Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 1605–1620
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Face recognition under arbitrary illumination using illuminated exemplars☆

Sang-Woong Lee, Song-Hyang Moon, Seong-Whan Lee*

Department of Computer Science and Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-713, Republic of Korea

Received 10 December 2005; received in revised form 21 September 2006; accepted 26 September 2006

Abstract

Recently, the importance of face recognition has been increasingly emphasized since popular CCD cameras are distributed to various applications. However, facial images are dramatically changed by lighting variations, so that facial appearance changes caused serious performance degradation in face recognition. Many researchers have tried to overcome these illumination problems using diverse approaches, which have required a multiple registered images per person or the prior knowledge of lighting conditions. In this paper, we propose a new method for face recognition under arbitrary lighting conditions, given only a single registered image and training data under unknown illuminations. Our proposed method is based on the illuminated exemplars which are synthesized from photometric stereo images of training data. The linear combination of illuminated exemplars can represent the new face and the weighted coefficients of those illuminated exemplars are used as identity signature. We make experiments for verifying our approach and compare it with two traditional approaches. As a result, higher recognition rates are reported in these experiments using the illumination subset of Max-Planck Institute face database and Korean face database.

Keywords: Face recognition; Illumination invariance; Illuminated exemplar; Photometric stereo; Face synthesis

1. Introduction

Lighting is the most significant factor affecting the appearance of an object. Changes in a person's appearance induced by illumination are larger than differences in appearance of individuals. Person identification from a facial image across lighting variations is still the most challenging problem in face recognition. Fig. 1 shows the 10 illuminated images of the same individual under varying lighting conditions from Yale database [1]. In the past few years, many methods have been proposed to solve this problem with improvements in recognition. Early works in illumination-invariant face recognition focused on image representations that are mostly insensitive to changes under various lighting [2]. Various image representations are compared by measuring distances on a controlled face database. Edge map, second derivatives and 2D Gabor filters are examples of the image representations used. However, these kinds of approaches have some drawbacks. First, the different representations of image can be only extracted once they overcome some degree of illumination variations. Second, features for the person's identity are weakened whereas the illumination-invariant features are extracted.

A photometric stereo method is an approach based on the low dimensionality of the image space. The images of one object with a Lambertian surface taken from a fixed viewpoint and varying illuminations lie in a linear subspace. We can classify the new probe image by checking if it lies in the linear subspace of the registered gallery images. These gallery images are composed of at least three images of the same person under different illuminations. The main restriction in these approaches is that multiple registered images

☆A preliminary version of this paper has been presented in 2005 International Conference on Audio- and Video-Based Biometric Person Authentication which was held in New York, USA, July 20–22, 2005.

* Corresponding author. Tel.: +82 2 3290 3197; fax: +82 2 926 2168.
E-mail address: swlee@image.korea.ac.kr (S.-W. Lee).

0031-3203/$30.00 © 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.09.016

Fig. 1. Example of facial images under different lighting conditions [1].

of the same person are required. Since it recognizes the new image by checking that it is spanned in a linear subspace of the multiple gallery images, it cannot handle the new images of a different person which is not included in the gallery set.

To solve the necessity of multiple gallery images, the bilinear analysis approach is proposed [3]. It applies singular value decomposition (SVD) to a variety of vision problems including identity and lighting on a collection of objects in the same class. For bilinear analysis, photometric stereo images of different people under the same set of illuminations are always required. A method based on quotient image was introduced; this method uses the basic idea of bilinear analysis [4]. Under the assumption of a Lambertian surface without shadow, quotient image is created, deduced from the ratio between a probe image and the linear combination of photometric stereo images. The quotient image should be invariant to varying illumination conditions of the input image. This method can synthesize an image under unknown lighting condition and recognize a new face by computing its quotient image. However, the assumption of a fixed shape is not valid for real faces and the quality of the quotient image is influenced by training data. A poor quality in the quotient image results when the training data is unknown or not controlled. The generalized photometric stereo method in the bilinear analysis approach has been proposed with a rank constraint on the product of albedo and surface normal [5]. The observation matrix can be factorized by a rank constraint. The main limitation of these bilinear analysis methods is that prior knowledge of the images like the lighting direction of training data is required.

Unlike the methods described above, Blanz and Vetter use 3D morphable models of a human head [6]. The 3D model is created using a database collected by Cyberware laser scanner. Both geometry and texture are linearly spanned by the training ensemble. This approach enables them to handle illumination, pose and expression variations. But it requires the external 3D model and high computational cost. For illumination-robust face recognition, we have to solve the following problem: Given a single image of a face under the arbitrary illumination, how can the same faces under different illuminations be recognized?

In this paper, we propose a new approach for solving this problem based on illuminated exemplars. The illuminated exemplars are synthesized from photometric stereo images of each person from training data and the new probe image can be represented by a linear combination of these illuminated exemplars. The weighted coefficients are estimated in this representation and can be used as the illumination-invariant identity signature. For face recognition, our proposed method has several distinct advantages over the previously proposed methods. First, the information regarding the lighting condition of training data is not required. We can synthesize the illuminated exemplars under other illumination which is not included in the training data. Second, we can perform recognition with only one gallery image by using linear analysis of illuminated exemplars in the same class. Third, the coefficients of illuminated exemplars are used as the identity signature for face recognition across variation in lighting, which results in high recognition rates.

This paper is organized as follows. Section 2 provides the results of a literature search for past studies related to this work. In Section 3, we describe how the illuminated exemplars are synthesized, how they are analyzed and what can be used as a signature identity for face recognition, and then we propose the face recognition method using the illuminated exemplars. Section 4 gives our experimental results on Max-Planck Institute face database (MPI DB) and Korean face database (KFDB). Finally, we make a conclusion regarding our approach and suggest some areas of further research in Section 5.

2. Background

We begin with a brief review of the photometric stereo method with Lambertian lighting model and bilinear analysis of illuminated training images. We will explain what the Lambertian reflectance is and how it can be used in the photometric stereo method for face recognition [7]. We will also explain recognition methods using the bilinear analysis of the training data [3–5].

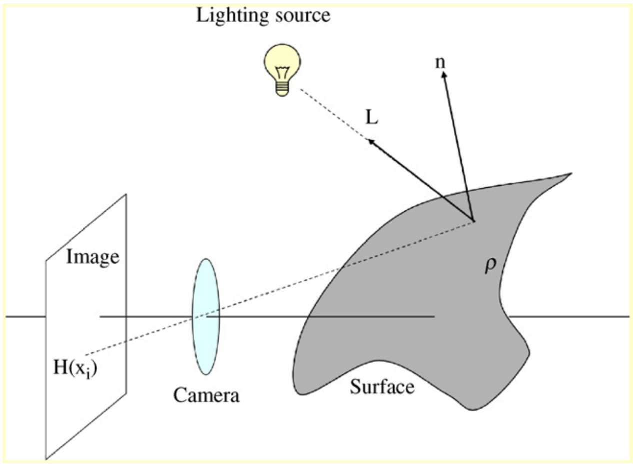

Fig. 2. Image acquisition of an object with Lambertian surface. The intensity of an image pixel depends on the surface normal and the location of the light source.

2.1. Lambertian lighting model

A Lambertian surface reflects light with an equal intensity in all directions. But this does not mean all surfaces appear equally bright. According to the Lambertian lighting model (see Fig. 2), the light reflectance properties of diffuse objects are directly related to the surface normal orientation [8].

We assume the face has the Lambertian surface. The outgoing radiance R from the surface is constant in all directions and proportional to the total irradiance E. Namely, $R=\rho E/\pi$ where ρ is the albedo of the surface. For a distant light source the total irradiance onto a surface patch will be proportional to the cosine of the angle between the surface normal and the incident light [9]. Hence, the intensity of pixel can be represented by the product of the albedo, the surface normal and a light source. The intensity of a pixel $H(x_i)$ is represented as

$$H(x_i)=[\rho(x_i)n(x_i)]\cdot[\tilde{\mathrm{l}},\hat{\mathrm{l}}]=T(x_i)^{\mathrm{T}}L(x_i), \tag{1}$$

where $L=[\tilde{\mathrm{l}},\hat{\mathrm{l}}]$ is the light source vector including the light surface vector $\hat{\mathrm{l}}$ and its intensity $\tilde{\mathrm{l}}$. The $n(x_i)$ is the surface normal vector, $\rho(x_i)$ is the albedo of the surface projecting to ith pixel x_i and $T(x_i)$ is the product of the albedo and the surface normal. The illuminated image H can represented by

$$H_{d\times 1}=\begin{pmatrix} T(x_1)\\ T(x_2)\\ \vdots\\ T(x_d)\end{pmatrix}^{\mathrm{T}} L=T_{d\times 3}{}^{\mathrm{T}}L_{3\times 1}, \tag{2}$$

where d is the number of pixels. The object-specific matrix T includes albedo and surface normal information of object. The T means the own peculiar matrix except the illumination information.

2.2. Photometric stereo

We have n images $\{H_1, H_2, \ldots, H_n\}$ of one object which has the Lambertian surface under varying lighting conditions. These images, called photometric stereo images, were observed at a fixed pose and different lighting sources. We assume these images have the same surface normal and albedo, with differences among them caused only by lighting conditions. Assuming that they are from the same object with a single viewpoint and various illuminations, the following can be expressed:

$$\mathbf{H}=\begin{pmatrix} H_1\\ H_2\\ \vdots\\ H_n\end{pmatrix}=\begin{pmatrix} T^{\mathrm{T}}L_1\\ T^{\mathrm{T}}L_2\\ \vdots\\ T^{\mathrm{T}}L_n\end{pmatrix}=T^{\mathrm{T}}\begin{pmatrix} L_1\\ L_2\\ \vdots\\ L_n\end{pmatrix}=T^{\mathrm{T}}\mathbf{L}, \tag{3}$$

where $\mathbf{H}$ is the observation matrix which is the collection of images $\{H_1, H_2, \ldots, H_n\}$ of the same object under different lighting conditions. The $\mathbf{L}=\{L_1, L_2, \ldots, L_n\}$ is the light source matrix. We rewrite Eq. (3) with dimension information as

$$\mathbf{H}_{d\times n}=T^{\mathrm{T}}_{d\times 3}\mathbf{L}_{3\times n}. \tag{4}$$

Hence, photometric stereo is rank 3 constrained. Therefore, at least three images for one object under different lighting conditions are required to determine the identity of new probe image by checking if it spanned in the linear subspace of the photometric stereo images [5]. In this approach, the training set is equivalent to the gallery set. The gallery set of one object has multiple images which is large restriction for a real face recognition system. If the new probe image belongs to a person who is not included in the gallery set, we cannot determine the identity of that person.

2.3. Bilinear models

Bilinear models offer a powerful framework for extracting the two-factor structure, identity and lighting [3–5]. For bilinear analysis, training images of each object under the same set of illuminations are required. These approaches also assume that the Lambertian surface and the image space $T^{\mathrm{T}}L$ are unknown and changed, respectively. Let $\{L_1, L_2, \ldots, L_n\}$ be a basis of linearly independent vectors, thus $L=\sum_{j=1}^{n}\beta_j L_j$ for some coefficients $\vec{\beta}=\{\beta_1, \beta_2, \ldots, \beta_n\}$. Let $\{T_1, T_2, \ldots, T_m\}$ be a basis for spanning all the possible products between albedo and surface normal of the class of objects, thus $T=\sum_{k=1}^{m}\alpha_k T_k$ for some coefficients $\vec{\alpha}=\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$. Let I be the new probe image as follows:

$$I=\rho_H N^{\mathrm{T}}L=T_H^{\mathrm{T}}L=\left(\sum_{k=1}^{m}\alpha_k T_k\right)\left(\sum_{j=1}^{n}\beta_j L_j\right). \tag{5}$$

Let $\mathbf{W} = \{\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_m\}$ be the observation matrix which is constructed from multiple images of m persons. Training images of each object $\mathbf{H}_k = \{H_{k_1}, H_{k_2}, \ldots, H_{k_n}\} = T_k\{L_1, L_2, \ldots, L_n\}$ are observed at a fixed pose and different illuminations, but each object has the same illumination set. The column of $\mathbf{H}_k$, H_{k_j}, means the image of kth object under jth illumination. We can represent the input image by linear combination of the images in $\mathbf{W}$ with the bilinear coefficients $\vec{\alpha}$ and $\vec{\beta}$. A new input image can now be represented by the following equation:

$$I = \left(\sum_{k=1}^{m} \alpha_k T_k\right)\left(\sum_{j=1}^{n} \beta_j L_j\right) = \vec{\alpha}\mathbf{W}\vec{\beta}. \tag{6}$$

The goal is now to apply the illuminated face interaction model $\mathbf{W}$ learned during training to recover appropriate face and lighting coefficients $\vec{\alpha}$ and $\vec{\beta}$ for new image which are commensurable with the face and lighting conditions learned for the training data [3]. The bilinear problem in the $m + n$ unknowns is finding $\vec{\alpha}$ and $\vec{\beta}$. If we solve these unknowns, we can generate the image space of object H from any desired lighting condition simply by keeping $\vec{\alpha}$ fixed and varying $\vec{\beta}$. But these approaches require the same set of illuminations per object, so that we have to know about the lighting condition of training data in advance. We implement this method to compare with our proposed method in Section 4.

3. Linear analysis of illuminated exemplars

We propose an illumination-invariant face recognition method based on synthesizing illuminated exemplars. We synthesize the illuminated exemplars from photometric stereo images and then we can represent a new probe image by linear combination of illuminated exemplars and the weighted coefficients offer an illumination-invariant identity signature. By computing the correlations between the set of coefficients for gallery image and that for probe image, we can determine the identity of new probe image as the gallery image which has the largest correlation. The whole procedure of the proposed face recognition algorithm is illustrated in Fig. 3.

The procedure has two phases: training and testing. Images in the database are separated into training or testing sets. In the training procedure, the training data consist of at least three illuminated images per person. However, we do not know the lighting conditions of training data. Unlike bilinear analysis methods, the training data can be constructed using different objects and different sets of illuminations. In our experiments, we construct the train matrix as m people under n different illuminated images. This is followed by computing the orthogonal basis images by the principal component analysis (PCA) for inverting the observation matrix per person. The orthogonal basis images of one person are used to synthesize the illuminated exemplars. We can then reconstruct a novel illuminated image using these basis images of the same face. In the testing procedure, the lighting condition of input image is referred to synthesize illuminated exemplars. The lighting conditions of these m synthesized exemplars and input images are same. The input image can be represented by the linear combination of illuminated exemplars and the weighted coefficients are used as those signature identities for face recognition. In the registration, those gallery images are saved. During the recognition process, we find the facial image that has the nearest set of coefficients by computing the correlation.

3.1. Synthesis of the illuminated exemplars

We assume that the face has a Lambertian surface and is illuminated by a point light, of which locations are not precisely known. However, it emits light equally in all directions. Then, the ith pixel value in an image $H(x_i)$ is given by

$$H(x_i) = \rho(x_i)n(x_i)L(x_i) = T^{\mathrm{T}}(x_i)L(x_i), \tag{7}$$

where $\rho(x_i)$ and $n(x_i)$ are the albedo and surface normal of the object's normal surface corresponding to the ith pixel x_i and L is an illumination vector including its intensity and direction of light source. An image, a collection of d pixels, H is represented as

$$H = \rho N L = T^{\mathrm{T}}L. \tag{8}$$

In the case of photometric stereo images, we have n images of the same object under different illuminations. The matrix $\mathbf{H}$ that made n images can be represented by $T^{\mathrm{T}}\mathbf{L}$ as shown in Eq. (3). The photometric stereo images are from the same object, we can assume that they have the same object-specific matrix T and different illumination vectors L. If the light source matrix $\mathbf{L}$ is non-singular ($|\mathbf{L}| \neq 0$) and $\{L_1, L_2, \ldots, L_n\}$ are linearly independent, the matrix $\mathbf{L}$ is invertible and then T can be expressed by the product of matrix $\mathbf{H}$ and the pseudo-inverse of $\mathbf{L}$ as follows:

$$T = \mathbf{H}\mathbf{L}^{+}. \tag{9}$$

The light source matrix $\mathbf{L}$ can be invertible when $\{L_1, L_2, \ldots, L_n\}$ are linearly independent of each other. To make the images independent from each other, we transform the photometric stereo images into the orthogonal basis images $\{B_1, B_2, \ldots, B_{n-1}\}$ by PCA for each person. By applying PCA to photometric stereo images, we can express a new illuminated image S of the same object by changing the coefficients $\vec{\alpha}$. The orthogonal basis images can be obtained in off-line training. The synthesized image S is determined by the coefficients multiplied to the orthogonal basis images,

$$S = \overline{B} + \sum_{j=1}^{n-1} \alpha_j B_j = \overline{B} + \vec{\alpha}B, \tag{10}$$

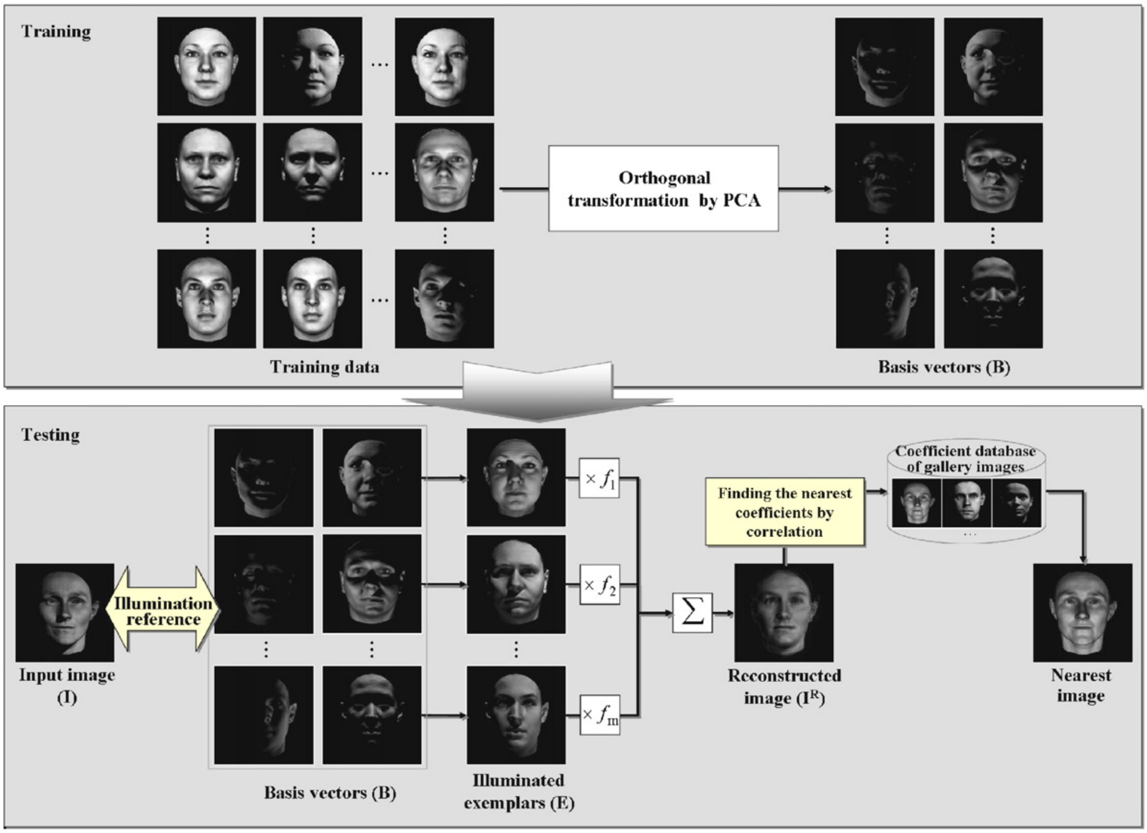

Fig. 3. Overview of the proposed face recognition procedure.

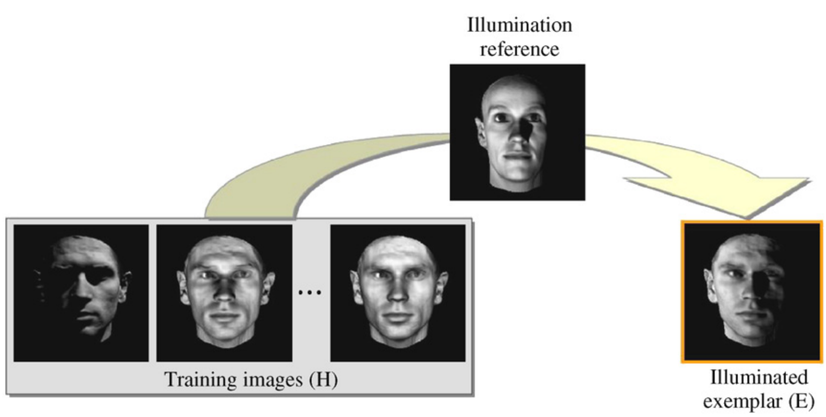

Fig. 4. Basic idea for synthesis of illuminated exemplar using reference image.

where $\overline{B}$ represents the mean of orthogonal basis images per object and $\vec{\alpha} \in \Re^{n-1}$. We propose that the input image is used as a reference to synthesize the new illuminated image S.

Fig. 4 shows the basic idea for synthesizing a new illuminated image, called '*Illuminated exemplar*', from training images under different lighting conditions. '*Illumination reference*' is the image which is referred by photometric stereo image per person to synthesize the illuminated exemplar. From training images, we can extract the basis vectors which are orthogonal to each other and then we can represent a new image using those basis and illumination reference. Illumination of new synthesized image becomes similar to that of the reference. Since photometric stereo images have the same surface (only different lighting condition) and we refer the lighting condition of input image, we can synthesize the image under the similar lighting condition of input. We assume that the input image can be represented using the set of coefficients $\vec{\alpha}$ from orthogonal basis images as follows:

$$I = \overline{B} + \vec{\alpha}\mathbf{B}. \tag{11}$$

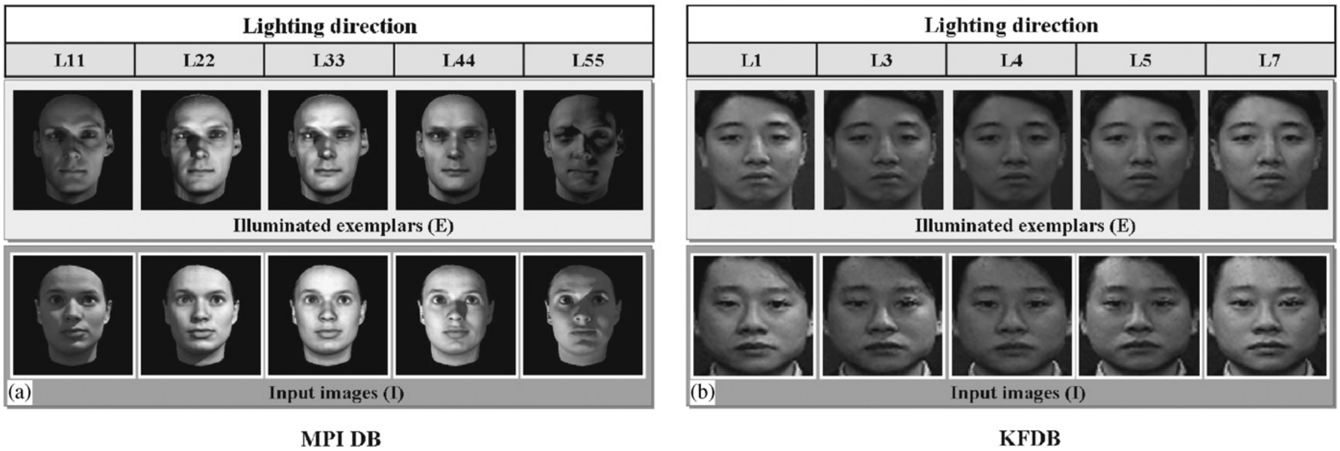

Fig. 5. Example of the illuminated exemplars. (a) MPI DB and (b) KFDB.

In this case, $\vec{\alpha}$ makes the synthesized image to have similar lighting condition as input image. The columns of matrix are orthogonal to each other, the transpose is the inverse and we can now easily find the optimal set of coefficients $\vec{\alpha}^*$ by transpose instead inverse. We can find the optimal set of coefficients $\vec{\alpha}^*$ as follows:

$$\vec{\alpha}^* = \mathbf{B}^{-1}(I - \overline{B}) = \mathbf{B}^{\mathrm{T}}(I - \overline{B}). \tag{12}$$

Using the optimal set of coefficients, we synthesize the illuminated exemplars E with the same lighting condition as input image. The synthesized exemplar E is represented as

$$E = \overline{B} + \sum_{j=1}^{n-1} \alpha_j^* B_j = \overline{B} + \vec{\alpha}^* B. \tag{13}$$

According to both the photometric stereo and bilinear analysis approaches, we need at least three images per object to reconstruct a new illuminated image. In the photometric stereo images, we choose three images of random lighting directions $\{\tilde{H}_1, \tilde{H}_2, \tilde{H}_3\}$ and we transform those images into the orthogonal coordinate system, which is defined by eigenvectors $\{\tilde{B}_1, \tilde{B}_2\}$ using PCA, where $\overline{\tilde{B}}$ is the mean of $\{\tilde{B}_1, \tilde{B}_2\}$ and $\vec{\tilde{\alpha}}^* = \{\tilde{\alpha}_1, \tilde{\alpha}_2\}$ is the coefficient for synthesizing the illuminated exemplar $\tilde{E}$. An illuminated exemplar using three images is represented as follows:

$$\tilde{E} = \overline{\tilde{B}} + \sum_{j=1}^{2} \alpha_j^* B_j = \overline{\tilde{B}} + \vec{\tilde{\alpha}}^* \tilde{B}. \tag{14}$$

Fig. 5 shows example of the synthesized exemplars from the training data of MPI DB and KFDB. We choose three images under random illumination of each person and those chosen images for each person are different sets. The upper row shows examples of the illuminated exemplars using the images from the top row. The lower row shows examples of the different illuminated input images. Each illuminated exemplar (upper row) is synthesized by referring the lighting condition of input image found directly below it. As shown, the illuminated exemplars have very similar lighting conditions to that of the input image. The illuminated exemplar is synthesized from three different illuminated images of each person. One illuminated exemplar is synthesized per object, so there are m illuminated exemplar images under the same lighting condition of the input image where the training data are collected by the images of m objects.

3.2. Linear combination of illuminated exemplars

In the previous section, we described that how the illuminated exemplar is synthesized. Using both the photometric stereo images and input image, m illuminated exemplars are synthesized per person. The illuminated exemplar of kth person $\tilde{E}_k$ can be represented as

$$\tilde{E}_k = \overline{\tilde{B}}_k + \sum_{j=1}^{2} \tilde{\alpha}_{k_j}^* B_{k_j} = \overline{\tilde{B}}_k + \vec{\tilde{\alpha}}_k^* \tilde{B}_k, \tag{15}$$

where $\overline{\tilde{B}}_k$ is the mean of orthogonal basis images $\{\tilde{B}_{k_1}, \tilde{B}_{k_2}\}$ from three photometric stereo images $\{\tilde{H}_{k_1}, \tilde{H}_{k_2}, \tilde{H}_{k_3}\}$. The column of $\tilde{\mathbf{H}}_k$, $\tilde{H}_{k_j}$, is the image under jth illumination of kth person. After synthesizing m illuminated exemplars from the training data, we analyze the input image using them. The input image is represented well by the linear combination of the illuminated exemplars. At this time, the weighted coefficients are estimated from the illuminated exemplars under the same illumination as the input image. That means, the coefficients depend on the m illuminated exemplars but not on the lighting conditions. Because the illuminated exemplars are for the object class only, the coefficients provide a signature identity that is invariant to illumination. The set of coefficients $\vec{f}$ is computed by the following equation:

$$I = \sum_{k=1}^{m} f_k \tilde{E}_k = \vec{f} \tilde{E}, \tag{16}$$

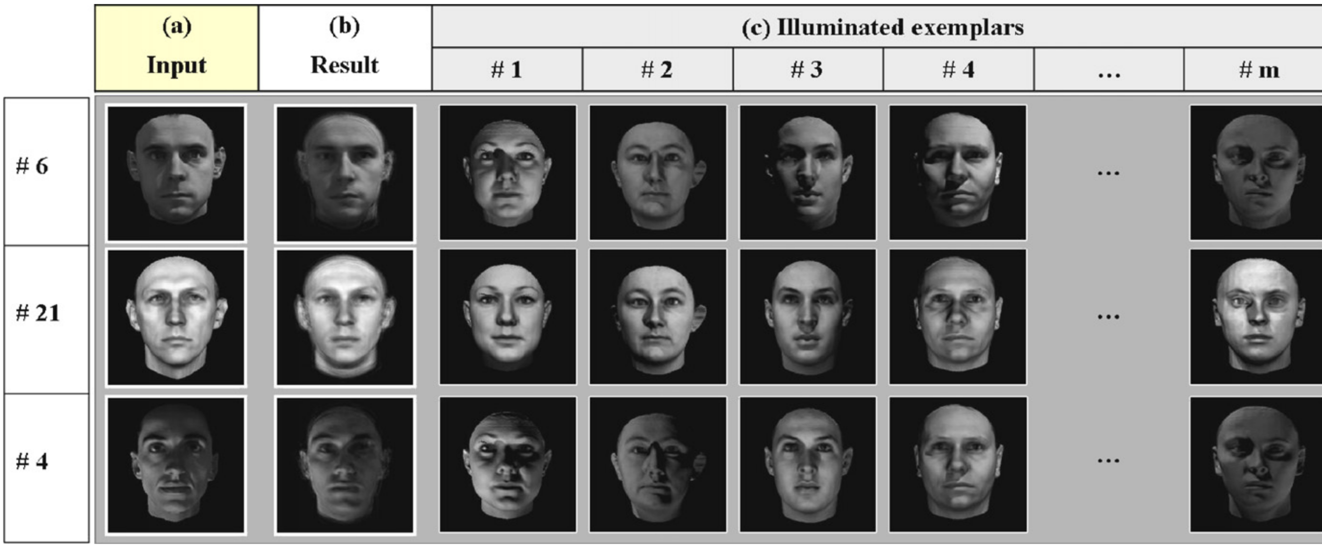

Fig. 6. Reconstructed examples with input images from MPI DB using m illuminated exemplars.

where $\vec{f}=\{f_1, f_2, \ldots, f_m\}$ is the set of coefficients from the m illuminated exemplars and used for face recognition. The f_k is the coefficient for the illuminated exemplar kth object. The $\tilde{\mathbf{E}} = \{\tilde{E}_1, \tilde{E}_2, \ldots, \tilde{E}_m\}$ is the matrix of the illuminated exemplars which are synthesized from the previous section. The problem is to choose $\vec{f}$ so as to minimize the cost function $C(\vec{f})$. We define the cost function as the sum of square errors which measures the difference between the input image and the linear sum of the illuminated exemplars. We can find the optimal set of coefficients which minimizes the cost function,

$$\vec{f}^* = \arg\min_{\vec{f}} C(\vec{f}), \tag{17}$$

with the cost function

$$C(\vec{f}) = \sum_{i=1}^{d} \left(I(x_i) - \sum_{k=1}^{m} f_k \tilde{E}_k(x_i) \right)^2. \tag{18}$$

The solution to Eq. (17) is obtained by the least square minimization method. We assume that the number of illuminated exemplars is m and the number of pixels is d. We represent the kth illuminated exemplar, $\tilde{E} = \{\tilde{E}(x_1), \tilde{E}(x_2), \ldots, \tilde{E}(x_d)\}$ where $\tilde{E}(x_i)$ is the intensity of a pixel, x_i. To represent the input image H using illuminated exemplars, we have to find $\vec{f}$ by the following equation:

$$\begin{pmatrix} I(x_1) \\ \vdots \\ I(x_d) \end{pmatrix} = \begin{pmatrix} \tilde{E}_1(x_1) & \ldots & \tilde{E}_m(x_1) \\ \vdots & \vdots & \vdots \\ \tilde{E}_1(x_d) & \ldots & \tilde{E}_m(x_d) \end{pmatrix} \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}. \tag{19}$$

This can be rewritten with dimensional information as follows:

$$I_{d\times 1} = \tilde{\mathbf{E}}_{d\times m} \vec{f}_{m\times 1}, \tag{20}$$

where

$$\tilde{\mathbf{E}} = \begin{pmatrix} \tilde{E}_1(x_1) & \ldots & \tilde{E}_m(x_1) \\ \vdots & \vdots & \vdots \\ \tilde{E}_1(x_d) & \ldots & \tilde{E}_m(x_d) \end{pmatrix}, \tag{21}$$

$$\vec{f} = (f_1, \ldots, f_m)^{\mathrm{T}},$$

$$I = (I(x_1), \ldots, I(x_d))^{\mathrm{T}}.$$

The least square solution satisfies $\tilde{\mathbf{E}}^{\mathrm{T}} I = \tilde{\mathbf{E}}^{\mathrm{T}} \tilde{\mathbf{E}} \vec{f}$. If the columns of $\tilde{\mathbf{E}}$ are linearly independent, then $\tilde{\mathbf{E}}^{\mathrm{T}}\tilde{\mathbf{E}}$ is non-singular and has an inverse. Then the optimal set of coefficients is

$$\vec{f}^* = \left(\tilde{\mathbf{E}}^{\mathrm{T}}\tilde{\mathbf{E}}\right)^{-1} \tilde{\mathbf{E}}^{\mathrm{T}} I. \tag{22}$$

We can express the input image I using the computed $\vec{f}^*$ on the assumption that the columns of matrix $\tilde{\mathbf{E}}$ are linearly independent. If they are not independent, the solution $\vec{f}^*$ will not be unique. In this case, the solution can be solved by the pseudo-inverse of $\tilde{\mathbf{E}}$, $\tilde{\mathbf{E}}^{+}$. But, that is unlikely to happen for proposed method. The reconstructed image I^R of the input image is represented as follows:

$$I^R = \sum_{k=1}^{m} f_k^* \tilde{E}_k = \vec{f}^* \tilde{\mathbf{E}}. \tag{23}$$

By using Eq. (23), we can get the optimal set of coefficients to represent the input image. To verify the coefficients as the signature identity, we reconstruct the input image using the computed optimal set of coefficients. Figs. 6 and 7 show the examples of reconstructed images of MPI DB and KFDB using the optimal set of coefficients $\vec{f}^*$. In these figures, (a) shows the input image under the arbitrary lighting condition, (b) shows the reconstructed images using the linear combination of illuminated exemplars and (c) is the

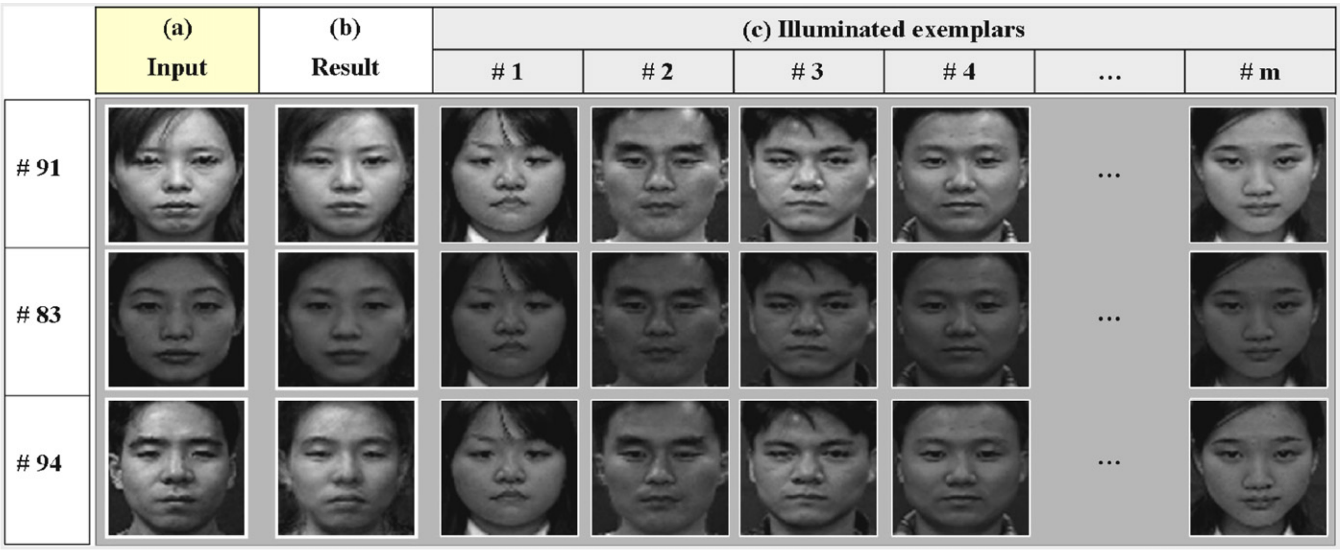

Fig. 7. Reconstructed examples with input images from KFDB using m illuminated exemplars.

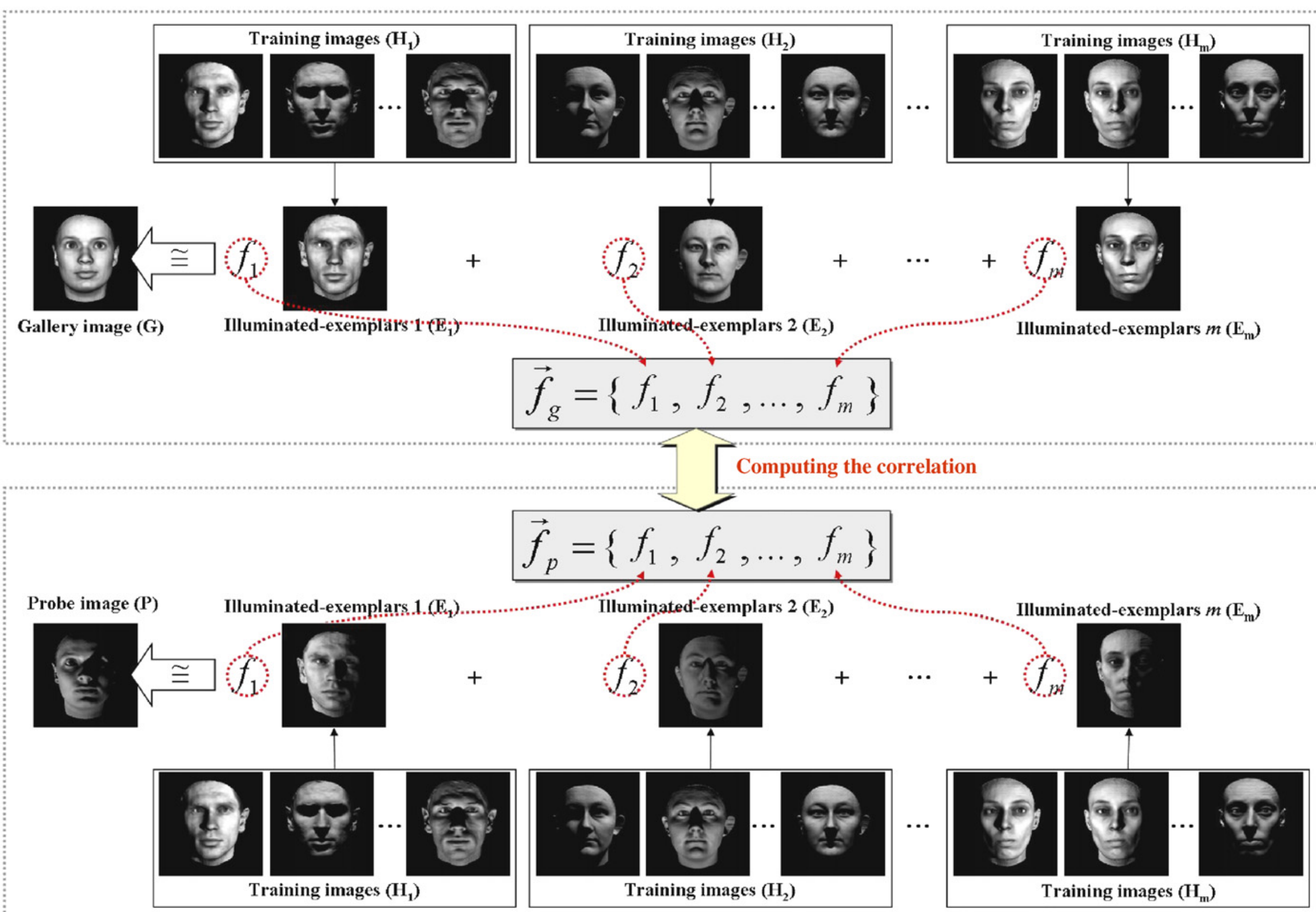

Fig. 8. The recognition principle for one gallery and probe image under arbitrary illumination.

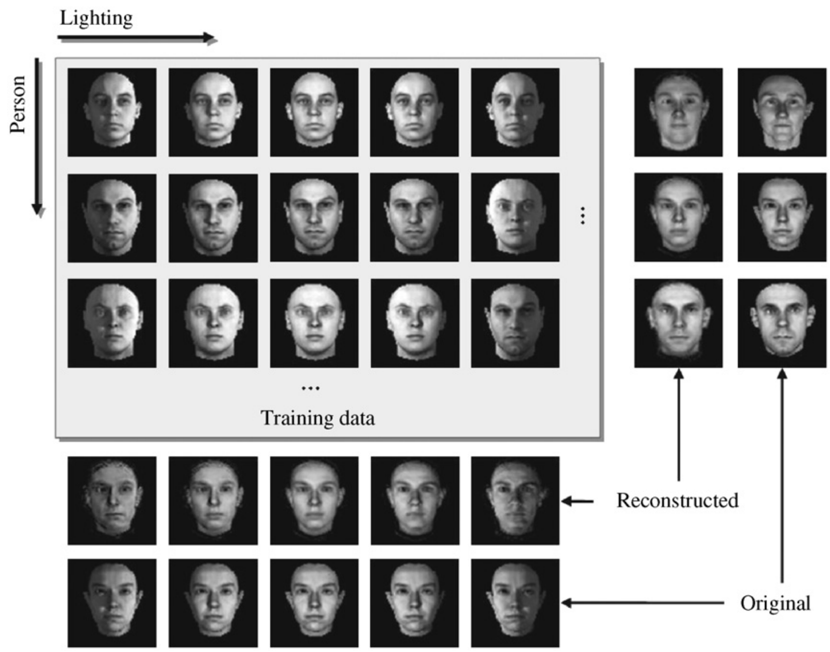

Fig. 9. Results of reconstructed images by bilinear analysis method.

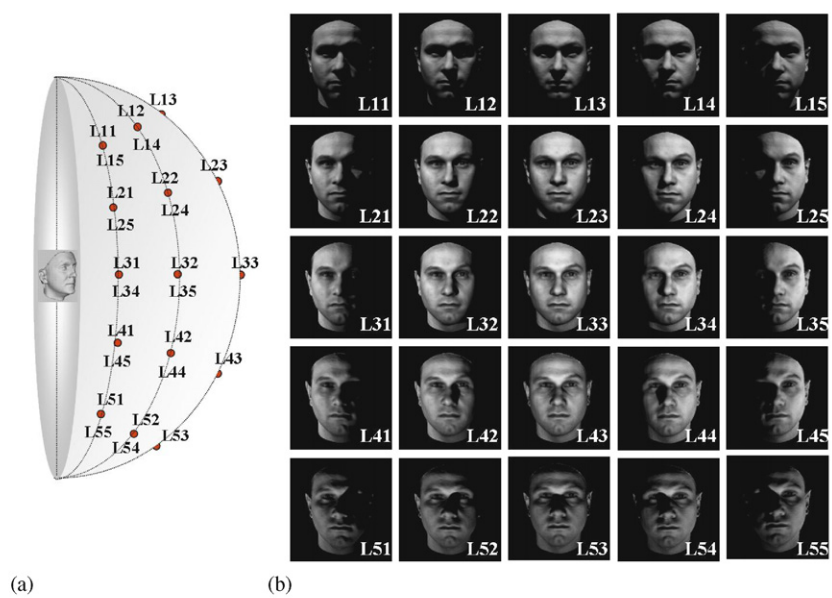

Fig. 10. MPI face database. (a) The position configuration of lights and (b) example of data set.

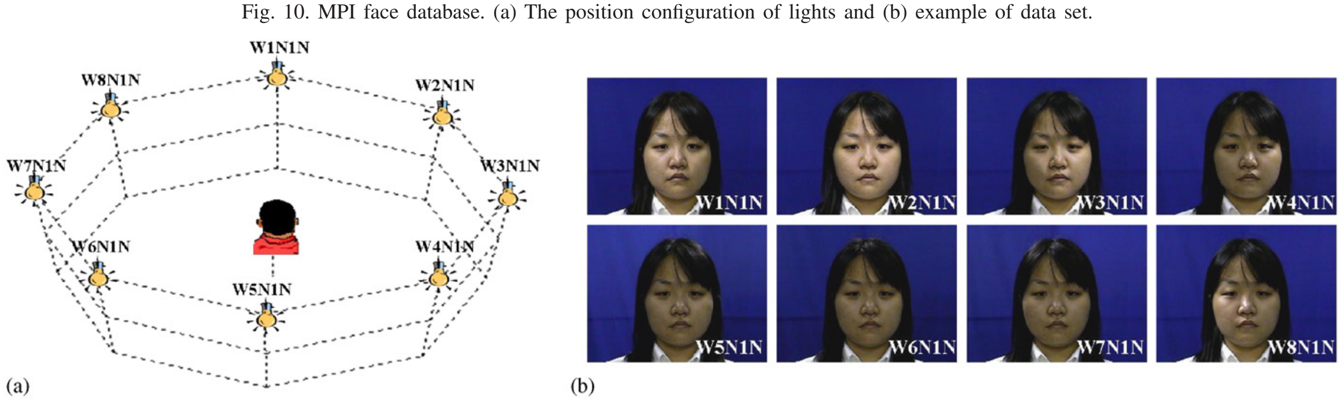

Fig. 11. Korean face database. (a) The position configuration of lights and (b) example of data set.

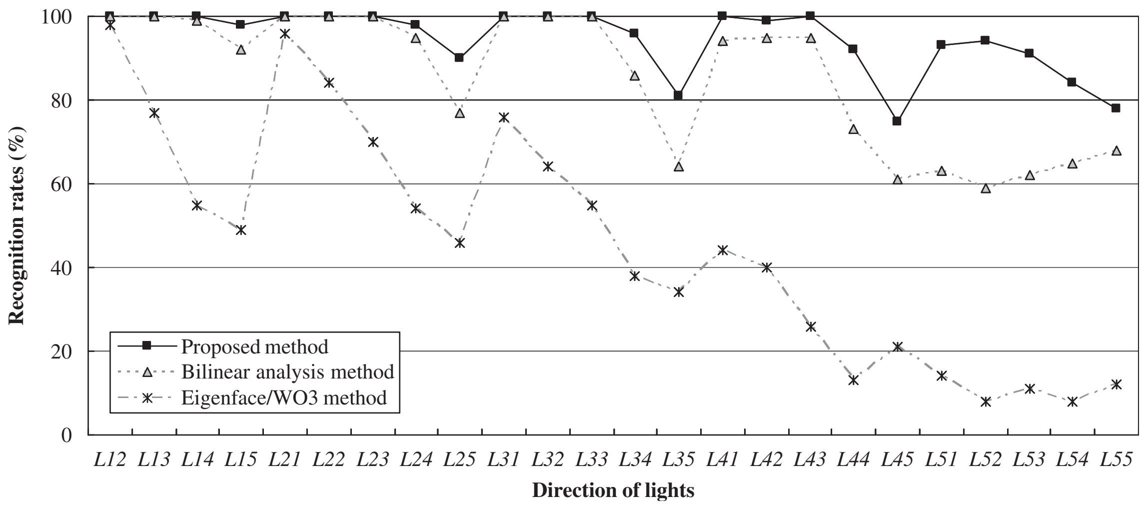

Fig. 12. Recognition rates for all the probe sets with a fixed gallery set.

illuminated exemplars which are synthesized using the input image to refer lighting condition.

3.3. Recognition

In this section, we describe what kind of signature is used for recognizing the face. We recognize the new face by the coefficients of the illuminated exemplars which are synthesized using photometric stereo images and input image. In the previous section, we had shown the computing procedure of the coefficients and the reconstruction results using those coefficients.

Fig. 8 shows the basic idea of the face recognition principle for one gallery image and probe image under arbitrary illumination. When the gallery or probe image is taken, we synthesize the illuminated exemplars from the photometric stereo images with each gallery or probe image. We analyze the input image, gallery and probe image by the linear combination of those synthesized exemplars. We can then get the set of coefficients which represent the gallery image and probe image, these two sets of coefficients are used as the signatures for face recognition. As shown Fig. 8, we perform the recognition by calculating the normalized correlations between coefficients for the probe image and the registered gallery images. We compute the normalized correlation by finding the facial image with the nearest correlation coefficient vector and recognized as that person.

Suppose that a gallery image G has its signature $\vec{f}_{\mathrm{g}}^{*}$ and a probe image P has its signature $\vec{f}_{\mathrm{p}}^{*}$. Then the optimal sets of coefficients for each image are

$$\vec{f}_{\mathrm{g}}^{*} = (\tilde{\mathbf{E}}_{\mathrm{g}}^{\mathrm{T}}\tilde{\mathbf{E}}_{\mathrm{g}})^{-1}\tilde{\mathbf{E}}_{\mathrm{g}}^{\mathrm{T}}G, \quad \vec{f}_{p}^{*} = (\tilde{\mathbf{E}}_{\mathrm{p}}^{\mathrm{T}}\tilde{\mathbf{E}}_{\mathrm{p}})^{-1}\tilde{\mathbf{E}}_{\mathrm{p}}^{\mathrm{T}}P, \tag{24}$$

where $\tilde{\mathbf{E}}_{\mathrm{g}}$ and $\tilde{\mathbf{E}}_{\mathrm{p}}$ are the matrices of illuminated exemplars using G and P as illumination reference of lighting condition. The normalized correlation between a gallery and probe image is

$$norr\_corr(G, P) = \frac{Cov(\vec{f}_{\mathrm{g}}^{*}, \vec{f}_{\mathrm{p}}^{*})}{sd(\vec{f}_{\mathrm{g}}^{*})\, sd(\vec{f}_{\mathrm{p}}^{*})}, \tag{25}$$

where $sd(a)$ is the standard deviation of a and $Cov(a, b)$ means the covariance of a and b.

4. Experiments

We have conducted a number of experiments with our approach using the MPI DB [6] and KFDB [10]. We performed experiments for face recognition using the illumination subset of the whole database. In these experiments, we compared the proposed method with eigenface/WO3 [11,12] method and bilinear analysis method [5].

Eigenface/WO3: *Eigenface* is a method commonly used in face recognition. We have implemented the eigenface approach by training the eigenvector from the same training data. To solve the illumination problem, this method is applied without three principal components, the most influential factor in degradation of performance. This means the three most significant components are discarded on the assumption that the first few principal components capture only variation in lighting.

Bilinear analysis: We also implemented the bilinear analysis method for comparison which is mentioned in Section 2. Fig. 9 shows the results of rendering the novel face under several illuminations and rendering the 3 out of 100 faces under the new illumination. We worked with the MPI DB of face images under well-controlled variation in lighting. For training, we took the face database of 100 faces under varying lighting conditions. We fit the bilinear model to this training data using the iterated SVD procedure. For testing, the single image per person which is not included in database is used. We use these results to compare with

Table 1
Recognition rates comparison

G/P	$L11$	$L22$	$L33$	$L44$	$L55$	Avg.
$L11$	—	100/100	99/100	92/73	81/68	94.4/88.2
$L22$	100/100	—	100/100	100/79	86/51	97.2/86.0
$L33$	99/100	100/100	—	100/100	99/100	99.6/100
$L44$	85/77	99/87	100/100	—	100/100	96.8/92.8
$L55$	63/43	79/47	95/97	100/100	—	87.4/77.4
Avg.	89.4/84.0	95.6/86.8	98.8/99.4	98.4/90.4	93.2/71.0	95.1/88.9

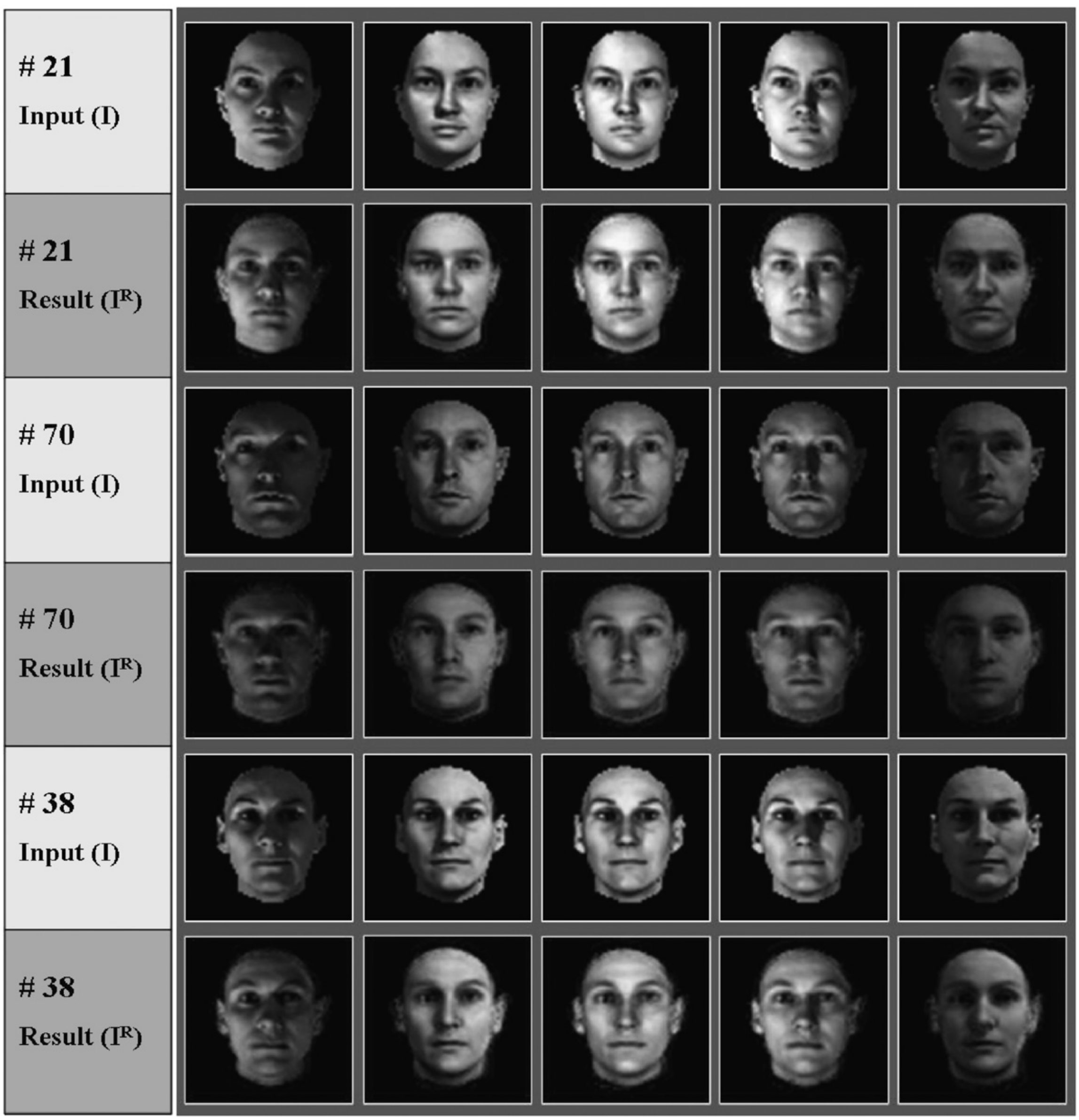

Fig. 13. Example of the reconstructed input images from MPI DB (64×64). The upper row shows the original input images (I) and lower row shows the reconstructed image (I^{R}) of each set (two rows). Each column represents the same lighting conditions.

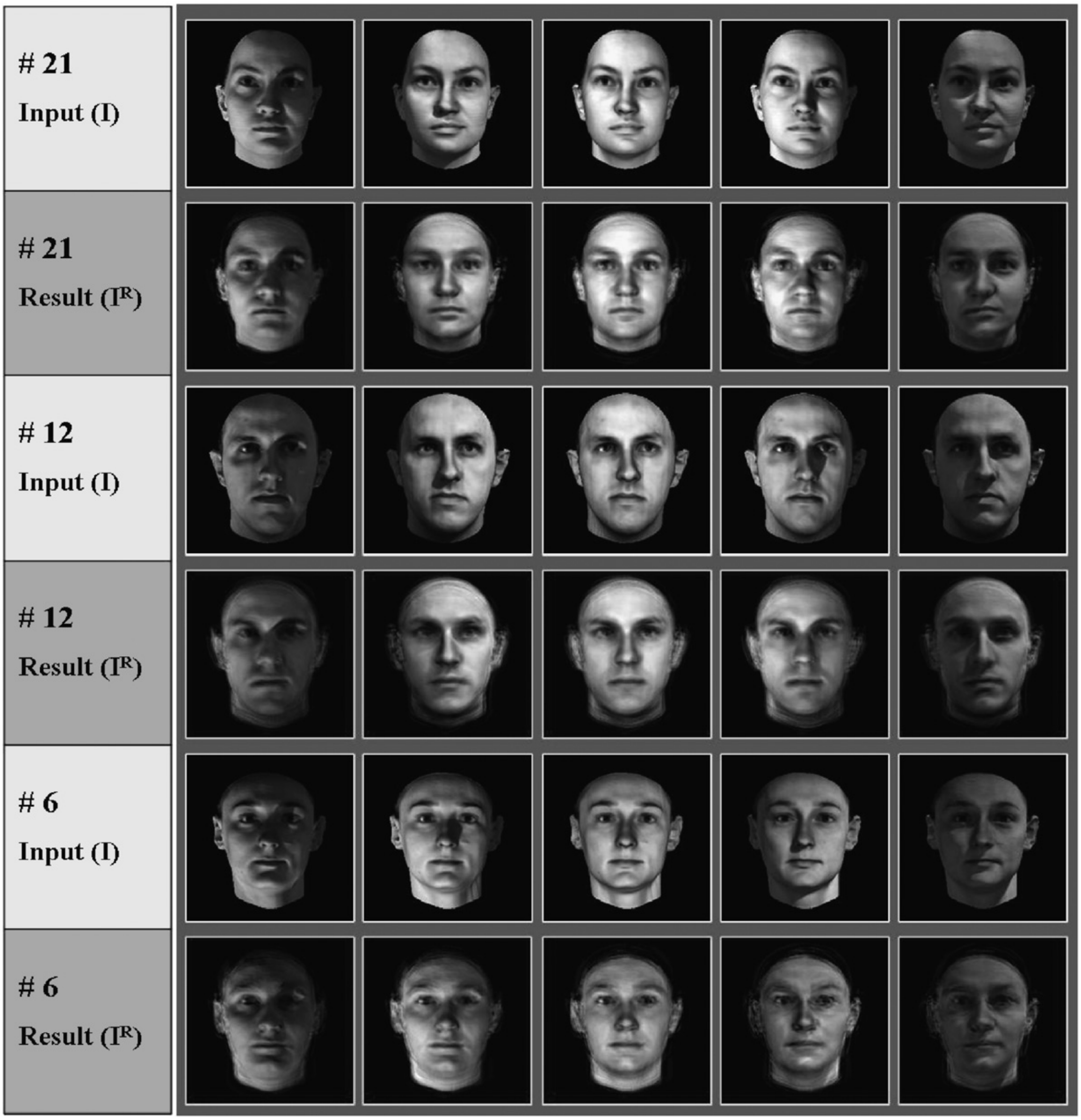

Fig. 14. Example of the reconstructed input images from MPI DB (256×256). The upper row shows the original input images (I) and lower row shows the reconstructed image (I^{R}) of each set (two rows). Each column represents the same lighting conditions.

our proposed method for face recognition (see Section 4.1.2) and face reconstruction (see Section 4.2) using same data set.

4.1. Face database

4.1.1. Max-Planck Institute face database (MPI DB)

The MPI DB is used to demonstrate our proposed approach. We use 200 2D images of Caucasian faces that were rendered from a database of 3D head models recorded with a laser scanner (*Cyberware*^{TM}) [13,14]. The images were rendered from a viewpoint 120 cm in front of each face with ambient light only. For training, we use the images of 100 people. The images of the other 100 people were used for testing our algorithm. Images having 25 changes $\{L11, L12, \ldots, L45, L55\}$ in illumination directions are used (see Fig. 10). The images had been collected for psychophysical experiments from males and females between 20 and 40 years of age. No glasses, earrings or beards were presented and all head hair was removed from the images. The resolution of the images is 256×256 pixels and the color images were converted to 8-bit gray level images. Fig. 10 shows the position lights and example of data set. In Fig. 10, (a) shows the position configuration of the lights and (b) shows example images of lighting condition changes.

4.1.2. Korean face database (KFDB)

We also used KFDB to demonstrate the effectiveness of our proposed approach. Images of 200 people were selected

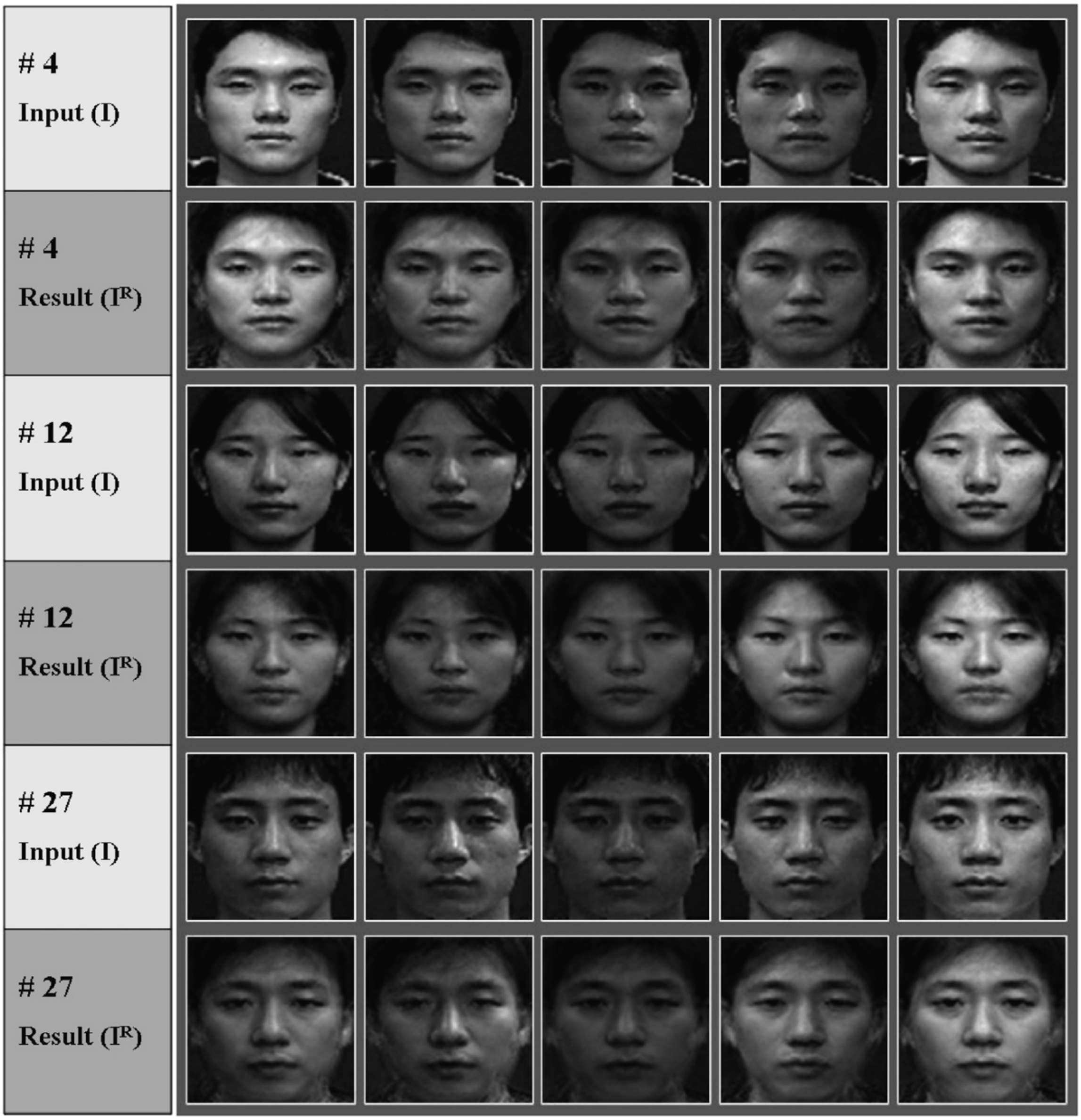

Fig. 15. Example of the reconstructed input images from KFDB (64×64). The upper row shows the original input images (I) and lower row shows the reconstructed image (I^{R}) of each set (two rows). Each column represents the same lighting conditions.

from 1000. For training, we use the images of 100 people. The images of the other 100 people were used for testing our algorithm. We also selected frontal facial images without glasses having eight changes $\{W1N1N, W2N1N, \ldots, W8N1N\}$ in their illumination directions from a total of 52 images per person. The images are 640×480 in resolution. For the experiments, we convert each color images to an 8-bit gray level image and resize to 64×64 pixels. Fig. 11 shows the position of lights, along with the entire configuration of the studio platform and example of data set. In Fig. 11, (a) shows the position configuration of the lights and (b) shows example images of lighting condition changes.

4.2. Recognition results and analysis

In this section, we describe the recognition experiments and their performances. First, we performed experiments for face recognition using the set of coefficients from synthesized illuminated exemplars. This was then followed with experiments for reconstruction of input images for verifying the set of coefficients used as the signature identity, in next section. We present the recognition results when the images of training and testing sets are taken from the same database, MPI DB. We have conducted two experiments by changing the lighting directions of the gallery and probe set.

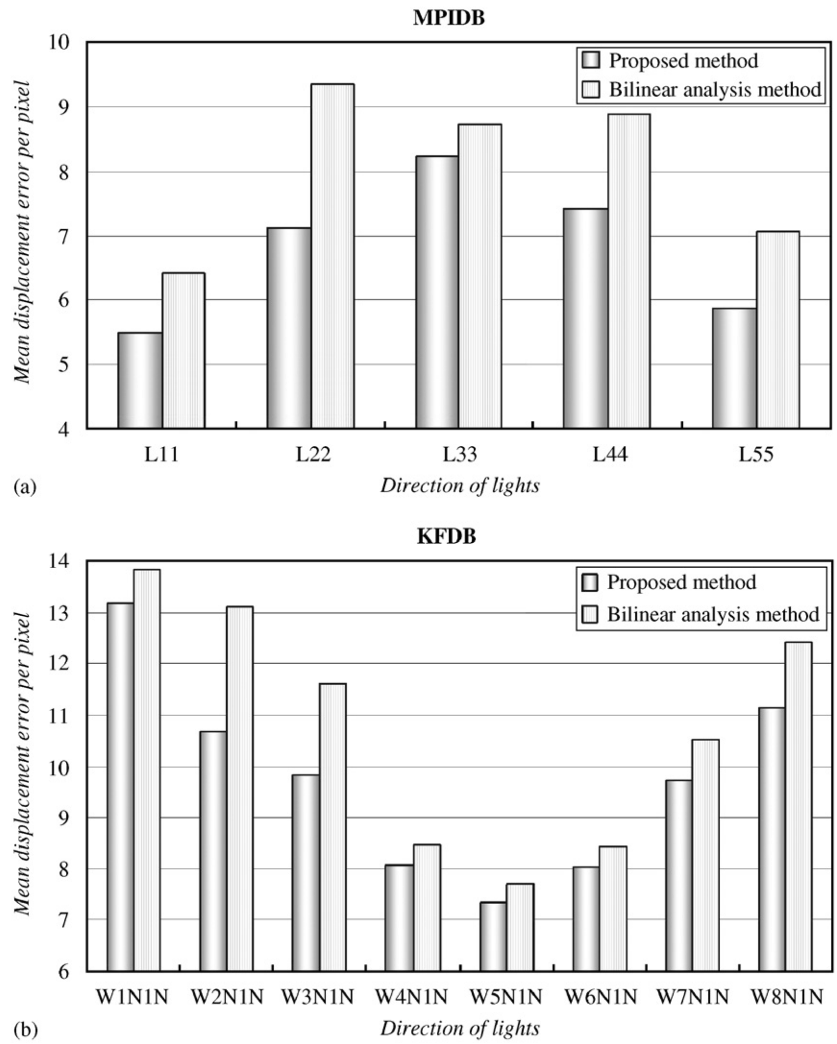

Fig. 16. Comparisons of mean reconstruction errors. (a) MPI DB and (b) KFDB.

4.2.1. Experiment in the gallery set of fixed direction and probe set of all lighting directions

Graph in Fig. 12 shows the recognition rates for the fixed gallery set of lighting conditions (100 images) with the probe sets of varying lighting conditions (100 images under each illumination). We use the gallery set under the first lighting condition, $L11$ and the probe sets under the other 24 lighting conditions from $L12$ to $L55$ in the testing set. In this experiment, we obtain good recognition results although the illumination changes are rapid. As shown in Fig. 12, when the distance between the light sources of the gallery and probe sets is small, the recognition performance is high, conversely when the distance between the two is large, the recognition results are of lower quality, especially when using the eigenface/WO3 and the bilinear methods. The bilinear analysis method allows higher recognition rates than the eigenface/WO3 method, though neither method results in as high a performance as our proposed method.

4.2.2. Experiment in the gallery set and probe set of varying lighting directions

The next experiment is designed for the gallery and probe sets both under different lighting conditions. Table 1 represents the comparison results between our approach and bilinear analysis approach for the gallery and probe sets under the different directions of lighting $\{L11, L22, L33, L44, L55\}$. The P means the probe sets and G means the gallery sets. The number on the right is for bilinear analysis method and the number on the left is for our proposed method. The average rates obtained by bilinear analysis are 88.9%, while our approach outperforms it at an average of 95.1%. When the distance between the light sources of the gallery and probe sets is small, the recognition performance of both methods is high. Nevertheless, in cases where facial images are under difficult illumination condition such as $L11$ and $L55$, we cannot avoid performance degradation but our proposed method is much less.

As shown Fig. 12 and Table 1, our approach outperforms the both approaches. The eigenface/WO3 method uses the principal components of training data but not the invariant of illumination changes and the bilinear analysis method takes a large number of iterations to solve concurrent lighting and identity problem. If we take one of the iterations in the bilinear analysis, the quality of the results drops rapidly, conversely if a large number of iterations are performed to convergence, this produces an improvement in the results at the cost of CPU time. We compared the recognition rate of our approach with one of the bilinear analysis methods after three iterations, which has similar time as ours. As a result, we can recognize the new face under the arbitrary illumination given by just one registered image and training data under also arbitrary illuminations without recursive iterations.

4.3. Reconstruction results and analysis

We use the coefficients of illuminated exemplars as the identity signature for face recognition. To verify the set of coefficients as the signature identity, we reconstruct the gallery or probe image. We have reconstruction experiments using MPI DB and KFDB. Figs. 13 and 14 show examples of the reconstructed input images (gallery or probe sets) using the coefficients of illuminated exemplars from MPI DB. Each set (two rows) shows image set of one person. The upper row shows the original input images and the other row shows the reconstructed images. Each column represents the images under lighting changes. The original images were color images, set to the size of 256×256 pixels. They were converted to an 8-bit gray level and resized to 64×64 pixels. Experiments were performed on both the sets of original images and resized images. First set (top two rows) in Figs. 13 and 14 is the same person and the other sets are different. The performance degradation of recognition and reconstruction with high-resolution images (256×256) were not much as compared with that of low-resolution images (64×64).

Fig. 15 shows examples of the reconstructed input images (gallery or probe sets) using the coefficients of illuminated exemplars from KFDB. The original images were color images, set to the size of 640×480 pixels. In our experiments, all images were cropped to include only the face with little hair and background. They were converted to an 8-bit gray level and resized to 64×64 pixels. We align the face on the images by fixing two eye locations to observe only variation of lighting (scale and rotate). Each set (two rows) shows image set of one person. The upper row shows the original input images and the other row shows the reconstructed images. Each column represents the images under lighting changes.

Fig. 16 represents the reconstruction errors in input image per pixel. We use the resized image (64×64). The horizontal axes represent the directions of illumination while the vertical axes represent the mean displacement error in input image per pixel. Fig. 16(a) shows reconstruction error rates of images obtained from MPI DB and Fig. 16(b) shows reconstruction error rates of images obtained from KFDB. As shown, the reconstruction errors of the proposed method are smaller than bilinear analysis method.

5. Conclusions and further research

We have addressed a new approach for illumination-invariant face recognition. The idea here is to synthesize illuminated exemplars using photometric stereo images and apply them to represent the new input image under the arbitrary illumination. This method only requires one input image and one registered image per person for recognition. The weighted coefficients are used as the signature identity, so that a new image can be represented as a linear combination of a small number of illuminated exemplars. Since we synthesize the illuminated exemplars under the same lighting conditions for the input image using any three or more images under different illuminations, we have no need for illumination information of the training data. We require three images at least per person, which need not be collected under the same conditions as the data set of other persons. In our experiments, we choose images under the random direction of lighting to synthesize illuminated exemplars. This allowed the photometric stereo images to have more lighting information than using images under the fixed directions of lighting. Experimental results on various face images have shown a good performance when compared with the previous approaches. Our approach also shows a stable recognition performance even under the large illumination changes. However, we incur a high computational complexity due to the synthesizing of illuminated exemplars for all input images. The cast shadow and the case of multiple light sources are not handled in this paper. This will be investigated in the future. Furthermore, it can become particularly difficult when illumination is coupled with pose variation. Because there are the extreme lighting changes which are caused by pose variation, we are also trying to treat not only lighting changes but also pose changes.

Acknowledgments

This research was supported by 2005 Seoul R&BD Program. And we would like to thank the Max-Planck Institute for providing the MPI DB.

References

[1] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination Cone Models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[2] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 721–732.
[3] W.T. Freeman, J.B. Tenenbaum, Learning bilinear models for two-factor problems in vision, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, June 1997, pp. 554–560.
[4] A. Shashua, T.R. Raviv, The quotient image: class based re-rendering and recognition with varying illuminations, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 129–139.
[5] K. Zhou, R. Chellappa, Rank constrained recognition under unknown illuminations, in: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Nice, France, October 2003, pp. 11–18.
[6] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.
[7] R. Basri, D. Jacobs, Photometric stereo with general, unknown lighting, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, USA, December 2001, pp. 374–381.
[8] B.K.P. Horn, Understanding image intensities, Artif. Intell. 8 (2) (1977) 1–31.
[9] P. Belhumeur, D. Kriegman, What is the set of an object under all possible illumination conditions?, Int. J. Comput. Vision 28 (3) (1998) 245–260.
[10] B.-W. Hwang, M.-C. Roh, S.-W. Lee, S.-W. Lee, Performance evaluation of Face Recognition Algorithms on Asian face Database, KFDB, in: Proceedings of the Fourth International Conference on Audio- and Video-Biometric Person Authentication, Lecture Notes in Computer Science, Guildford, UK, June 2003, pp. 557–565.
[11] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognit. Neurosci. 3 (1991) 72–86.
[12] L. Sirovitch, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Am. 4 (3) (1987) 519–524.
[13] T. Vetter, N.E. Troje, Separation of texture and shape in images of faces for image coding and synthesis, J. Opt. Soc. Am. 14 (9) (1997) 2152–2161.
[14] V. Blanz, S. Romdhani, T. Vetter, Face identification across different poses and illuminations with a 3D Morphable Model, in: Proceedings of the Fifth International Conference on Automatic Face and Gesture Recognition, Washington DC., USA, May 2002, pp. 202–207.

About the Author—SANG-WOONG LEE received his BS degree in Electronics and Computer Engineering from Korea University, Seoul, Republic of Korea, in 1996 and his MS and PhD degrees in Computer Science and Engineering from Korea University, Seoul, Republic of Korea, in 2001 and 2006, respectively. Currently, he is a visiting researcher in Carnegie Mellon University. His research interests include computer vision, robotics, pattern recognition and related applications.

About the Author—SONG-HYANG MOON received her BS degree in Computer Science from Sookmyung Women's University, Seoul, Republic of Korea, in 2003 and her MS degree in Computer Science and Engineering from Korea University, Seoul, Republic of Korea, in 2005. Currently, she is a researcher in LG electronics, Inc. Her research interests include face recognition, computer vision and visual communication.

About the Author—SEONG WHAN LEE received his BS degree in Computer Science and Statistics from Seoul National University, Seoul, Republic of Korea, in 1984, and his MS and PhD degrees in Computer Science from KAIST in 1986 and 1989, respectively. From February 1989 to February 1995, he was an assistant professor in the Department of Computer Science at Chungbuk National University, Cheongju, Republic of Korea. In March 1995, he joined the faculty of the Department of Computer Science and Engineering at Korea University, Seoul, Republic of Korea, as an associate professor, and he is now a full professor. He was the winner of the Annual Best Paper Award of the Korea Information Science Society in 1986. He obtained the First Outstanding Young Researcher Award at the Second International Conference on Document Analysis and Recognition in 1993, and the First Distinguished Research Professor Award from Chungbuk National University in 1994. He also obtained the Outstanding Research Award from the Korea Information Science Society in 1996. He has been the founding coeditor-in-chief of the International Journal on Document Analysis and Recognition and the associate editor of the Pattern Recognition Journal, the International Journal of Pattern Recognition and Artificial Intelligence and the International Journal of Computer Processing of Oriental Languages since 1997. He was the program cochair of the Sixth International Workshop on Frontiers in Handwriting Recognition, the Second International Conference on Multimodal Interface, the 17th International Conference on the Computer Processing of Oriental Languages, the Fifth International Conference on Document Analysis and Recognition, and the Seventh International Conference on Neural Information Processing. He was the workshop cochair of the Third International Workshop on Document Analysis Systems and the First IEEE International Workshop on Biologically Motivated Computer Vision. He served on the program committees of several well-known international conferences. He is a fellow of IAPR, a senior member of the IEEE Computer Society and a life member of the Korea Information Science Society. His research interests include pattern recognition, computer vision and neural networks. He has published more than 200 publications in these areas in international journals and conference proceedings, and has authored five books.