1 without GNR	53.6	83.6	59.6	58.3		
FDA-L1 without GNR	57.8	84.5	60.6	59.4		
PCA-L2 without GNR	48.1	77.9	59.8	60.0	63.7	12.2
FDA-L2 without GNR	50.3	79.1	61.2	62.1		
PCA-Cos without GNR	56.8	86.4	55.9	53.4		
FDA-Cos without GNR	58.7	88.1	57.4	55.5		
PCA-L1 with GNR	93.9	96.1	90.4	98.7		
FDA-L1 with GNR	94.6	96.7	91.3	98.7		
PCA-L2 with GNR	83.6	83.6	88.5	92.3	92.6	4.9
FDA-L2 with GNR	85.2	85.4	90.4	93.3		
PCA-Cos with GNR	95.3	96.1	87.9	98.7		
FDA-Cos with GNR	95.6	97.1	89.5	98.7		

Two recognition methods, principle component analysis (PCA) and Fisher's discrimination analysis (FDA), and three similarity measures, L1: σ_{L_1} metric in (21), L2: σ_{L_2} metric in (22), Cos: $\sigma_{\cos}$ metric in (23), are trained with and without using the GNR model

Table 2 Error rates for different postures against different lighting conditions (slight and extreme) using the recognition methods with our proposed GNR model

Lighting variations	Posture variations (%)			
	Posture 1 (frontal)	Postures 2, 3, 4, 5, 6	Postures 7, 8, 9	All postures
Slight lighting conditions (up to 20°)	0.1	3.8	0.74	2.5
Extreme lighting conditions (up to 45°)	8.5	11.7	11.9	11.7

represent all the illumination scenarios. The accuracy of recognition methods can then be dramatically improved by training these synthesis images even though the training size is rather greater than the original. Figure 9 shows the results from these simulations. Notice that the face recognition rates in extreme lighting cases (i.e., Subset 4) were about 88 and 90%, respectively, corresponding to the PCA and the FDA representations by training the 3,200 synthesis images, whereas the recognition rate was about 76% by only training the original images in Subset 1. The results support that whenever using this GNR model, good recognition rates are achieved by means of face representation in low-dimensional subspaces in approximation of the different illumination conditions.

4.3 Experimental results of variation in postures

We conducted recognition experiment under variations in pose using images from all nine postures in the database. Figure 10 shows the images of a same person in these nine postures. The goal of this experiment is to test the performance of the face representation processed

by the GNR model together with the 3D rigid transformation onto the face surfaces. We performed the recognition methods using the first five postures for training and the other four postures for testing. Images in training set of the five postures were taken under light source directions with up to 15° of the optical axis and images in testing set of the other four postures were captured under up to 25° lighting directions. Similar to the experiment for variation in lighting, we used the 3D rigid transformation to generate face surfaces of the other four postures and subsequently synthesized face images in all nine postures under different lighting conditions. In our approach, the training size was increased up to 1,800 face images (9×20 images per person). We compared two recognition methods (i.e., PCA and FDA) with and without using the GNR model and the rigid transformation. The recognition results are tabulated in Table 1. The overall recognition rate, in average about 92%, obtained by the different recognition methods using GNR model is much better than that, in average about 64%, obtained by those methods without GNR model. This demonstrates that the need for the GNR model explicitly overcomes the face image variability due to changes in lighting and posture simultaneously. In particular, the method of FDA in $\sigma_{\cos}$ similarity measure using the GNR approach performs reasonably well in comparison to the same method without using the GNR approach. Moreover, Table 2 presents the error rates by our approach for all postures with different lighting conditions, except for the most extreme cases in the whole database. This shows that our approach performs with less than 3% error for all postures in slight lighting condition and approximately 10% error in extreme variations in lighting directions up to 45° of the optical axis.

5 Conclusion

In this paper, we addressed the major problems in face recognition by variations in lighting and postures. We presented a framework of face recognition which requires a small number of images of a face in several fixed postures and illuminated by a single point light source at unknown positions to generate a rich representation of the face useful for recognition. The main idea of our method is to make use of the proposed GNR model to transform the given faces into reconstructed 2.5D face surfaces and use the 3D rigid transformation to transform the reconstructed face surfaces from given postures into pre-defined target postures. Using all these reconstructed face surfaces, a full set of face images can then be synthesized under different illumination conditions and postures. These synthesized images are used for training in many cases in variation of lighting and posture. The recognition method, in this paper, simply uses the PCA to reduce the dimensionality of face representation and further uses the FDA to enhance the classification. The experimental results demonstrate that the performance of face recognition is markedly improved after using the proposed GNR reconstruction and transformation. We believe that our method can be applied to practical cases of face recognition under variations in postures and illuminations, although the PCA/FDA may not be ideal solution for huge number of face database. In fact, we can use other more advanced recognition methods instead of PCA/FDA to achieve results reasonably. We also believe that this method is applicable to other object recognitions in industrial applications where similar representations are used.

Acknowledgements The authors would like to thank the anonymous reviewers for their valuable comments. The work was supported by School of Computer Engineering, Nanyang Technological University, Singapore under a start-up research grant (CE-SUG 10/03).

References

1. Pentland A, Choudhury T (2000) Face recognition for smart environments. Computer 33(2):50–55
2. Samil A, Iyengar P (1992) Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognit 25:65–77
3. Chellappa R, Wilson CL, Sirohey S (1995) Human and machine recognition of faces: a survey. Proc IEEE 83:705–740
4. Zhao W, Chellappa R, Rosenfeld A (2000) Face recognition: a literature survey. Technical Report, CAR-TR948, University of Maryland
5. Ekman P (1992) Facial expressions of emotion: an old controversy and new findings. Philos Trans R Soc Lond 335:63–69
6. Kanade T (1973) Picture processing by computer complex and recognition of human faces. PhD dissertation, Kyoto University
7. Brunelli R, Poggio T (1993) Face recognition: features versus templates. IEEE Trans Pattern Anal Mach Intell 15:1042–1052
8. Gao Y, Leung MKH (2002) Face recognition using line edge map. IEEE Trans Pattern Anal Mach Intell 24(6):764–779
9. Sutherlan K, Renshaw D, Denyer PB (1992) Automatic face recognition. In: Proceedings of the 1st international conference on intelligent systems engineering. IEEE Press, Piscataway, pp 29–34
10. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognit Neurosci 3:71–86
11. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs Fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
12. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolution neural-network approach. IEEE Trans Neural Netw 8:98–113
13. Lin S-H, Kung S-Y, Lin L-J (1997) Face recognition/detection by probabilistic decision-based neural network. IEEE Trans Neural Netw 8:114–132
14. Er MJ, Wu S, Lu J (2002) Face recognition with radial basis function (RBF) neural networks. IEEE Trans Neural Netw 13(3):697–710
15. Pankanti S, Bolle RM, Jain A (2000) Guest editors' introduction: biometrics—the future of identification. Computer 33(2):46–49
16. Hallinan P (1994) A low-dimensional representation of human faces for arbitrary lighting conditions. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 995–999
17. Hallinan P (1995) A deformable model for face recognition under arbitrary lighting conditions. PhD thesis, Harvard University

18. Adini Y, Moses Y, Ullman S (1997) Face recognition: the problem of compensating for changes in illumination direction. IEEE Trans Pattern Anal Mach Intell 19:721–732
19. Sali E, Ullman S (1998) Recognizing novel 3-D objects under new illumination and viewing position using a small number of example views or even a single view. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 153–161
20. Cho SY, Chow TWS (2001) Enhanced 3D shape recovery using the neural-based hybrid reflectance model. Neural Comput 13:2617–2637
21. Hallinan P, Gordon G, Yuille A, Mumford D (1999) Two- and three-dimensional patterns of the face. A.K. Peters, Wellesley
22. Beymer D (1994) Face recognition under varying pose. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 756–761
23. Huang F, Zhou Z, Zhang H, Chen T (2000) Pose invariant face recognition. In: Proceedings of IEEE international conference on automatic face and gesture recognition, pp 245–250
24. Georghiades AS, Belhumeur PN, Kriegman DJ (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans On Pattern Anal Mach Intell 23(6):643–660
25. Healy G, Binford TO (1988) Local shape from specularity. Comput Vis Graph Image Process 42:62–86
26. Torrance KE, Sparrow EM (1967) Theory for off-specular reflection from roughened surfaces. J Opt Soc Am 57:1105–1114
27. Nayar SK, Ikeuchi K, Kanade T (1991) Surface reflection: physical and geometrical perspectives. IEEE Trans Pattern Anal Mach Intell 13:611–634
28. Nayar SK, Ikeuchi K, Kanade T (1990) Determining shape and reflectance of hybrid surfaces by photometric sampling. IEEE Trans Robot Autom 6:418–430
29. Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximation. Neural Netw 2:359–366
30. Park J, Sandberg IW (1991) Universal approximation using radial basis function networks. Neural Comput 3:261–257
31. Cho SY, Chow TWS (2002) Neural computation approach for developing a 3-D shape reconstruction model. IEEE Trans Neural Netw 12(5):1204–1214
32. Atick JJ, Griffin PA, Redlich AN (1996) Statistical approach to shape from shading: reconstructions of three-dimensional face surfaces from single two-dimensional images. Neural Comput 8:1321–40
33. Swets DL, Weng J (1996) Using discriminant eigenfeatures for image retrieval. IEEE Trans Pattern Anal Mach Intell 18:831–836