International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 22, No. 3 (2008) 389–410

NONITERATIVE 3D FACE RECONSTRUCTION BASED ON PHOTOMETRIC STEREO

SANG-WOONG LEE

Department of Computer Science and Engineering
Chosun University, 375 Seosuk-dong
Dong-gu, Gwangju 501-759, Korea
swlee@chosun.ac.kr

PATRICK S. P. WANG

College of Computer and Information Science
Northeastern University
360 Huntington Avenue, 260 WVH
Boston, MA 02115, USA
pwang@ccs.neu.edu

SVETLANA N. YANUSHKEVICH

Department of Electrical and Computer Engineering
Schulich School of Engineering University of Calgary
2500 University Drive NW Calgary
Alberta T2N 1N4 Canada
syanshk@ucalgary.ca

SEONG-WHAN LEE*

Division of Computer and Communication Engineering
Korea University, Anam-dong
Seongbuk-ku, Seoul 136-713, Korea
swlee@image.korea.ac.kr

3D face reconstruction is a popular area within the computer vision domain. 3D face reconstruction should ideally be achieved easily and cost-effectively, without requiring specialized equipment to estimate 3D shapes. As a result of this, many techniques for retrieving 3D shapes from 2D images have been proposed. In this paper, a novel method for 3D face reconstruction based on photometric stereo, which estimates the surface normal from shading information in multiple images, hence recovering the 3D shape of a face, is proposed. In order to overcome the problems of previous approaches related to prior-knowledge regarding lighting conditions and iterative algorithms, the exemplar is synthesized with known lighting conditions from at least three images, under arbitrary lighting conditions and using an illumination reference. Experiments in 3D face reconstruction were made by verifying the proposed approach using the illumination subset of

*Author for correspondence

the Max-Planck Institute face database and Yale face database B. Experimental results demonstrate that the proposed method is effective for 3D shape reconstruction of faces from 2D images.

Keywords: 3D face reconstruction; illuminated exemplar; photometric stereo; face synthesis.

1. Introduction

Even though face recognition systems are becoming more popular, their low performance prevents face technologies from being applied in real life. Most researchers have tried to improve performance using newly-developed pattern recognition approaches such as Principle Component Analysis and Support Vector Machines,[11] which have resulted in better systems. However, there still remain several problems which decrease the recognition ratio drastically. They are commonly caused by smiling, aging, facial accessories, and various external conditions such as illumination and pose changes.[10]

Among these problems, the two most critical factors are pose and illumination. Therefore, it follows logically that the optimum solution to this problem is to acquire full 3D information regarding the face as represented, for example, using a 3D shape-mesh and a 2D texture-map. While 2D view-based[8] and other appearance-based approaches have significant merits, they suffer from a fundamentally limited representation (a collection of 2D appearance subspaces). 3D data can be used to produce invariant measures out of 2D data. It is believed that an intrinsic 3D model is the only way to properly tackle the complications that arise due to pose, illumination and perhaps expression.

In this paper, an impressive 3D reconstruction of human faces is presented. A high-resolution 3D model of the shape space was obtained by employing laser scanning of a large face database *a priori*. In this database, face images of a number of people were collected in a set of different orientations, in order to sample the orientation space. Then, a new face given in one of these orientations was synthesized in any orientation, as a linear combination of faces in the desired orientation. This idea was subsequently extended by Blanz[4] to incorporate a 3D face model, enabling the synthesized face to be more person-specific and realistic. This capability can be used directly for face recognition. Beymer and Poggio[3] reported results along the same lines for pose-invariant face recognition. The results varied depending on the available number of real views of a person to be recognized. Without view synthesis, this ranged from 66.7% when only one real view was available to 98.7% when 15 real views were available. However, these approaches require a sensor, which acquires depth information, usually referred to as depth or range camera, thus limiting applicability. Therefore, many solutions that recover a 3D shape from 2D image sequences have been proposed.

The methods to recover shape are called shape-from-X techniques, where X can be shading, stereo, motion, texture, multi-views, and so on. The traditional approach is of course that of Structure-from-Motion (SFM).[9] This deals with the problem of recovering 3D points on a rigid object from 2D correspondences of points across images. SFM is a direct method of obtaining 3D points, which are often not very dense or accurate; thus a post-processing phase is required. Shape from shading (SFS) deals with the recovery of shape from a gradual variation of shading in the image. To solve the SFS problem, it is important to study how images are formed. A simple model of image formation is the Lambertian model, in which the gray level at a pixel in the image depends on the light source direction and the surface normal. In SFS, given a gray level image, the aim is to recover the light source and the surface shape at each pixel in the image. However, real images do not always follow the Lambertian model. If the Lambertian reflectance and known light source direction is assumed, and if the brightness can be described as a function of the surface shape and light source direction, the problem is still not simple. This is because if the surface shape is described in terms of the surface normal, a linear equation with three unknowns is produced, and if the surface shape is described in terms of the surface gradient, a nonlinear equation with two unknowns is produced. Therefore, finding a unique solution to SFS is difficult. SFS algorithms solve the surface gradients using several different methods, because the mapping is not invertible locally using a single measurement of image intensity. A class of algorithms uses two additional constraints on the surface and minimizes the energy function. Other algorithms use propagation of shape information from reference points to iteratively solve the surface gradients, local surface assumptions or linear approximations, in order to obtain the reflectance map. A single image has natural limitations in revealing 3D information about the objects, given the reduction to one dimension. Therefore, as an extension of SFS, the photometric stereo method (PSM) is proposed. The PSM is a shape from shading algorithm using several images to invert the reflectance map, and has been an ongoing research problem in the computer vision community. The basic algorithm for PSM estimates the surface gradients locally for each pixel without using global constraints. Among the more advanced methods are those that use local confidence measures to account for surface interreflections and shadowing.

In this paper, a novel method for 3D face reconstruction based on photometric stereo is proposed: a reconstruction method for estimating the surface normal and hence the shape of a face from the shading information in multiple images. The illuminated exemplars are synthesized with known lighting conditions from input images. The illuminated images of one person (reference person) are used as an illumination reference for synthesizing exemplars. The surface normal for 3D depth values can be estimated as a noniterative 3D face reconstruction algorithm, by using synthesized exemplars.

This paper is organized as follows. In the next section, a brief review of related literature is introduced. In Sec. 3, the process of synthesizing images using illuminated exemplar is described and the method of recovering the surface normal and 3D depth map is explained. Section 4 presents the experimental results from the

Yale face database B[6] and MPI face database.[7] Finally, the conclusion regarding the proposed approach is presented, and areas of future work are suggested.

2. Background

2.1. *Shape from shading*

The basic idea of SFS is to infer the 3D surface of an object from shading information in an image. In order to infer such information, a reflectance model under which the given image is generated from a 3D shape is assumed. There are many illumination models available. Among these models, the Lambertian model represents the most popular model, and has been used extensively in the computer vision community for the SFS problem. The key equation in the SFS problem follows the irradiance equation,[7]

$$I(x,y) = R(p(x,y), q(x,y)) \tag{1}$$

where $I(x,y)$ is the image, R is the reflectance map and $p(x,y)$ and $q(x,y)$ are the shape gradients (partial derivatives of the depth map). With the assumption of a Lambertian surface and a single distant light source, the equation can be written as follows,

$$I = \rho \cos\theta = \rho N^T L \tag{2}$$

or

$$I = \frac{1 + pP_s + qQ_s}{\sqrt{1+p^2+q^2}\sqrt{1+P_s+Q_s}}, \tag{3}$$

where θ is the angle between the outward normal to the surface, $N = (p, q, 1)$ and the negative illumination vector, $-L = (P_s, Q_s, 1)$ represents the direction opposite to the distant light source, and ρ is the albedo. Since the SFS algorithm provides face shape information, illumination and pose problems can be solved simultaneously. For example, the illumination problem can be solved by rendering the prototype image from a given input image, I_p. In order to evaluate existing SFS algorithms, Zhao[12] applies several SFS algorithms to (1) synthetic face images generated based on the Lambertian model and constant albedo, and (2) real face images. The experiment results demonstrate that these algorithms are not sufficient for real face images, such that a significant improvement in face recognition can be achieved. The reason is that the face is composed of materials with different reflecting properties: cheek skin, lip skin, eye, and so on. Hence, Lambertian model and constant albedo cannot provide sufficient approximation. Once the light source direction is known, the 3D shape can be estimated. Most SFS algorithms assume that the light source direction is known, high computational complexity and the re-rendering image result in poor quality, because of reconstruction from a singular image. The nature of SFS makes it an ill-posed problem in general. In other words, the reconstructed 3D surface cannot synthesize images under different lighting angles.

2.2. *Photometric stereo*

The term Photometric Stereo Method (PSM) or just Photometric Stereo (PS) refers to the extension of shape from shading class methods which use three or more images for shading, based on 3D shape recovery. SFS methods are extended to PSM if several irradiances are known for every image point and corresponding surface point. A reduction in the necessary assumption can be expected, because of the larger amount of data and improvement in reconstruction results. Photometric stereo methods firstly recover surface orientations and can be combined with an integration method to calculate a height or depth map. Assuming the face has a Lambertian surface, the illuminated image can represented by

$$H = \rho N^T L = T^T L, \tag{4}$$

where N is the surface normal and ρ is the albedo, a material dependant coefficient. The object-specific matrix T includes the albedo and surface normal information of the object. A collection of images, $\{H_1, H_2, \ldots, H_n\}$ of the same object under varying illumination is created. These images, called photometric stereo images, are observed at a fixed pose and different lighting sources. Therefore, more than one irradiance value can be assigned to a projected surface point without encountering a correspondence problem. Each image corresponds to one light source. Assuming that these light sources are from the same object with a single viewpoint and various illuminations, these images can be expressed as

$$\mathbf{H} = \begin{Bmatrix} H_1 \\ H_2 \\ \vdots \\ H_n \end{Bmatrix} = \begin{Bmatrix} T^T L_1 \\ T^T L_2 \\ \vdots \\ T^T L_n \end{Bmatrix} = T^T \mathbf{L}, \tag{5}$$

where $\mathbf{H}$, the collection of images $\{H_1, H_2, \ldots, H_n\}$ of same object under different lighting conditions, is the observation matrix, and $\mathbf{L} = \{L_1, L_2, \ldots, L_n\}$ is the light source matrix. If lighting parameters are known, the surface normal orientation for objects can be extracted. In order to solve Eq. (5), a least squares cost function is defined,

$$C(T, L) = \sum_{i=1}^{d} \sum_{k=1}^{n} (H_k(x_i) - T_k(x_i) L_k)^2, \tag{6}$$

where d is the number of pixels. The object-specific matrix, T includes the albedo and surface normal information of an object. It is possible to minimize this cost function to solve T and L up to a constant linear transform using Singular Value Decomposition (SVD),[1] $H = SVD(H) = UDV^T$ and retain the top three components as $H = U_{d\times3} D_{3\times3} V^T_{3\times3}$. Then T and L can be recovered up to an invertible matrix A with $T = T^* A$ and $L = A^{-1} L^*$, where $L^* = V$, $T^* = UD$, $\rho = \|T\|$ and then the surface normal, N can be represented as $N = T/\|T\|$. In enforcing integrability, it is guaranteed that the vector field of surface normal is a gradient field that corresponds to a surface. The surface of a face is denoted by $z(x, y)$, the surface normal can be rewritten as

$$N = (p, q, 1), \tag{7}$$

and

$$p = -\frac{n_x}{n_z} = z_x, \quad q = -\frac{n_y}{n_z} = z_y, \tag{8}$$

where $N = (n_x, n_y, n_z)$, the $z_x = \frac{\partial z(x,y)}{\partial x}$ and $z_y = \frac{\partial z(x,y)}{\partial y}$ are the partial gradients of $z(x, y)$. A photometric stereo is used to acquire surface gradient information and suggests the use of features derived from the gradient space. A common assumption to ensure unambiguous surface reconstruction from a single image (i.e. SFS) or several images (i.e. PSM) are to ensure that the Lambertian surface, single point light source and the directions of illumination are known.

2.3. *Illumination cone*

In earlier work, it is demonstrated that the images under an arbitrary combination of light sources form a convex cone in an image space. This cone, called the illumination cone, can be constructed from as few as three images. The reconstructed 3D face surface and illumination cones can be combined together to synthesize images under different illumination and pose. Georghiades *et al.*[6] use prior knowledge of the shape of the face to resolve Generalized Bas-Relief (GBR)[2] ambiguity. Once GBR parameters are calculated, it is a simple matter of rendering synthetic images under different illumination and poses. For reconstruction of the 3D face surface, this method includes the use of boundary conditions for the reflectance map, a very frequent iteration between the surface model and the reflectance map and the use of a multigrid scheme during optimization. The illumination cone method demonstrates an improvement in recognition performance, under the variation of pose and illumination. A problem is that it requires many images per person to construct the subspace. For commercial applications, the use of many images per person is not feasible due to the cost consideration. The other issue is the computational cost when the database contains a large number of images. If it took too much time to recognize a face when the database is large, it is discouraged from further development and the approach is an iterative solution technique, employed for finding the surface normal and 3D surface caused by high computational cost.

3. Noniterative 3D Face Reconstruction

In this paper, a method of 3D face reconstruction based on photometric stereo, representing a face reconstruction method with facial images under various illuminations, is proposed. Unlike previous photometric stereo, the 3D shape of faces can be recovered from images under unknown illuminations. One additional process is possible; synthesizing the illuminated exemplar to known lighting conditions from

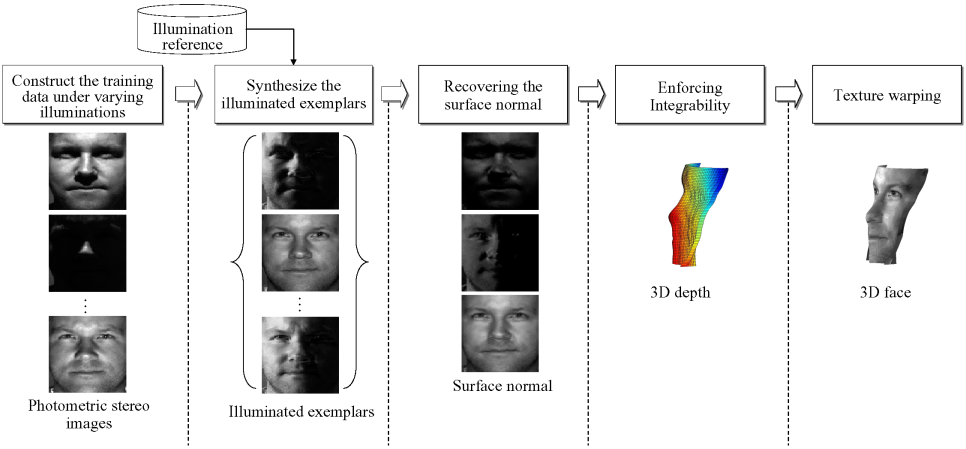


Fig. 1. Overview of the proposed face reconstruction procedure.

unknowns. The entire procedure of the proposed face reconstruction method and example of each process are illustrated in Fig. 1. The proposed method is based on photometric stereo; however the illuminated exemplar is synthesized in order to solve the problem of prior-knowledge of lighting conditions in the photometric stereo method. First, the illuminated exemplars are synthesized under known lighting conditions, from input images under arbitrary lighting conditions. The input is constructed from at least three facial images under the same viewpoint and same person, but different lighting conditions (these images are called photometric stereo images). If the light source of photometric stereo images is known, then the surface normal can be estimated using the least squares solution for the observation and light source matrix. The depth of face and 3D surface can be reconstructed using the 2D texture image, by enforcing integrability to the surface normal. In the previous photometric stereo method, knowledge of the lighting conditions like direction and intensity of lighting source is required for face reconstruction. This constraint is solved using illumination reference, by synthesizing an illuminated exemplar under known illuminations from input images. The surface normal can be extracted from this synthesized exemplar, and the 3D face can be reconstructed by integration of the surface normal.

3.1. *Synthesis exemplar*

It is assumed that the face has a Lambertian surface and a point light source, with locations not precisely known, emitting light equally in all directions from a single point. The ith pixel value in an image, $H(x_i)$ is given by

$$H(x_i) = \rho(x_i)n(x_i)L(x_i), \tag{9}$$

where $\rho(x_i)$ is the albedo and $n(x_i)$ is the surface normal of the object corresponding to the ith pixel, x_i and L is an illumination vector including its intensity and direction of the light source. An image, a collection of d pixels, H is represented in Eq. (4). In the case of photometric stereo images, n images of the same object under different illuminations are retrieved. The matrix H that made n images can be represented by $T^T L$. Since the photometric stereo images are from the same object, it can be assumed that they have the same object-specific matrix, T and a different illumination vector, L. If the light source matrix, L is nonsingular ($\|L\| \neq 0$) and $\{L_1, L_2, \ldots, L_n\}$ are linearly independent, the matrix L is invertible and T can be expressed by the product of matrix H and the pseudo-inverse of L, L^+ as follows

$$T = \mathbf{HL}^+. \tag{10}$$

The light source matrix, L can be invertible when $\{L_1, L_2, \ldots, L_n\}$ are linearly independent of each other. To make the images independent of each other, the photometric stereo images are transformed into the orthogonal basis images, $\{B_1, B_2, \ldots, B_{n-1}\}$, by Principal Component Analysis (PCA) for each person. When applying PCA to photometric stereo images, a new illuminated image of the same object can be expressed using orthogonal basis images, $\mathbf{B}$, and the proper coefficients, $\boldsymbol{\alpha}$. By changing $\boldsymbol{\alpha}$, many images across lighting variations can be synthesized. The synthesized image S is determined by the coefficients multiplied by the orthogonal basis images,

$$S = \bar{B} + \sum_{j=1}^{n-1} \alpha_j B_j = \bar{B} + \boldsymbol{\alpha}\,\mathbf{B} \tag{11}$$

where $\bar{B}$ represents the mean of orthogonal basis images per object and $\boldsymbol{\alpha} \in \Re^{n-1}$. In this paper, it is proposed that illuminated variation images of one person be used as an illumination reference to synthesize a new illuminated image. The lighting conditions of these illumination reference images are already known, therefore the lighting condition of the synthesized exemplar is identical to reference images.

Figure 2 presents the basic idea for synthesizing a new illuminated image, called "***exemplar***," from input images under different lighting conditions. "***Illumination reference***" is the image under known lighting conditions of the reference person. The set of images under lighting variations of only one person is required. From input images, the basis vectors, which are orthogonal to each other, are extracted, and then a new image is represented using these basis vectors and illumination references. Illumination of a new synthesized image is similar to that of the reference image. Since input images have the same surface (only different lighting conditions) and the lighting condition of illumination reference images (already known lighting condition) are referred to, the image can be synthesized under similar lighting conditions. It is assumed that the illumination reference can be represented using a set of coefficients, $\boldsymbol{\alpha}$, from orthogonal basis images as follows

$$R = \bar{B} + \boldsymbol{\alpha}\mathbf{B}. \tag{12}$$

In this case, $\boldsymbol{\alpha}$ is a set of the coefficients for the synthesized image, having similar lighting conditions of the illumination reference. The columns of the matrix

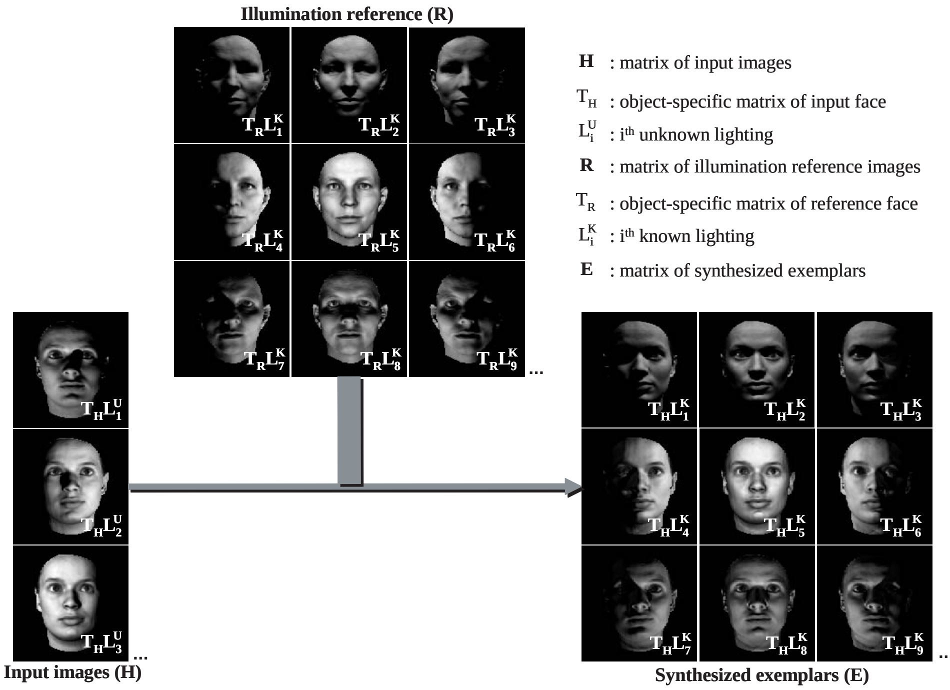


Fig. 2. Basic idea for synthesis of exemplars using illumination reference.

are orthogonal to each other, the transpose is the inverse and the optimal set of coefficients, $\boldsymbol{\alpha}^*$ can now easily be found by transposition, instead of inversion. The optimal set of coefficients, $\boldsymbol{\alpha}^*$, can be found as follows

$$\boldsymbol{\alpha}^* = \mathbf{B}^{-1}(R - \bar{B}) = \mathbf{B}^{\mathrm{T}}(R - \bar{B}). \tag{13}$$

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

$$E = \bar{B} + \sum_{j=1}^{n-1} \alpha_j^* B_j = \bar{B} + \boldsymbol{\alpha}^* \mathbf{B}. \tag{14}$$

In the photometric stereo images of the input face, eight images of random lighting directions are used, $\{\tilde{H}_1, \tilde{H}_2, \ldots, \tilde{H}_8\}$ and these images are transformed into the orthogonal coordinate system by PCA with eigenvectors, $\{\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_7\}$. $\bar{\tilde{B}}$ is the mean of $\{\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_7\}$ and $\tilde{\boldsymbol{\alpha}}^* = \{\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots \tilde{\alpha}_7\}$ are the coefficients for synthesizing the illuminated exemplar $\tilde{E}$, an illuminated exemplar using three images is as follows

$$\tilde{E} = \bar{\tilde{B}} + \sum_{j=1}^{n-1} \tilde{\alpha}_j^* B_j = \bar{\tilde{B}} + \boldsymbol{\alpha}^* \mathbf{B}. \tag{15}$$

Since the illuminated-exemplar is synthesized from unknowns to known illumination, an iterative algorithm for finding a lighting source vector L or factorization of T and L is not required.

Figure 3 presents examples of the synthesized exemplars from photometric stereo images of the input. Eight images are chosen under random illumination of each person and the selected images for each person are in a different set. The top row presents examples of different illuminated images of the same person (input). The middle row presents examples of the synthesized exemplars using the images from the top and bottom rows, presenting examples of the different illumination references. Each synthesized exemplar image (middle row) references the illumination of the reference image found directly below it. As shown, the synthesized exemplars have very similar lighting conditions to those of the reference images. The lighting conditions of the illumination references are known, and then the lighting condition of synthesized exemplars can be used. If the lighting conditions are known, the surface normal of face can be estimated.

3.2. *Noniterative recovery of surface normals*

It is assumed that n images are possessed for each face with equal poses and different light sources. Under this assumption, the face albedo and surface normal can be modeled with linear ambiguity. In fact, this ambiguity can be discarded in the face model, just as in the formation of the illumination cone,[6] by an iterative algorithm. However, if this ambiguity is not solved, the light direction of input images must be extracted once for each face model, leading to a high computation expense in recognition. The light direction of the input images is estimated by synthesizing illuminated exemplars under known lighting conditions using illumination reference images. The proposed approach is a simple noniterative method

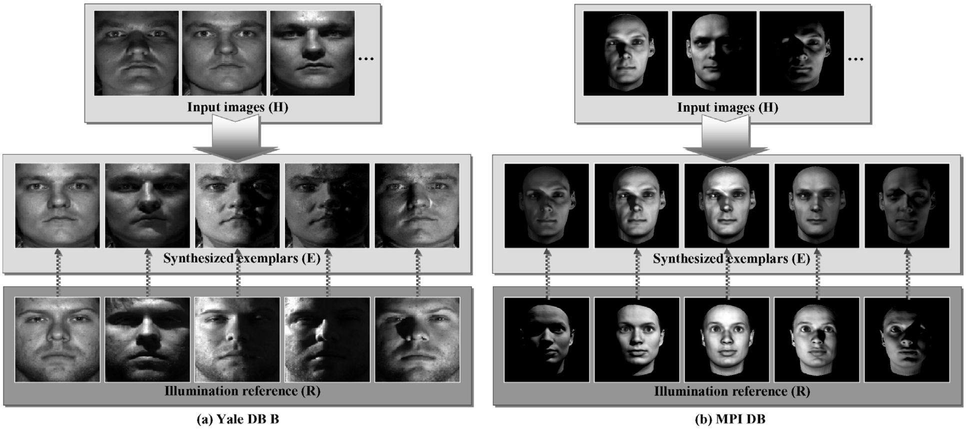


Fig. 3. Example of the synthesized exemplars.

of estimating surface normal for 3D face reconstruction, using synthesized exemplars. The well-known photometric stereo approaches are able to rapidly obtain dense local 3D surface orientations from intensity images of the face illuminated by calibrated light sources; however, uncalibrated images can be obtained under arbitrary illumination in the proposed approach. Let the synthesized exemplar be $\tilde{E}$,

$$\tilde{E} = T^T L. \tag{16}$$

If the light source L is known, T to the surface normal can be calculated for reconstruction of the depth map (3D surface). The exemplars are synthesized using the illumination reference, the light directions of synthesized exemplars can be assumed to be the same as directions of reference images. The light source vector that includes information regarding the light direction, is represented as L_R, and can be expressed as $\tilde{E} = T^T L_R$. The problem lies in selecting T, to minimize the cost function, $C(T)$. The cost function is defined as the sum of square errors, which measures the difference between the synthesized exemplar image and the product of T and L_R. The optimal object-specific matrix, T, which minimizes the cost function, $C(T)$, is the following,

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

while the cost function is

$$C(T) = \sum_{i=1}^{d} (\tilde{E}(x_i) - T^T(x_i) L_R)^2. \tag{18}$$

The solution to Eq. (17) is obtained using the least square minimization method. It is assumed that the number of exemplars is m and the number of pixels is d. To represent the image $\tilde{E}$ by the product of T and L_R, T must be found, using the following equation,

$$\begin{pmatrix} \tilde{E}_1(x_1), \ldots, \tilde{E}_m(x_1) \\ \vdots \quad \ddots \quad \vdots \\ \tilde{E}_1(x_d), \ldots, \tilde{E}_m(x_d) \end{pmatrix} = \begin{pmatrix} T(x_1) \\ \vdots \\ T(x_d) \end{pmatrix} \begin{pmatrix} L_{R1} \\ \vdots \\ L_{Rm} \end{pmatrix}^{\mathrm{T}}. \tag{19}$$

The least square solution satisfies $\tilde{\mathbf{E}}\mathbf{L_R} = \mathbf{T}\mathbf{L_R^T}\mathbf{L_R}$. If the columns of $\mathbf{L_R}$ are linearly independent, then $\mathbf{L_R}$ is nonsingular and has an inverse

$$T^* = \left(\mathbf{L_R^T}\mathbf{L_R}\right)^{-1} \tilde{\mathbf{E}}\mathbf{L_R}. \tag{20}$$

The surface normal N can now be estimated as follows

$$N^* = \frac{T^*}{\|T\|}, \tag{21}$$

where $N^* = \{n_x^*, n_y^*, n_z^*\}$ and the albedo is assumed to equal $\|T^*\|$.

3.3. *3D face reconstruction from surface gradients*

In this section a method for imposing integrability is introduced, demonstrating that it will indeed work in practice. A surface function $f = z(x, y)$ is characterized by the validity or invalidity of the surface normal integration. The shading based shape recovery techniques include the proposed method, normally providing gradient values for a discrete set of visible points on object surfaces. This requires a subsequent integration step, in order to achieve the specified alternative goal of depth or height maps. The general considerations regarding integration paths support local integration. A scan algorithm, which passes through all image points of the image grid, is assumed. Starting with initial depth values, this algorithm can be used to propagate depth values according to a local approximation rule, using the given gradient data. A calculation with relative depth values can be achieved within repeated scans. The resulting depth values can be determined by averaging operations. The initial depth values are required to be provided or assumed for the start positions of the different runs. However, this method is based on the global integration for the task of calculating depth from gradients. The solution calculated by the *Frankot-Chellappa algorithm*,[5] is optimal in the sense of the quadratic error function between the ideal and given gradient values. It only provides a relative depth function up to an additive constant. The depth $z(x, y)$ and gradient citation $\{p(x, y)q(x, y)\}$ mentioned in Eqs. (7) and (8), are functions over the image grid $\{(x, y) : 1 \le x \le d_w \wedge 1 \le y \le d_h\}$ where d_w is the number of pixels for width and d_h is the number of pixels for height. The surface $z(x, y)$ is then obtained by numerical integration of the surface gradients,

$$z(x, y) = z(x_0, y_0) + \int p(x, y)dx + q(x, y)dx. \tag{22}$$

It has been observed that the shape from shading can be expressed as a problem of solving a first-order partial differential equation in x and y.[7] In deriving solutions to Eq. (22) by the calculus of variations, it appears to be much more straightforward to solve for surface orientation than to solve directly for $z(x, y)$. The question of consistency between $z_x(x, y)$ and $z_y(x, y)$ then arises. A reasonable consistency constraint to place on the surface slopes is that they can be integrated, defined by

$$z_x(x, y) = z_y(x, y), \tag{23}$$

where

$$\begin{aligned} z_x(x, y) &= \frac{\partial z(x, y)}{\partial x} \approx z(x + 1, y) - z(x, y) \\ z_y(x, y) &= \frac{\partial z(x, y)}{\partial y} \approx z(x, y + 1) - z(x, y) \end{aligned}. \tag{24}$$

The cost function $C_f(z(x, y))$ is minimized by setting its derivatives with respect to the surface gradients p and q, to zero. The resulting rule is applied pixelwise. The surface depth values recovered by integration of the surface normal vectors are scaled with respect to the depth values of the true 3D surface,

$$C_f(z(x,y)) = \sum_{x,y} [(z_x(x,y) - p(x,y))^2 + (z_y(x,y) - q(x,y))^2]. \tag{25}$$

The depth map (surface depth) is defined relatively to an assumed background plane (of depth zero), which is parallel to the image plane. In the ideal case at each point (x, y), the depth value is equal to the (scaled) depth of those surface points projected onto this image point. The depth surface is measured with respect to the chosen background plane. A given depth map allows reconstruction of the faces in 3D space within a subsequent computation step of a general back projection approach.

4. Experiments

We have conducted a number of experiments with our approach for reconstruction of 3D face using the illumination subset of the Yale face database B and MPI face database. The results demonstrate that the proposed reconstruction algorithm can recover surface normal and 3D shape with Lambertian reflectance assumptions.

4.1. *Face database*

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

The MPI face database is used to demonstrate the proposed approach. 200 two-dimensional images of Caucasian faces that were rendered from a database of three-dimensional head models recorded with a laser scanner were used (Cyberware$^{\mathrm{TM}}$).[7,12] The images were rendered from a viewpoint 120 cm in front of each face, with ambient light only. For training, the images of 100 people were used. The images of the remaining 100 people were used for testing the proposed algorithm. Images consisting of 25 changes, {L11, L12, ..., L45, L55}, in illumination directions are used (see Fig. 4). 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 was 256×256 pixels and the color images were converted into 8-bit gray level images. Figure 4 presents the position configuration of the lights (a) and example images of lighting condition changes (b). Since it is convenient to exclude background, we used a MPI DB in most experiments and figures (Figs. 8–11).

4.1.2. *Yale face database B*

The Yale face database[6] is used to demonstrate the proposed approach. In order to capture the images in this database, a geodesic lighting rig was constructed with 64 computer controlled xenon strobes whose positions in spherical coordinates are

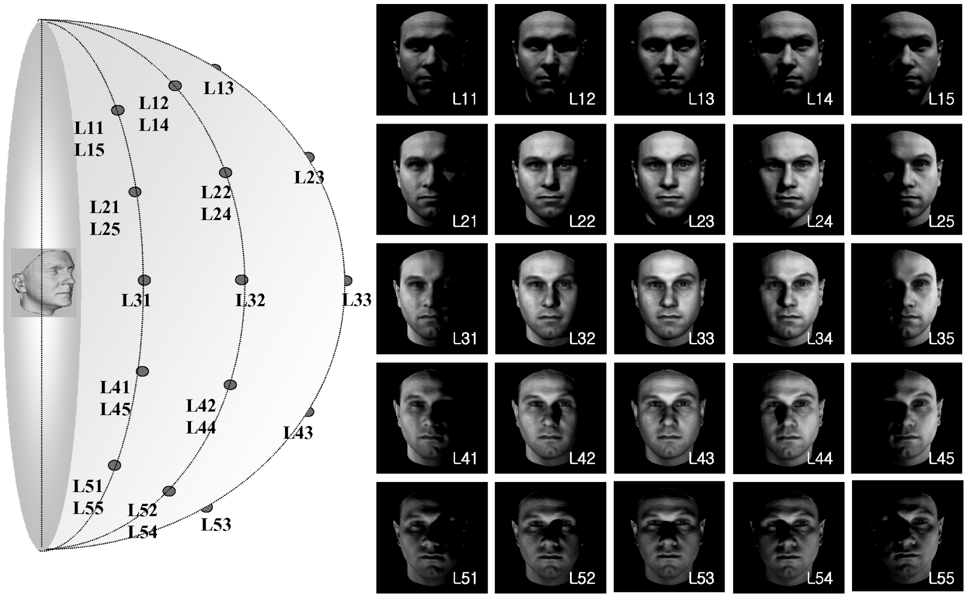
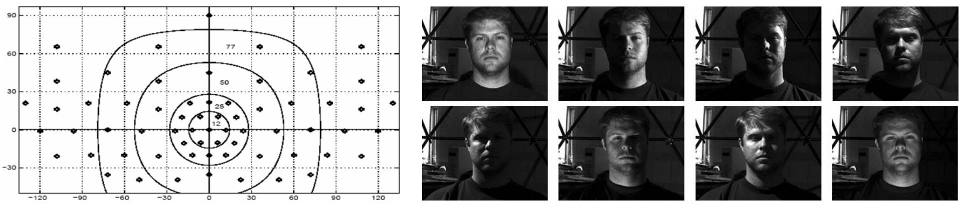


Fig. 4. MPI face database. (a) The position configuration of lights. (b) Example of data set.

presented in Fig. 5. Images of ten individuals were acquired under 64 different lighting conditions in 9 poses. The 64 images of a face in a particular pose are acquired in approximately 2 s. Therefore, there is only minimal change in head position and facial expression in those 64 images. Of the 64 images per person in each pose, 45 were used in the experiments, for a total of 4050 images (9 poses $\times$ 45 illumination conditions $\times$ 10 faces). The images from each pose were divided into four subsets (12, 25, 50, and 77) according to the angle the light source direction makes with the camera's axis. The original size of the images was 640 $\times$ 480 pixels. In the experiments, all images were manually cropped to include only the face, with as little hair and background as possible, at 128 $\times$ 128 pixels. Figure 5 presents the position configuration of the lights (a) and example images of lighting condition changes (b). Each annulus contains the positions of the strobes (azimuth and elevation of the 64 strobes) corresponding to the images of each illumination subset.

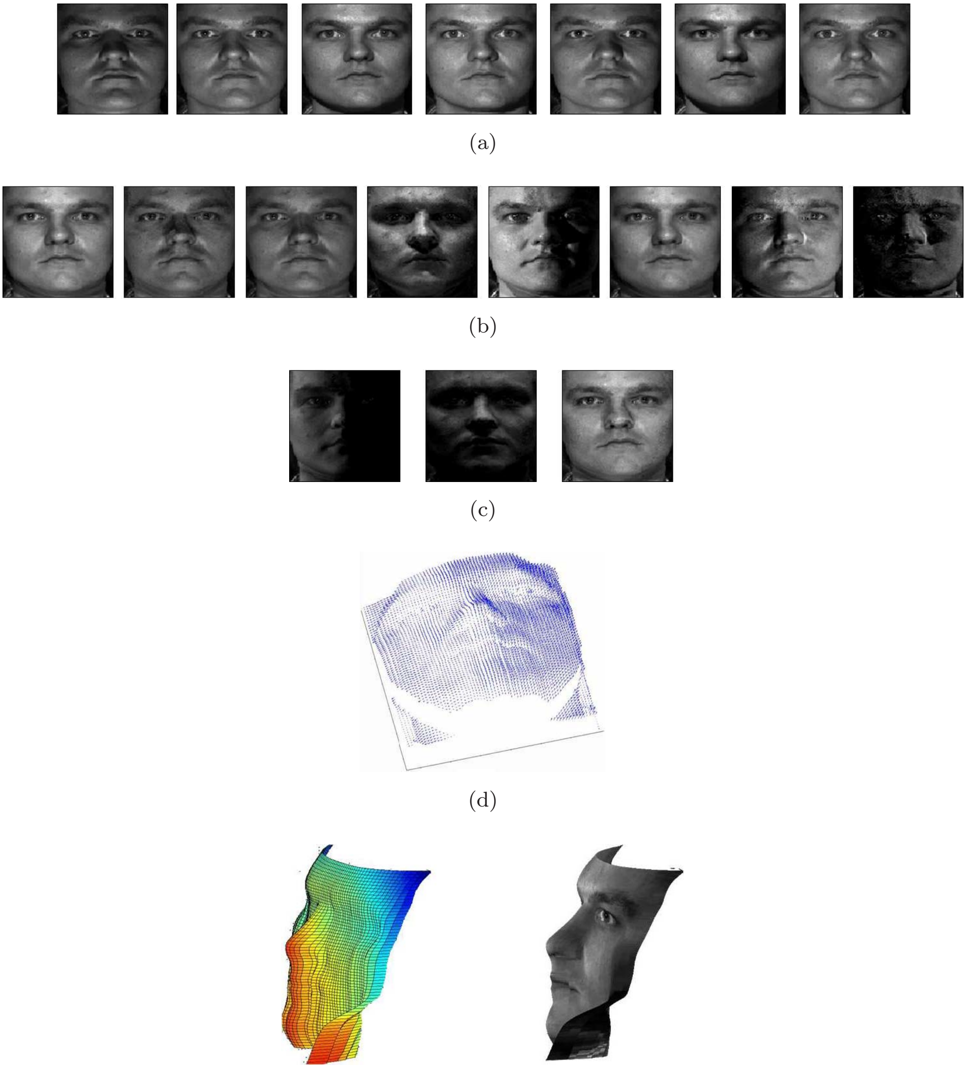


(a) (b)

Fig. 5. Yale face database B. (a) The position configuration of lights. (b) Example of data set.

4.2. *Reconstruction results and analysis*

In this section the reconstruction results for faces are shown. Figure 6 demonstrates the results of each process for constructing the 3D face shape using Yale face database B, seven images are used per person under arbitrary illuminations to recover surface normal in the illumination subset. Figure 6(a) shows the seven images for synthesizing illuminated exemplars and (b) shows examples of



(a)

(b)

(c)

(d)

(e)

Fig. 6. Face reconstruction of Yale face database B. (a) Training images. (b) Example of illuminated exemplars. (c) Basis images which are transformed from illuminated exemplars. (d) The surface normal. (e) 3D depth and surface of face.

synthesized exemplar under known lighting conditions. The illuminated exemplars can be synthesized under the lighting of all directions. Figure 6(c) shows the basis images which will be used to recover the surface normal, (d) shows the x, y, z components of the surface normal in Eq. (17) and (e) shows the 3D depth and 3D surface of faces with 2D texture image.

Figure 7 also presents the results of 3D face reconstruction of the MPI face database. As demonstrated, the 3D shape of face can be reconstructed only using shading information of 2D images under arbitrary illumination. In these experiments, the first image is used as texture of the 3D shape, but is not a good choice. For the probe image under arbitrary illumination, some process is required to handle the illumination problem between the new probe image and projected image of the 3D face (see next section).

Image synthesis of the same individual under lighting and pose variation: if the 3D information (depth) can be obtained, images can be synthesized under arbitrary illumination and the images in the database can be used more. Figure 8 presents synthetic images of a face under novel variations in lighting and pose. These images are projected from a reconstructed 3D face using synthesized exemplars.

4.3. *Fitting for lighting variation (illumination fitting)*

In this section, the proposed method of illumination fitting is introduced. For face recognition using the reconstructed 3D face under varying pose and illumination, the poses and lighting condition of the input image should be known. It is assumed that the pose of a new probe image is already estimated. The recognition procedure is performed by computing the correlation between the probe and projected image of the 3D face. The 3D face is constructed using photometric stereo images, the image under front lighting or first gallery image is used as a texture image. Therefore, the fitting procedure should be required for improvement in face recognition

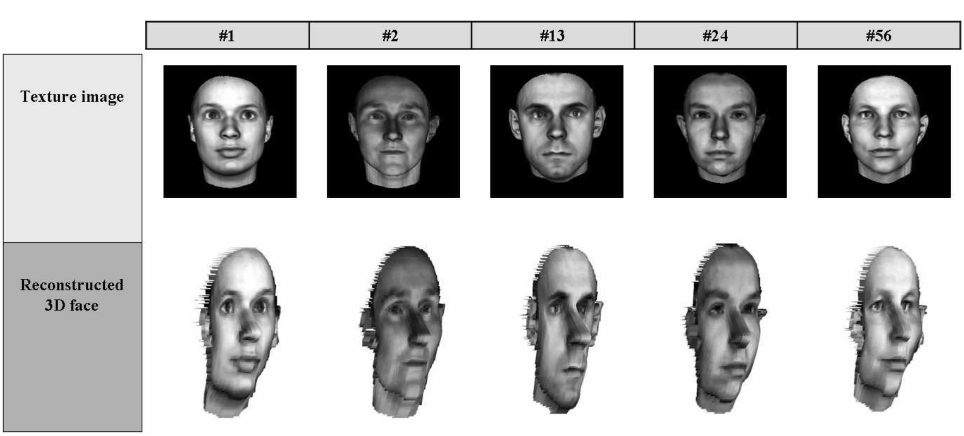


Fig. 7. Results of 3D face reconstruction of MPI face database B.

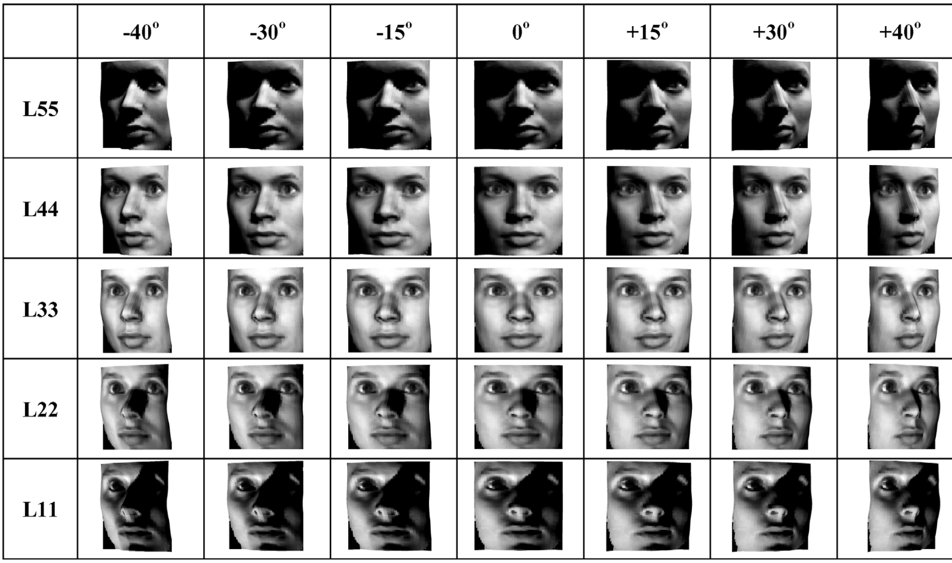


Fig. 8. Synthesized images of the same individual under novel variations in illumination and pose. Each row represents the variation in lighting (from the direction of L11 to the direction of L55) and each column represents the variation in pose (from −40 to 40).

performance. Figure 9 presents the basic idea of the proposed illumination fitting procedure. n gallery images under varying lighting conditions exist, and are used. After reconstructing the 3D face, these images are used as texture images and then n projected images are obtained. Using linear combination of these projected images, any image under arbitrary illumination (based on photometric stereo method[13]) can be represented.

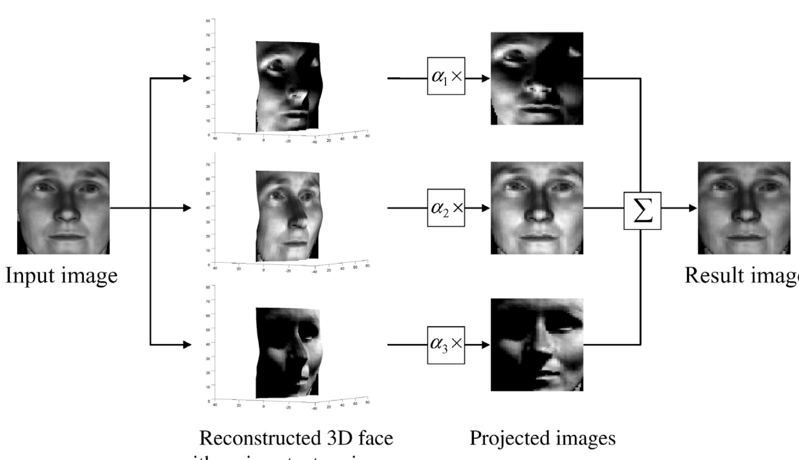


Fig. 9. The basic idea for illumination fitting.

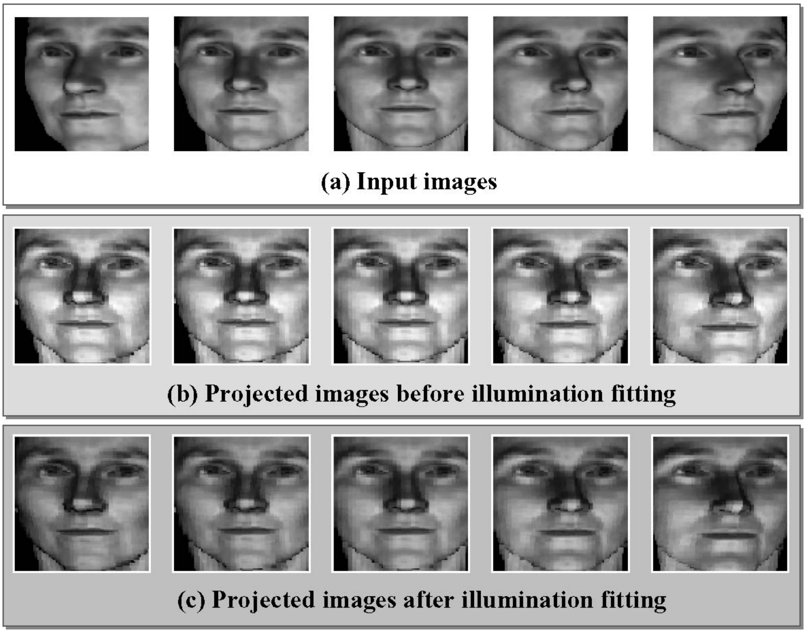


Fig. 10. Examples of projected images before and after illumination fitting.

Figure 10 presents the results of fitting for lighting variation. Figure 10(a) represents the input image under arbitrary illumination and varying pose, (b) represents the images before illumination fitting and (c) represents the resulting images after illumination fitting. As presented in Fig. 10, the projected images after illumination fitting are reconstructed better. The results of reconstruction and recognition are compared by computing reconstruction error and correlation between the input image and projected image (see Fig. 11). The reconstruction error means the mean displacement error per pixel, the horizontal axes represent the direction of pose variations while the vertical axes represents the mean displacement error in input image per pixel.

5. Conclusion and Further Research

In this paper, a simple and practical approach for 3D face reconstruction has been proposed. The proposed approach is based on photometric stereo, which recovers the surface normal from multiple images under arbitrary illuminations. The illuminated exemplars are synthesized under known lighting conditions from gallery images under unknown lighting conditions using illumination reference. In using synthesized exemplars, the surface normal for 3D depth values of the face can be estimated without iteration and correspondence information. A noniterative 3D face reconstruction algorithm can be used for a wide range of practical applications regarding pose estimation of face and pose or illumination invariant face recognition

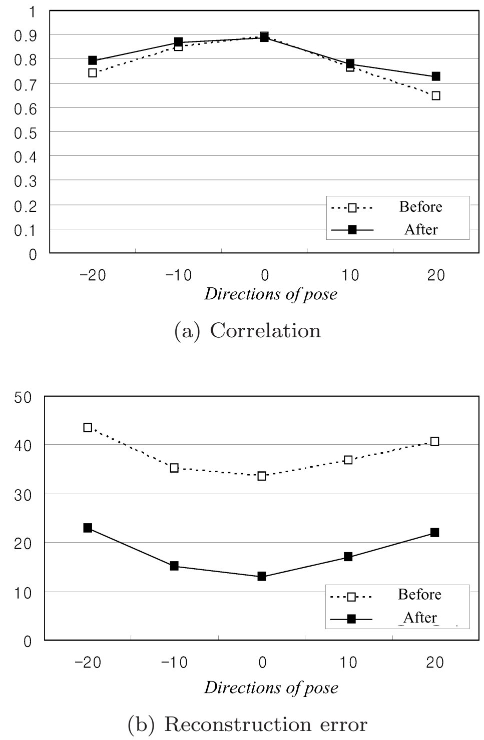


Fig. 11. Comparisons of recognition results and mean reconstruction errors.

systems. In the experiments, a small number of images under the random direction of lighting are chosen to synthesize exemplars. Experimental results regarding various face images have demonstrated good reconstruction results when compared with the previous approaches. In the future, arbitrary views will be synthesized using the 3D face shape, which is reconstructed from the proposed approach for pose invariant face recognition of face images in a real environment. An attempt is also being made to handle cast shadow and cases of multiple light sources.

Acknowledgments

The authors would like to thank the Max-Planck-Institute for providing the MPI Face Database. This research was supported by the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea and in part by research funds from Chosun University, 2007.

References

1. R. Basriand and D. Jacobs, Photometric stereo with general, unknown lighting, *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition* (2001), pp. 374–381.

2. P. N. Belhumeur, D. J. Kriegman and A. L. Yuille, The bas-relief ambiguity, *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition* **17** (1997) 1060–1066.
3. D. Beymer and T. Poggio, *Face Recognition from One Example View*, Artificial Intelligence, Memo No. 1536, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1995).
4. V. Blanz, S. Romdhani and T. Vetter, Face identification across different poses and illuminations with a 3D morphable model, *Proc. 5th Int. Conf. Automatic Face and Gesture Recognition* (2002), pp. 202–207.
5. R. Frankot and R. Chellappa, A method for enforcing integrability in shape from shading algorithms, *IEEE Trans. Patt. Anal. Mach. Intell.* **10** (1988) 439–451.
6. A. Georghiades, P. Belhumeur and D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Patt. Anal. Mach. Intell.* **23**(6) (2001) 643–660.
7. B. Horn and J. Brooks, *Shape from Shading* (MIT Press, 1989).
8. A. Pentland, B. Moghaddam and T. Starner, View-based and modular eigenspaces for face recognition, *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition* (1994).
9. R. Szeliski and S. Kang, *Recovering 3D Shape and Motion from Image Streams Using Non-Linear Least Squares*, Technical Report, Robotics Institute, Carnegie Mellon University (March 1993).
10. P. S. P. Wang and S. N. Y. Yanushkevich, Biometric technologies and applications, *Proc. 25th IASTED Int. Multi-Conf. Aritificial Intelligence and Applications* (2007), pp. 226–231.
11. X. You, D. Zhang, Q. Chen, P. Wang and Y. Y. Tang, Face representation by using non-tensor product wavelets, *Int. Conf. Pattern Recognition* **1** (2006) 502–505.
12. W. Zhao, *Robust Image Based 3D Face Recognition*, Ph.D. thesis, University of Maryland (1999).
13. K. Zhou and R. Chellappa, Rank constrained recognition under unknown illuminations, *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures* **17**(17) (2003) 11–18.

Sang-Woong Lee received his B.S. degree in electronics and computer engineering from Korea University, Seoul, Korea, in 1996 and his M.S. and Ph.D. degrees in computer science and engineering from Korea University, Seoul, Korea, in 2001 and 2006, respectively. From June 2006 to May 2007, he was a visiting scholar in Robotics Institute, Carnegie Mellon University. Currently, he is a full-time instructor in the Department of Computer Science and Engineering at Chosun University, Gwangju, Korea.

His present research interests include face recognition, gesture recognition and related realtime applications.

Patrick S. P. Wang is tenured full professor of computer science at College of Computer and Information Science of Northeastern University, Boston, MA, USA. He is *Fellow* of IAPR (International Association of Pattern Recognition) and co-founding chief editor of IJPRAI (*International Journal of Pattern Recognition and Artificial Intelligence*) and WSP MPAI Book Series. Dr. Wang has been elected as Otto-Von-Guericke Distinguished Guest Professor at the Graphics and Simulations Lab of the University of Magdeburg (near Berlin) of Deutschland (Germany). Dr. Wang has been visiting at University of Calgary's Biometrics Technology Laboratory, Calgary, Canada as *iCORE* (Informatics Circle of Research Excellence) visiting Professor, where he published his 23rd book in July 2007, entitled: *Image Pattern Recognition — Synthesis and Analysis in Biometrics* by World Scientific Publishing Co. He is now *Zi-Jiang Visiting Chair Professor* at the Department of Computer Science and Technology, East China Normal University, Shanghai, China. Professor Wang has also served as Honorary Advisor Professor of Xiamen University and Sichuan University of China.

Dr. Wang got his B.S. in electronic engineering from National Chiao Tung University (Hsin Chu), M.S. in electrical engineering from National Taiwan University, M.S. in information and computer science from Georgia Institute of Technology, Ph.D. in computer science from Oregon State University. He has been Visiting Scientist at MIT AI Lab, and Research Consultant at MIT Sloan School, as well as adjunct faculty of Computer Science at Harvard University Extension School. For the past decades, he has been invited to give lectures in many countries in America, Asia and Europe. He has served as the Chair of IAPR-SSPR TC-2 (Structural and Syntactic Pattern Recognition) for 4 years (1992–1996). He has also organized many international conferences, including Data Mining (MLDM2003 and 2005) at Leipzig and BT2004, Int. Workshop on Biometric Technology, 2004, Calgary, Canada. Dr. Wang was general conference co-chair of the IAPR's 18th ICPR International Conference on Pattern Recognition, August, 2006 at Hong Kong, China. Prof. Wang serves as Program Vice-Chair and Distinguished International Advisor of the *IEEE-BIBE2007* at Harvard Medical. He gave a keynote plenary lecture on *Biometrics Intelligence Information Systems and Applications* and was honored and received a Plaque of *Outstanding Achievement Award* from Conference General Chair Prof. Jack Yang and IEEE-SMC President Prof. Larry Hall. He also gave a tutorial *on IPR and Biometrics* at the *IEEE-SMC 2007*, Montreal, Canada, October 2007. Dr. Wang was also invited to give a keynote speech and tutorial at IAS 2007, Manchester, UK, and e-Forensics 2008 at Adelaide, Australia.

His other literature works on music and poems have received many awards including the prestigious World Journal Prose Competition Award, and his book *Harvard Meditation Melody* (哈佛冥想曲) has won the prestigious Taiwan Provincial Better Publication Award.

Svetlana N. Yanushkevich received M.S. (1989) and Ph.D. (1992) degrees in electrical and computer engineering from the State University of Informatics and Radioelectronics, Minsk, Belarus, and she received her Habilitation degree (1999) in engineering science from the University of Technology, Warsaw, Poland. She joined the Department of Electrical and Computer Engineering, University of Calgary, Canada, in 2001. She is the founding director of the "Biometric Technologies: Modeling and Simulatio" Laboratory at Calgary. Dr. Yanushkevich is a Senior Member of the IEEE, a Member of IAPR, and a Member of the Institute of Electronics, Information and Communication Engineers (IEICE), Japan.

She also acts as a consultant to several government and commercial institutions. Dr. Yanushkevich is a Member of the IEEE Technical Committee on Multiple-Valued Logic.

She serves as a Guest Editor of several international journals, and she was a general chair and co-chair of about 20 international conferences, symposia, and workshops. Dr. Yanushkevich published over 200 technical papers, including 15 patents, 8 books, co-authored, and 3 edited books. She co-authored (with V. P. Shmerko, D. Popel, and A. Stoica) the monograph *Biometric Inverse Problems* (Taylor&Francis/CRC Press, 2005).

Dr. Yanushkevich's current research interests include artificial intelligence in decision making, biometrics, computer aided design of systems, and nanocomputing in 3D.

Seong-Whan Lee received his B.S. degree in computer science and statistics from Seoul National University, Korea, in 1984, and his M.S. and Ph.D. 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, Korea. In March 1995, he joined the faculty of the Department of Computer Science and Engineering at Korea University, Seoul, 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 associate editor of the *Pattern Recognition Journal*, the *International Journal of Pattern Recognition and Artificial Intelligence*, the *International Journal on Document Analysis and Recognition*, and the *International Journal of Computer Processing of Oriental Languages*. He is a fellow of IAPR, a senior member of the IEEE Computer Society and a life member of the Korea Information Science Society and the Oriental Languages Computer Society.

His research interests include pattern recognition, computer vision and intelligent robots. He has published more than 200 publications in these areas in international journals and conference proceedings, and has authored five books.