Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2008, Article ID 748483, 18 pages
doi:10.1155/2008/748483

Research Article

Pose-Encoded Spherical Harmonics for Face Recognition and Synthesis Using a Single Image

Zhanfeng Yue,[1] Wenyi Zhao,[2] and Rama Chellappa[1]

[1] *Center for Automation Research, University of Maryland, College Park, MD 20742, USA*
[2] *Vision Technologies Lab, Sarnoff Corporation, Princeton, NJ 08873, USA*

Correspondence should be addressed to Zhanfeng Yue, zyue@cfar.umd.edu

Received 1 May 2007; Accepted 4 September 2007

Recommended by Juwei Lu

Face recognition under varying pose is a challenging problem, especially when illumination variations are also present. In this paper, we propose to address one of the most challenging scenarios in face recognition. That is, to identify a subject from a test image that is acquired under different pose and illumination condition from only one training sample (also known as a gallery image) of this subject in the database. For example, the test image could be semifrontal and illuminated by multiple lighting sources while the corresponding training image is frontal under a single lighting source. Under the assumption of Lambertian reflectance, the spherical harmonics representation has proved to be effective in modeling illumination variations for a fixed pose. In this paper, we extend the spherical harmonics representation to encode pose information. More specifically, we utilize the fact that 2D harmonic basis images at different poses are related by close-form linear transformations, and give a more convenient transformation matrix to be directly used for basis images. An immediate application is that we can easily synthesize a different view of a subject under arbitrary lighting conditions by changing the coefficients of the spherical harmonics representation. A more important result is an efficient face recognition method, based on the orthonormality of the linear transformations, for solving the above-mentioned challenging scenario. Thus, we directly project a nonfrontal view test image onto the space of frontal view harmonic basis images. The impact of some empirical factors due to the projection is embedded in a sparse warping matrix; for most cases, we show that the recognition performance does not deteriorate after warping the test image to the frontal view. Very good recognition results are obtained using this method for both synthetic and challenging real images.

1. INTRODUCTION

Face recognition is one of the most successful applications of image analysis and understanding [1]. Given a database of training images (sometimes called a gallery set, or gallery images), the task of face recognition is to determine the facial ID of an incoming test image. Built upon the success of earlier efforts, recent research has focused on robust face recognition to handle the issue of significant difference between a test image and its corresponding training images (i.e., they belong to the same subject). Despite significant progress, robust face recognition under varying lighting and different pose conditions remains to be a challenging problem. The problem becomes even more difficult when only one training image per subject is available. Recently, methods have been proposed to handle the combined pose and illumination problem when only one training image is available, for example, the method based on morphable models [2] and its extension [3] that proposes to handle the complex illumination problem by integrating spherical harmonics representation [4, 5]. In these methods, either arbitrary illumination conditions cannot be handled [2] or the expensive computation of harmonic basis images is required for each pose per subject [3].

Under the assumption of Lambertian reflectance, the spherical harmonics representation has proved to be effective in modelling illumination variations for a fixed pose. In this paper, we extend the harmonic representation to encode pose information. We utilize the fact that all the harmonic basis images of a subject at various poses are related to each other via close-form linear transformations [6, 7], and derive a more convenient transformation matrix to analytically synthesize basis images of a subject at various poses from just one set of basis images at a fixed pose, say, the frontal

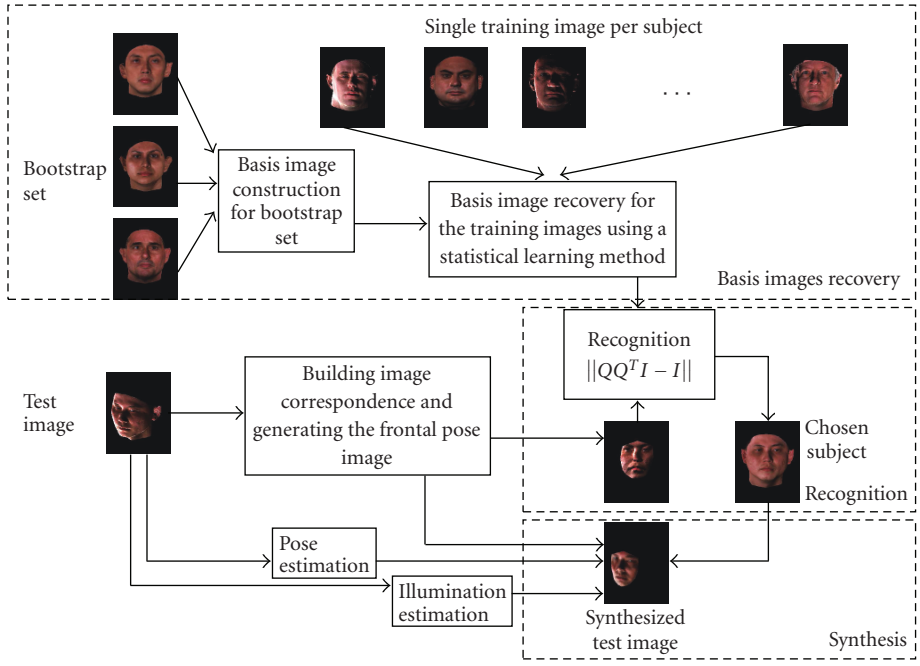


Figure 1: The proposed face synthesis and recognition system.

view [8]. We prove that the derived transformation matrix is consistent with the general rotation matrix of spherical harmonics. According to the theory of spherical harmonics representation [4, 5], this implies that we can easily synthesize from one image under a fixed pose and lighting to an image acquired under different poses and arbitrary lightings. Moreover, these linear transformations are orthonormal. This suggests that recognition methods based on projection onto fixed-pose harmonic basis images [4] for test images under the same pose can be easily extended to handle test images under various poses and illuminations. In other words, we do not need to generate a new set of basis images at the same pose as that of test image. Instead, we can warp the test images to a frontal view and directly use the existing frontal view basis images. The impact of some empirical factors (i.e., correspondence and interpolation) due to the warping is embedded in a sparse transformation matrix; for most cases, we show that the recognition performance does not deteriorate after warping the test image to the frontal view.

To summarize, we propose an efficient face synthesis and recognition method that needs only one single training image per subject for novel view synthesis and robust recognition of faces under variable illuminations and poses. The structure of our face synthesis and recognition system is shown in Figure 1. We have a single training image at the frontal pose for each subject in the training set. The basis images for each training subject are recovered using a statistical learning algorithm [9] with the aid of a bootstrap set consisting of 3D face scans. For a test image at a rotated pose and under an arbitrary illumination condition, we manually establish the image correspondence between the test image and a mean face image at the frontal pose. The frontal view image is then synthesized from the test image. A face is identified for which there exists a linear reconstruction based on basis images that is the closest to the test image. Note that although in Figure 1 we only show the training images acquired at the frontal pose, it does not exclude other cases when the available training images are at different poses. Furthermore, the user is given the option to visualize the recognition result by comparing the synthesized images of the chosen subject against the test image. Specifically, we can generate novel images of the chosen subject at the same pose as the test image by using the closed-form linear transformation between the harmonic basis images of the subject across poses. The pose of the test image is estimated from a few manually selected main facial features.

We test our face recognition method on both synthetic and real images. For synthetic images, we generate the training images at the frontal pose and under various illumination conditions, and the test images at different poses, under arbitrary lighting conditions, all using Vetter's 3D face database [10]. For real images, we use the CMU-PIE [11] database which contains face images of 68 subjects under 13 different poses and 43 different illumination conditions. The test images are acquired at six different poses and under twenty one different lighting sources. High recognition rates are achieved on both synthetic and real test images using the proposed algorithm.

The remainder of the paper is organized as follows. Section 2 introduces related work. The pose-encoded spherical harmonic representation is illustrated in Section 3 where we derive a more convenient transformation matrix to

analytically synthesize basis images at one pose from those at another pose. Section 4 presents the complete face recognition and synthesis system. Specifically, in Section 4.1 we briefly summarize a statistical learning method to recover the basis images from a single image when the pose is fixed. Section 4.2 describes the recognition algorithm and demonstrates that the recognition performance does not degrade after warping the test image to the frontal view. Section 4.3 presents how to generate the novel image of the chosen subject at the same pose as the test image for visual comparison. The system performance is demonstrated in Section 5. We conclude our paper in Section 6.

2. RELATED WORK

As pointed out in [1] and many references cited therein, pose and/or illumination variations can cause serious performance degradation to many existing face recognition systems. A review of these two problems and proposed solutions can be found in [1]. Most earlier methods focused on either illumination or pose alone. For example, an early effort to handle illumination variations is to discard the first few principal components that are assumed to pack most of the energy caused by illumination variations [12]. To handle complex illumination variations more efficiently, spherical harmonics representation was independently proposed by Basri and Jacobs [4] and Ramamoorthi [5]. It has been shown that the set of images of a convex Lambertian face object obtained under a wide variety of lighting conditions can be approximated by a low-dimensional linear subspace. The basis images spanning the illumination space for each face can then be rendered from a 3D scan of the face [4]. Following the statistical learning scheme in [13], Zhang and Samaras [9] showed that the basis images spanning this space can be recovered from just one image taken under arbitrary illumination conditions for a fixed pose.

To handle the pose problem, a template matching scheme was proposed in [14] that needs many different views per person and does not allow lighting variations. Approaches for face recognition under pose variations [15, 16] avoid the strict correspondence problem by storing multiple normalized images at different poses for each person. View-based eigenface methods [15] explicitly code the pose information by constructing an individual eigenface for each pose. Reference [16] treats face recognition across poses as a bilinear factorization problem, with facial identity and head pose as the two factors.

To handle the combined pose and illumination variations, researchers have proposed several methods. The synthesis method in [17] can handle both illumination and pose variations by reconstructing the face surface using the illumination cone method under a fixed pose and rotating it to the desired pose. The proposed method essentially builds illumination cones at each pose for each person. Reference [18] presented a symmetric shape-from-shading (SFS) approach to recover both shape and albedo for symmetric objects. This work was extended in [19] to recover the 3D shape of a human face using a single image. In [20], a unified approach was proposed to solve the pose and illumination problem. A generic 3D model was used to establish the correspondence and estimate the pose and illumination direction. Reference [21] presented a pose-normalized face synthesis method under varying illuminations using the bilateral symmetry of the human face. A Lambertian model with a single light source was assumed. Reference [22] extended the photometric stereo algorithms to recover albedos and surface normals from one image illuminated by unknown single or multiple distant illumination source.

Building upon the highly successful statistical modeling of 2D face images [23], the authors in [24] propose a 2D + 3D active appearance model (AAM) scheme to enhance AAM in handling 3D effects to some extent. A sequence of face images (900 frames) is tracked using AAM and a 3D shape model is constructed using structure-from-motion (SFM) algorithms. As camera calibration and 3D reconstruction accuracy can be severely affected when the camera is far away from the subjects, the authors imposed these 3D models as soft constraints for the 2D AAM fitting procedure and showed convincing tracking and image synthesis results on a set of five subjects. However, this is not a true 3D approach with accurate shape recovery and does not handle occlusion.

To handle both pose and illumination variations, a 3D morphable face model has been proposed in [2], where the shape and texture of each face is represented as a linear combination of a set of 3D face exemplars and the parameters are estimated by fitting a morphable model to the input image. By far the most impressive face synthesis results were reported in [2] accompanied by very high recognition rates. In order to effectively handle both illumination and pose, a recent work [3] combines spherical harmonics and the morphable model. It works by assuming that shape and pose can be first solved by applying the morphable model and illumination can then be handled by building spherical harmonic basis images at the resolved pose. Most of the 3D morphable model approaches are computationally intense [25] because of the large number of parameters that need to be optimized. On the contrary, our method does not require the time-consuming procedure of building a set of harmonic basis images for each pose. Rather, we can analytically synthesize many sets of basis images from just one set of basis images, say, the frontal basis images. For the purpose of face recognition, we can further improve the efficiency by exploring the orthonormality of linear transformations among sets of basis images at different poses. Thus, we do not synthesize basis images at different poses. Rather, we warp the test image to the same pose as that of the existing basis images and perform recognition.

3. POSE-ENCODED SPHERICAL HARMONICS

The spherical harmonics are a set of functions that form an orthonormal basis for the set of all square-integrable functions defined on the unit sphere [4]. Any image of a Lambertian object under certain illumination conditions is a linear combination of a series of spherical harmonic basis images $\{b_{lm}\}$. In order to generate the basis images for the object, 3D

information is required. The harmonic basis image intensity of a point p with surface normal $n = (n_x, n_y, n_z)$ and albedo λ can be computed as the combination of the first nine spherical harmonics, shown in (1), where $n_{x^2} = n_x n_x$. n_{y^2}, n_{z^2}, n_{xy}, n_{xz}, n_{yz} are defined similarly. $\lambda.*t$ denotes the componentwise product of λ with any vector t. The superscripts e and o denote the even and the odd components of the harmonics, respectively:

$$
\begin{aligned}
b_{00} &= \frac{1}{\sqrt{4\pi}}\lambda, \qquad b_{10} = \sqrt{\frac{3}{4\pi}}\lambda.*n_z,\\
b_{11}^e &= \sqrt{\frac{3}{4\pi}}\lambda.*n_x, \qquad b_{11}^o = \sqrt{\frac{3}{4\pi}}\lambda.*n_y,\\
b_{20} &= \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda.*(2n_{z^2} - n_{x^2} - n_{y^2}),\\
b_{21}^e &= 3\sqrt{\frac{5}{12\pi}}\lambda.*n_{xz}, \qquad b_{21}^o = 3\sqrt{\frac{5}{12\pi}}\lambda.*n_{yz},\\
b_{22}^e &= \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda.*(n_{x^2} - n_{y^2}), \qquad b_{22}^o = 3\sqrt{\frac{5}{12\pi}}\lambda.*n_{xy}.
\end{aligned}
\tag{1}
$$

Given a bootstrap set of 3D models, the spherical harmonics representation has proved to be effective in modeling illumination variations for a fixed pose, even in the case when only one training image per subject is available [9]. In the presence of both illumination and pose variations, two possible approaches can be taken. One is to use a 3D morphable model to reconstruct the 3D model from a single training image and then build spherical harmonic basis images at the pose of the test image [3]. Another approach is to require multiple training images at various poses in order to recover the new set of basis images at each pose. However, multiple training images are not always available and a 3D morphable model-based method could be computationally expensive. As for efficient recognition of a rotated test image, a natural question to ask is that can we represent the basis images at different poses using one set of basis images at a given pose, say, the frontal view. The answer is yes, and the reason lies on the fact that 2D harmonic basis images at different poses are related by close-form linear transformations. This enables an analytic method for generating new basis images at poses different from that of the existing basis images.

Rotations of spherical harmonics have been studied by researchers [6, 7] and it can be shown that rotations of spherical harmonic with order l are linearly composed entirely of other spherical harmonics of the same order. In terms of group theory, the transformation matrix is the $(2l + 1)$-dimensional representation of the rotation group SO (3) [7]. Let $Y_{l,m}(\psi, \varphi)$ be the spherical harmonic, the general rotation formula of spherical harmonic can be written as $Y_{l,m}(R_{\theta,\omega,\beta}(\psi, \varphi)) = \sum_{m'=-l}^{l} D_{mm'}^{l}(\theta, \omega, \beta) Y_{l,m'}(\psi, \varphi)$, where θ, ω, β are the rotation angles around the Y, Z, and X axes, respectively. This means that for each order l, D^l is a matrix that tells us how a spherical harmonic transforms under rotation. As a matrix multiplication, the transformation is found to have the following block diagonal sparse form:

$$
\begin{bmatrix} Y'_{0,0} \\ Y'_{1,-1} \\ Y'_{1,0} \\ Y'_{1,1} \\ Y'_{2,-2} \\ Y'_{2,-1} \\ Y'_{2,0} \\ Y'_{2,1} \\ Y'_{2,2} \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & H_2 & H_3 & H_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & H_5 & H_6 & H_7 & 0 & 0 & 0 & 0 & 0 \\ 0 & H_8 & H_9 & H_{10} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & H_{11} & H_{12} & H_{13} & H_{14} & H_{15} \\ 0 & 0 & 0 & 0 & H_{16} & H_{17} & H_{18} & H_{19} & H_{20} \\ 0 & 0 & 0 & 0 & H_{21} & H_{22} & H_{23} & H_{24} & H_{25} \\ 0 & 0 & 0 & 0 & H_{26} & H_{27} & H_{28} & H_{29} & H_{30} \\ 0 & 0 & 0 & 0 & H_{31} & H_{32} & H_{33} & H_{34} & H_{35} \\ & & & \vdots & \vdots & \vdots & & & \end{bmatrix} \begin{bmatrix} Y_{0,0} \\ Y_{1,-1} \\ Y_{1,0} \\ Y_{1,1} \\ Y_{2,-2} \\ Y_{2,-1} \\ Y_{2,0} \\ Y_{2,1} \\ Y_{2,2} \\ \vdots \end{bmatrix},
\tag{2}
$$

where, $H_1 = D_{00}^0$, $H_2 = D_{-1,-1}^1$, $H_3 = D_{-1,0}^1$, $H_4 = D_{-1,1}^1$, $H_5 = D_{0,-1}^1$, $H_6 = D_{0,0}^1$, $H_7 = D_{0,1}^1$, $H_8 = D_{1,-1}^1$, $H_9 = D_{1,0}^1$, $H_{10} = D_{1,1}^1$, $H_{11} = D_{-2,-2}^2$, $H_{12} = D_{-2,-1}^2$, $H_{13} = D_{-2,0}^2$, $H_{14} = D_{-2,1}^2$, $H_{15} = D_{-2,2}^2$, $H_{16} = D_{-1,-2}^2$, $H_{17} = D_{-1,-1}^2$, $H_{18} = D_{-1,0}^2$, $H_{19} = D_{-1,1}^2$, $H_{20} = D_{-1,2}^2$, $H_{21} = D_{0,-2}^2$, $H_{22} = D_{0,-1}^2$, $H_{23} = D_{0,0}^2$, $H_{24} = D_{0,1}^2$, $H_{25} = D_{0,2}^2$, $H_{26} = D_{1,-2}^2$, $H_{27} = D_{1,-1}^2$, $H_{28} = D_{1,0}^2$, $H_{29} = D_{1,1}^2$, $H_{30} = D_{1,2}^2$, $H_{31} = D_{2,-2}^2$, $H_{32} = D_{2,-1}^2$, $H_{33} = D_{2,0}^2$, $H_{34} = D_{2,1}^2$, $H_{35} = D_{2,2}^2$. The analytic formula is rather complicated, and is derived in [6, equatioin (7.48)].

Assuming that the test image I_{test} is at a different pose (e.g., a rotated view) from the training images (usually at the frontal view), we look for the basis images at the rotated pose from the basis images at the frontal pose. It will be more convenient to use the basis image form as in (1), rather than the spherical harmonics form $Y_{l,m}(\psi, \varphi)$. The general rotation can be decomposed into three concatenated Euler angles around the X, Y, and Z axes, namely, elevation (β), azimuth (θ), and roll (ω), respectively. Roll is an in-plane rotation that can be handled much easily and so will not be discussed here. The following proposition gives the linear transformation matrix from the basis images at the frontal pose to the basis images at the rotated pose for orders $l = 0, 1, 2$, which capture 98% of the energy [4].

Proposition 1. *Assume that a rotated view is obtained by rotating a frontal view head with an azimuth angle $-\theta$. Given the correspondence between the frontal view and the rotated view, the basis images B' at the rotated pose are related to the basis images B at the frontal pose as*

$$
\begin{bmatrix} b'_{00} \\ b'_{10} \\ b'^{e}_{11} \\ b'^{o}_{11} \\ b'_{20} \\ b'^{e}_{21} \\ b'^{o}_{21} \\ b'^{e}_{22} \\ b'^{o}_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sin\theta & \cos\theta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_1 & C_2 & 0 & C_3 & 0 \\ 0 & 0 & 0 & 0 & C_4 & C_5 & 0 & C_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos\theta & 0 & -\sin\theta \\ 0 & 0 & 0 & 0 & C_7 & 0 & C_8 & C_9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} b_{00} \\ b_{10} \\ b^{e}_{11} \\ b^{o}_{11} \\ b_{20} \\ b^{e}_{21} \\ b^{o}_{21} \\ b^{e}_{22} \\ b^{o}_{22} \end{bmatrix},
\tag{3}
$$

where $C_1 = 1 - (3/2)\sin^2\theta$, $C_2 = -\sqrt{3}\sin\theta\cos\theta$, $C_3 = (\sqrt{3}/2)\sin^2\theta$, $C_4 = \sqrt{3}\sin\theta\cos\theta$, $C_5 = \cos^2\theta - \sin^2\theta$, $C_6 = -\cos\theta\sin\theta$, $C_7 = (\sqrt{3}/2)\sin^2\theta$, $C_8 = \cos\theta\sin\theta$, $C_9 = 1 - (1/2)\sin^2\theta$.

Further, if there is an elevation angle $-\beta$, the basis images B'' for the newly rotated view are related to B' in the following linear form:

$$\begin{bmatrix} b''_{00} \\ b''_{10} \\ b''^{e}_{11} \\ b''^{o}_{11} \\ b''_{20} \\ b''^{e}_{21} \\ b''^{o}_{21} \\ b''^{e}_{22} \\ b''^{o}_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\beta & 0 & \sin\beta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sin\beta & 0 & \cos\beta & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A_1 & 0 & A_2 & A_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cos\beta & 0 & 0 & \sin\beta \\ 0 & 0 & 0 & 0 & A_4 & 0 & A_5 & A_6 & 0 \\ 0 & 0 & 0 & 0 & A_7 & 0 & A_8 & A_9 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sin\beta & 0 & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} b'_{00} \\ b'_{10} \\ b'^{e}_{11} \\ b'^{o}_{11} \\ b'_{20} \\ b'^{e}_{21} \\ b'^{o}_{21} \\ b'^{e}_{22} \\ b'^{o}_{22} \end{bmatrix}, \tag{4}$$

where $A_1 = 1 - (3/2)\sin^2\beta$, $A_2 = \sqrt{3}\sin\beta\cos\beta$, $A_3 = (-\sqrt{3}/2)\sin^2\beta$, $A_4 = -\sqrt{3}\sin\beta\cos\beta$, $A_5 = \cos^2\beta - \sin^2\beta$, $A_6 = -\cos\beta\sin\beta$, $A_7 = (-\sqrt{3}/2)\sin^2\beta$, $A_8 = \cos\beta\sin\beta$, $A_9 = 1 - (1/2)\sin^2\beta$.

A direct proof (rather than deriving from the general rotation equations) of this proposition is given in the appendix, where we also show that the proposition is consistent with the general rotation matrix of spherical harmonics.

To illustrate the effectiveness of (3) and (4), we synthesized the basis images at an arbitrarily rotated pose from those at the frontal pose, and compared them with the ground truth generated from the 3D scan in Figure 2. The first three rows present the results for subject 1, with the first row showing the basis images at the frontal pose generated from the 3D scan, the second row showing the basis images at the rotated pose (azimuth angle $\theta = -30^\circ$, elevation angle $\beta = 20^\circ$) synthesized from the images at the first row, and the third row showing the ground truth of the basis images at the rotated pose generated from the 3D scan. Rows four through six present the results for subject 2, with the fourth row showing the basis images at the frontal pose generated from the 3D scan, the fifth row showing the basis images for another rotated view (azimuth angle $\theta = -30^\circ$, elevation angle $\beta = -20^\circ$) synthesized from the images at the fourth row, and the last row showing the ground truth of the basis images at the rotated pose generated from the 3D scan. As we can see from Figure 2, the synthesized basis images at the rotated poses are very close to the ground truth. Note in Figure 2 and the figures in the sequel the dark regions represent the negative values of the basis images.

Given that the correspondence between the rotated-pose image and the frontal-pose image is available, a consequence of the existence of such linear transformation is that the procedure of first rotating objects and then recomputing basis images at the desired pose can be avoided. The block diagonal form of the transformation matrices preserves the energy on each order $l = 0, 1, 2$. Moreover, the orthonormality of the transformation matrices helps to further simplify the computation required for the recognition of the rotated test image as shown in Section 4.2. Although in theory new basis images can be generated from a rotated 3D model inferred by the existing basis images (since basis images actually capture the albedo (b_{00}) and the 3D surface normal ($b_{10}, b^e_{11}, b^o_{11}$) of a given human face), the procedure of such 3D recovery is not trivial in practice, even if computational cost is taken out of consideration.

4. FACE RECOGNITION USING POSE-ENCODED SPHERICAL HARMONICS

In this section, we present an efficient face recognition method using pose-encoded spherical harmonics. Only one training image is needed per subject and high recognition performance is achieved even when the test image is at a different pose from the training image and under an arbitrary illumination condition.

4.1. Statistical models of basis images

We briefly summarize a statistical learning method to recover the harmonic basis images from only one image taken under arbitrary illumination conditions, as shown in [9].

We build a bootstrap set with fifty 3D face scans and corresponding texture maps from Vetter's 3D face database [10], and generate nine basis images for each face model. For a novel N-dimensional vectorized image I, let B be the $N \times 9$ matrix of basis images, α, a 9-dimensional vector, and e, an N-dimensional error term. We have $I = B\alpha + e$. It is assumed that the probability density functions (pdf's) of B are Gaussian distributions. The sample mean vectors $\mu_b(x)$ and covariance matrixes $C_b(x)$ are estimated from the basis images in the bootstrap set. Figure 3 shows the sample mean of the basis images estimated from the bootstrap set.

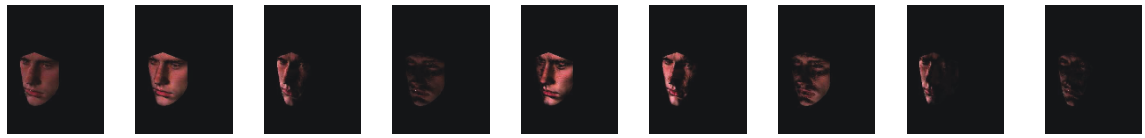
By estimating α and the statistics of $E(\alpha)$ in a prior step with kernel regression and using them consistently across all pixels to recover B, it is shown in [9] that for a given novel face image $i(x)$, the corresponding basis images $b(x)$ at each pixel x are recovered by computing the maximum a posteriori (MAP) estimate, $b_{\mathrm{MAP}}(x) = \arg_{b(x)}\max(P(b(x) \mid i(x)))$. Using the Bayes rule,

$$\begin{aligned} b_{\mathrm{MAP}}(x) &= \arg\max_{b(x)} P(i(x) \mid b(x)) P(b(x)) \\ &= \arg\max_{b(x)} \left\{ \mathcal{N}\left(b(x)^T\alpha + \mu_e, \sigma_e^2\right) \mathcal{N}\left(\mu_b(x), C_b(x)\right) \right\}. \end{aligned} \tag{5}$$

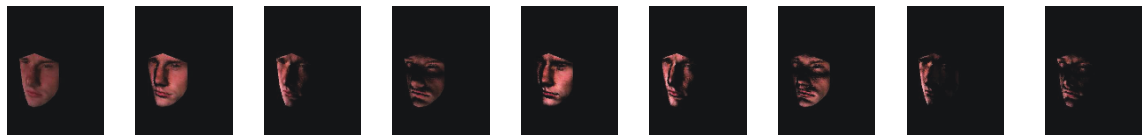
Taking logarithm, and setting the derivatives of the right-hand side of (5) (w.r.t. $b(x)$) to 0, we get $A * b_{\mathrm{MAP}} = U$, where $A = (1/\sigma_e^2)\alpha\alpha^T + C_b^{-1}$ and $U = ((i - \mu_e)/\sigma_e^2)\alpha + C_b^{-1}\mu_b$. Note that the superscript $(\cdot)^T$ denotes the transpose of the matrix here and in the sequel. By solving this linear equation, $b(x)$ of the subject can be recovered.

In Figure 4, we illustrate the procedure for generating the basis images at a rotated pose (azimuth angle $\theta = -30^\circ$)

(a) Subject 1: the basis images at the frontal pose generated from the 3D scan

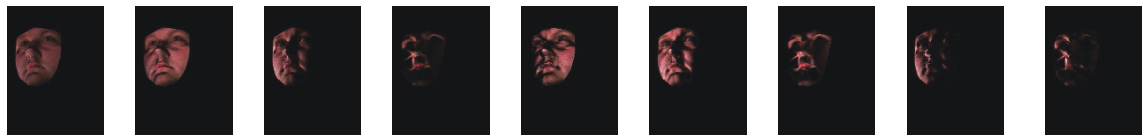


(b) Subject 1: the basis images at the rotated pose synthesized from (a)

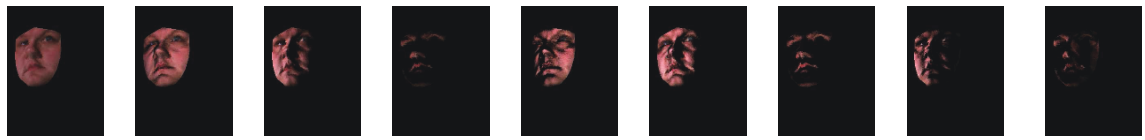


(c) Subject 1: the ground truth of the basis images at the rotated pose generated from the 3D scan

(d) Subject 2: the basis images at the frontal pose generated from the 3D scan



(e) Subject 2: the basis images at the rotated pose synthesized from (d)



(f) Subject 2: the ground truth of the basis images at the rotated pose generated from the 3D scan

Figure 2: (a)–(c) present the results of the synthesized basis images for subject 1, where (a) shows the basis images at the frontal pose generated from the 3D scan, (b) the basis images at a rotated pose synthesized from (a), and (c) the ground truth of the basis images at the rotated pose. (d)-(e) present the results of the synthesized basis images for subject 2, with (d) showing the basis images at the frontal pose generated from the 3D scan, (e) the basis images at a rotated pose synthesized from (d), and (f) the ground truth of the basis images at the rotated pose.

b_{00} b_{10} b_{11}^{e} b_{11}^{o} b_{20} b_{21}^{e} b_{21}^{o} b_{22}^{e} b_{22}^{o}

Figure 3: The sample mean of the basis images estimated from the bootstrap set [10].

from a single training image at the frontal pose. In Figure 4, rows one through three show the results of the recovered basis images from a single training image, with the first column showing different training images I under arbitrary illumination conditions for the same subject and the remaining nine columns showing the recovered basis images. We can observe from the figure that the basis images recovered from different training images of the same subject look very similar. Using the basis images recovered from any training image in row one through three, we can synthesize basis images at

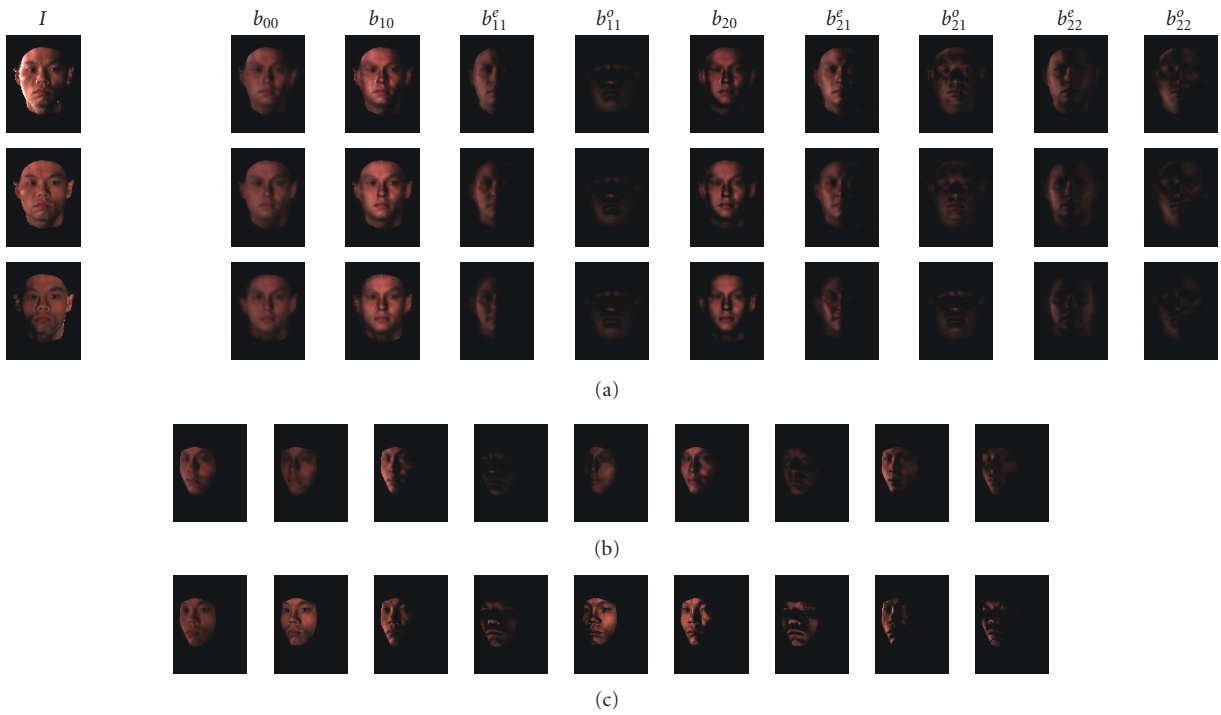


Figure 4: The first column in (a) shows different training images I under arbitrary illumination conditions for the same subject and the remaining nine columns in (a) show the recovered basis images from I. We can observe that the basis images recovered from different training images of the same subject look very similar. Using the basis images recovered from any training image I in (a), we can synthesize basis images at the rotated pose, as shown in (b). As a comparison, (c) shows the ground truth of the basis images at the rotated pose generated from the 3D scan.

the rotated pose, as shown in row four. As a comparison, the fifth row shows the ground truth of the basis images at the rotated pose generated from the 3D scan.

For the CMU-PIE [11] database, we used the images of each subject at the frontal pose (c27) as the training set. One hundred 3D face models from Vetter's database [10] were used as the bootstrap set. The training images were first rescaled to the size of the images in the bootstrap set. The statistics of the harmonic basis images was then learnt from the bootstrap set and the basis images B for each training subject were recovered. Figure 5 shows two examples of the recovered basis images from the single training image, with the first column showing the training images I and the remaining 9 columns showing the reconstructed basis images.

4.2. Recognition

For recognition, we follow a simple yet effective algorithm given in [4]. A face is identified for which there exists a weighted combination of basis images that is the closest to the test image. Let B be the set of basis images at the frontal pose, with size $N \times v$, where N is the number of pixels in the image and $v = 9$ is the number of basis images used. Every column of B contains one spherical harmonic image. These images form a basis for the linear subspace, though not an orthonormal one. A QR decomposition is applied to compute Q, an $N \times v$ matrix with orthonormal columns, such that $B = QR$, where R is a $v \times v$ upper triangular matrix.

For a vectorized test image I_{test} at an arbitrary pose, let B_{test} be the set of basis images at that pose. The orthonormal basis Q_{test} of the space spanned by B_{test} can be computed by QR decomposition. The matching score is defined as the distance from I_{test} to the space spanned by B_{test} : $s_{\text{test}} = \|Q_{\text{test}}Q_{\text{test}}^T I_{\text{test}} - I_{\text{test}}\|$. However, this algorithm is not efficient to handle pose variation because the set of basis images B_{test} has to be generated for each subject at the arbitrary pose of a test image.

We propose to warp the test image I_{test} at the arbitrary (rotated) pose to its frontal view image I_f to perform recognition. In order to warp I_{test} to I_f, we have to find the point correspondence between these two images, which can be embedded in a sparse $N \times N$ warping matrix K, that is, $I_f = KI_{\text{test}}$. The positions of the nonzero elements in K encode the 1-to-1 and many-to-1 correspondence cases (the 1-to-many case is same as 1-to-1 case for pixels in I_f) between I_{test} and I_f, and the positions of zeros on the diagonal line of K encode the no-correspondence case. More specifically, if pixel $I_f(i)$ (the ith element in vector I_f) corresponds to pixel $I_{\text{test}}(j)$ (the jth element in vector I_{test}), then $K(i, j) = 1$. There might be cases that there are more than one pixel in I_{test} corresponding to the same pixel $I_f(i)$, that is, there are more than one 1 in the ith row of K, and the column indices of these 1's are the corresponding pixel indices in I_{test}. For this case, although there are several pixels in I_{test} mapping to the same pixel $I_f(i)$, it can only have one reasonable intensity value. We compute a single "virtual" corresponding pixel in

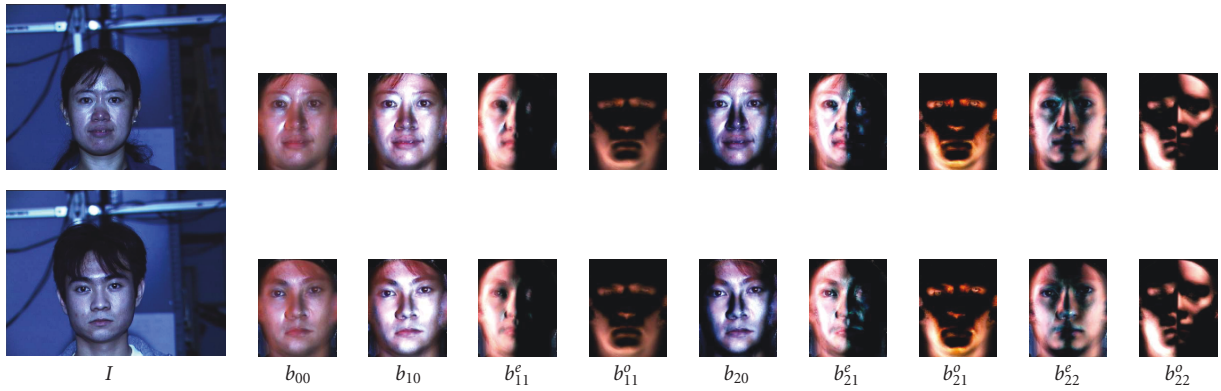


FIGURE 5: The first column shows the training images I for two subjects in the CMU-PIE database and the remaining nine columns show the reconstructed basis images.

I_{test} for $I_f(i)$ as the centroid of $I_f(i)$'s real corresponding pixels in I_{test}, and assign it the average intensity. The weight for each real corresponding pixel $I_{\text{test}}(j)$ is proportional to the inverse of its distance to the centroid, and this weight is assigned as the value of $K(i, j)$. If there is no correspondence in I_{test} for $I_f(i)$ which is in the valid facial area and should have a corresponding point in I_{test}, it means that $K(i, i) = 0$. This is often the case that the corresponding "pixel" of $I_f(i)$ falls in the subpixel region. Thus, interpolation is needed to fill the intensity for $I_f(i)$. Barycentric coordinates [26] are calculated with the pixels which have real corresponding integer pixels in I_{test} as the triangle vertices. These Barycentric coordinates are assigned as the values of $K(i, j)$, where j is the column index for each vertex of the triangle.

We now have the warping matrix K which encodes the correspondence and interpolation information in order to generate I_f from I_{test}. It provides a very convenient tool to analyze the impact of some empirical factors in image warping. Note that due to self-occlusion, I_f does not cover the whole area, but only a subregion, of the full frontal face of the subject it belongs to. The missing facial region due to the rotated pose is filled with zeros in I_f. Assume that B_f is the basis images for the full frontal view training images and Q_f is its orthonormal basis, and let b_f be the corresponding basis images of I_f and q_f its orthonormal basis. In b_f, the rows corresponding to the valid facial pixels in I_f form a submatrix of the rows in B_f corresponding to the valid facial pixels in the full frontal face images. For recognition, we cannot directly use the orthonormal columns in Q_f because it is not guaranteed that all the columns in q_f are still orthonormal.

We study the relationship between the matching score for the rotated view $s_{\text{test}} = \|Q_{\text{test}}Q_{\text{test}}^T I_{\text{test}} - I_{\text{test}}\|$ and the matching score for the frontal view $s_f = \|q_f q_f^T I_f - I_f\|$. Let subject a be the one that has the minimum matching score at the rotated pose, that is, $s^a_{\text{test}} = \|Q^a_{\text{test}}{Q^a_{\text{test}}}^T I_{\text{test}} - I_{\text{test}}\| \le s^c_{\text{test}} = \|Q^c_{\text{test}}{Q^c_{\text{test}}}^T I_{\text{test}} - I_{\text{test}}\|$, for all $c \in [1, 2, \ldots, C]$, where C is the number of training subjects. If a is the correct subject for the test image I_{test}, warping Q^a_{test} to q^a_f undertakes the same warping matrix K as warping I_{test} to I_f, that is, the matching score for the frontal view $s^a_f = \|q^a_f {q^a_f}^T I_f - I_f\| = \|KQ^a_{\text{test}}{Q^a_{\text{test}}}^T K^T K I_{\text{test}} - KI_{\text{test}}\|$. Note here that we only consider the correspondence and interpolation issues. Due to the orthonormality of the transformation matrices as shown in (3) and (4), the linear transformation from B_{test} to b_f does not affect the matching score. For all the other subjects $c \in [1, 2, \ldots C], c \neq a$, the warping matrix K^c for Q^c_{test} is different from that for I_{test}, that is, $s^c_f = \|K^c Q^c_{\text{test}}{Q^c_{\text{test}}}^T {K^c}^T K I_{\text{test}} - KI_{\text{test}}\|$. We will show that warping I_{test} to I_f does not deteriorate the recognition performance, that is, given $s^a_{\text{test}} \le s^c_{\text{test}}$, we have $s^a_f \le s^c_f$.

In terms of K, we consider the following cases.

Case 1. $K = \begin{pmatrix} E_k & 0 \\ 0 & 0 \end{pmatrix}$, where E_k is the k-rank identity matrix. It means that K is a diagonal matrix and the first k elements on the diagonal line are 1, all the rest are zeros.

This is the case when I_{test} is at the frontal pose. The difference between I_{test} and I_f is that there are some missing (nonvalid) facial pixels in I_f than in I_{test}, and all the valid facial pixels in I_f are packed in the first k elements. Since I_{test} and I_f are at the same pose, Q_{test} and q_f are also at the same pose. In this case, for subject a, the missing (nonvalid) facial pixels in q_f are at the same locations as in I_f since they have the same warping matrix K. On the other hand, for any other subject c, the missing (nonvalid) facial pixels in q_f are not at the same locations as in I_f since $K^c \neq K$. Apparently the 0's and 1's on the diagonal line of K^c has different positions from that of K, thus $K^c K$ has more 0's on the diagonal line than K.

Assume $K = \begin{pmatrix} E_k & 0 \\ 0 & 0 \end{pmatrix}$ and $V = Q_{\text{test}}Q_{\text{test}}^T = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$, where V_{11} is a $(k \times k)$ matrix. Similarly, let $I_{\text{test}} = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$, where I_1 is a $(k \times 1)$ vector. Then $KQ_{\text{test}}Q_{\text{test}}^T K^T = \begin{pmatrix} V_{11} & 0 \\ 0 & 0 \end{pmatrix}$, $KI_{\text{test}} = \begin{pmatrix} I_1 \\ 0 \end{pmatrix}$, and $KQ_{\text{test}}Q_{\text{test}}^T K^T K I_{\text{test}} - KI_{\text{test}} = \begin{pmatrix} V_{11}I_1 \\ 0 \end{pmatrix} - \begin{pmatrix} I_1 \\ 0 \end{pmatrix} = \begin{pmatrix} (V_{11} - E_k)I_1 \\ 0 \end{pmatrix}$. Therefore, $s^a_f = \|(V_{11} - E_k)I_1\|$. Similarly, $K^c Q_{\text{test}}Q_{\text{test}}^T K^{cT} = \begin{pmatrix} V^c_{11} & 0 \\ 0 & 0 \end{pmatrix}$, where V^c_{11} is also a $(k \times k)$ matrix that might contain rows with all 0's, depending on the locations of the 0's on the diagonal line of K^c. We have $K^c Q_{\text{test}}Q_{\text{test}}^T K^{cT} K I_{\text{test}} - KI_{\text{test}} = \begin{pmatrix} V^c_{11}I_1 \\ 0 \end{pmatrix} - \begin{pmatrix} I_1 \\ 0 \end{pmatrix} = \begin{pmatrix} (V^c_{11} - E_k)I_1 \\ 0 \end{pmatrix}$. Thus, $s^c_f = \|(V^c_{11} - E_k)I_1\|$.

If V^c_{11} has rows with all 0's in the first k rows, these rows will have -1's at the diagonal positions for $V^c_{11} - E_k$, which will increase the matching score s^c_f. Therefore, $s^a_f \le s^c_f$.

Table 1

Pose	$(\theta = 30^\circ, \beta = 0^\circ)$	$(\theta = 30^\circ, \beta = -20^\circ)$	$(\theta = -30^\circ, \beta = 0^\circ)$	$(\theta = -30^\circ, \beta = 20^\circ)$
mean $((s_f - s_{\text{test}})/s_{\text{test}})$	3.4%	3.9%	3.5%	4.1%
std $((s_f - s_{\text{test}})/s_{\text{test}})$	5.0%	5.2%	4.9%	5.1%

Case 2. K is a diagonal matrix with rank k, however, the k 1's are not necessarily the first k elements on the diagonal line.

We can use some elementary transformation to reduce this case to the previous case. That is, there exists a orthonormal matrix P, such that $\hat{K} = PKP^T = \begin{pmatrix} E_k & 0 \\ 0 & 0 \end{pmatrix}$.

Let $\hat{Q}_{\text{test}} = PQ_{\text{test}}P^T$ and $\hat{I}_{\text{test}} = PI_{\text{test}}$. Then

$$\begin{aligned} s_f^a &= ||P(KQ_{\text{test}}Q_{\text{test}}^T K^T K I_{\text{test}} - KI_{\text{test}})|| \\ &= ||\hat{K}\hat{Q}_{\text{test}}\hat{Q}_{\text{test}}^T\hat{K}^T\hat{K}\hat{I}_{\text{test}} - \hat{K}\hat{I}_{\text{test}}||. \end{aligned} \tag{6}$$

Note that elementary transformation does not change the norm. Hence, it reduces to the previous case. Similarly, we have that s_f^c stays the same as in Case 1. Therefore, $s_f^a \le s_f^c$ still holds.

In the general case, 1's in K can be off-diagonal. This means that I_{test} and I_f are at different poses. There are three subcases that we need to discuss for a general K.

Case 3. 1-to-1 correspondence between I_{test} and I_f. If pixel $I_{\text{test}}(j)$ has only one corresponding point in I_f, denoted as $I_f(i)$, then $K(i, j) = 1$ and there are no 1's in both the ith row and the jth column in K. Suppose there are only k columns of the matrix K containing 1. Then, by appropriate elementary transformation again, we can left multiply and right multiply K by an orthonormal transformation matrixes, W and V, respectively, such that $\tilde{K} = WKV$. If we define $\tilde{Q}_{\text{test}} = V^T Q_{\text{test}} W$ and $\tilde{I}_{\text{test}} = V^T I_{\text{test}}$, then

$$\begin{aligned} s_f^a &= ||KQ_{\text{test}}Q_{\text{test}}^T K^T K I_{\text{test}} - KI_{\text{test}}|| \\ &= ||W(KQ_{\text{test}}Q_{\text{test}}^T K^T K I_{\text{test}} - KI_{\text{test}})|| \\ &= ||WKVV^T Q_{\text{test}} W W^T Q_{\text{test}}^T V V^T K^T W^T W K V (V^T I_{\text{test}}) \\ &\quad - WKV(V^T I_{\text{test}})|| \\ &= ||\tilde{K}\tilde{Q}_{\text{test}}\tilde{Q}_{\text{test}}^T\tilde{K}^T\tilde{K}\tilde{I}_{\text{test}} - \tilde{K}\tilde{I}_{\text{test}}||. \end{aligned} \tag{7}$$

Under $\tilde{K}$, it reduces to Case 2, which can be further reduced to Case 1 by the aforementioned technique. Similarly, we have that s_f^c stays the same as in Case 2. Therefore, $s_f^a \le s_f^c$ still holds.

In all the cases discussed up to now, the correspondence between I_{test} and I_f is 1-to-1 mapping. For such cases, the following lemma shows that the matching score stays the same before and after the warping.

Lemma 1. *Given the correspondence between a rotated test image I_{test} and its geometrically synthesized frontal view image I_f is 1-to-1 mapping, the matching score s_{test} of I_{test} based on the basis images B_{test} at that pose is the same as the matching score s_f of I_f based on the basis images b_f.*

Let O be the transpose of the combined coefficient matrices in (3) and (4), we have $b_f = KB_{\text{test}}O = Q_{\text{test}}RO$ by QR decomposition, where K is the warping matrix from I_{test} to I_f with only 1-to-1 mapping. Applying QR decomposition again to RO, we have $RO = \tilde{q}\tilde{r}$, where $\tilde{q}_{v\times v}$ is an orthonormal matrix and $\tilde{r}$ is an upper triangular matrix. We now have $b_f = KQ_{\text{test}}\tilde{q}\tilde{r} = q_f\tilde{r}$ with $q_f = KQ_{\text{test}}\tilde{q}$. Since $Q_{\text{test}}\tilde{q}$ is the product of two orthonormal matrices, q_f forms a valid orthnormal basis for b_f. Hence the matching score is $s_f = \|q_f q_f^T I_f - I_f\| = \|KQ_{\text{test}}\tilde{q}\tilde{q}^T Q_{\text{test}}^T K^T K I_{\text{test}} - KI_{\text{test}}\| = \|Q_{\text{test}}Q_{\text{test}}^T I_{\text{test}} - I_{\text{test}}\| = s_{\text{test}}$.

If the correspondence between I_{test} and I_f is not 1-to-1 mapping, we have the following two cases.

Case 4. Many-to-1 correspondence between I_{test} and I_f.

Case 5. There is no correspondence for $I_f(i)$ in I_{test}.

For Cases 4 and 5, since the 1-to-1 correspondence assumption does not hold any more, the relationship between s_{test} and s_f is more complex. This is due to the effects of fortshortening and interpolation. Fortshortening leads to more contributions for the rotated view recognition but less in the frontal view recognition (or vice versa) because of the fortshortening. The increased (or decreased) information due to interpolation, and the assigned weight for each interpolated pixel, is not guaranteed to be the same as that before the warping. Therefore, the relationship between s_{test} and s_f relies on each specific K, which may vary significantly depending on the variation of the head pose. Instead of theoretical analysis, the empirical error bound between s_{test} and s_f is sought to give a general idea of how the warping affects the matching scores. We conducted experiments using Vetter's database. For the fifty subjects which are not used in the bootstrap set, we generated images at various poses and obtained their basis images at each pose. For each pose, s_{test} and s_f are compared, and the mean of the relative error and the relative standard deviation for some poses are listed in Table 1.

We can see from the experimental results although s_{test} and s_f are not exactly the same that the difference between s_{test} and s_f is very small. We examined the ranking of the matching scores before and after the warping. Table 2 shows the percentage that the top one pick before the warping still remains as the top one after the warping.

Thus, warping the test image I_{test} to its frontal view image I_f does not reduce the recognition performance. We now have a very efficient solution for face recognition to handle both pose and illumination variations as only one image I_f needs to be synthesized.

Now, the only remaining problem is that the correspondence between I_{test} and I_f has to be built. Although a

Table 2

Pose	$(\theta = 30^\circ, \beta = 0^\circ)$	$(\theta = 30^\circ, \beta = -20^\circ)$	$(\theta = -30^\circ, \beta = 0^\circ)$	$(\theta = -30^\circ, \beta = 20^\circ)$
percentage of the top one pick keeps its position	98.4%	97.6%	99.2%	97.9%

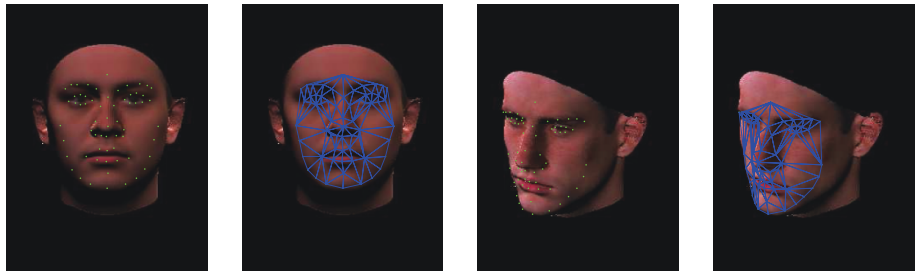


Figure 6: Building dense correspondence between the rotated view and the frontal view using sparse features. The first and second images show the sparse features and the constructed meshes on the mean face at the frontal pose. The third and fourth images show the picked features and the constructed meshes on the given test image at the rotated pose.

necessary component of the system, finding correspondence is not the main focus of this paper. Like most of the approaches to handle pose variations, we adopt the method to use sparse main facial features to build the dense cross-pose or cross-subject correspondence [9]. Some automatic facial feature detection/selection techniques are available, but most of them are not robust enough to reliably detect the facial features from images at arbitrary poses and are taken under arbitrary lighting conditions. For now, we manually pick sixty three designated feature points (eyebrows, eyes, nose, mouth, and the face contour) on I_{test} at the arbitrary pose. An average face calculated from training images at the frontal pose and the corresponding feature points were used to help to build the correspondence between I_{test} and I_f. Triangular meshes on both faces were constructed and barycentric interpolation inside each triangle was used to find the dense correspondence, as shown in Figure 6. The number of feature points needed in our approach is comparable to the 56 manually picked feature points in [9] to deform the 3D model.

4.3. View synthesis

To verify the recognition results, the user is given the option to visually compare the chosen subject and the test image I_{test} by generating the face image of the chosen subject at the same pose and under the same illumination condition as I_{test}. The desired N-dimensional vectorized image I_{des} can be synthesized easily as long as we can generate the basis images B_{des} of the chosen subject at that pose by using $I_{\text{des}} = B_{\text{des}}\alpha_{\text{test}}$. Assuming that the correspondence between I_{test} and the frontal pose image has been built as described in Section 4.2, then B_{des} can be generated from the basis images B of the chosen subject at the frontal pose using (3) and (4), given that the pose (θ, β) of I_{test} can be estimated as described later. We also need to estimate the 9-dimensional lighting coefficient vector α_{test}. Assuming that the chosen subject is the correct one, that is, $B_{\text{test}} = B_{\text{des}}$, we have $I_{\text{test}} = B_{\text{des}}\alpha_{\text{test}}$ by substituting $B_{\text{test}} = B_{\text{des}}$ into $I_{\text{test}} = B_{\text{test}}\alpha_{\text{test}}$. Recalling that $B_{\text{des}} = Q_{\text{des}}R_{\text{des}}$, we have $I_{\text{test}} = Q_{\text{des}}R_{\text{des}}\alpha_{\text{test}}$ and then $Q_{\text{des}}^T I_{\text{test}} = Q_{\text{des}}^T Q_{\text{des}}R_{\text{des}}\alpha_{\text{test}} = R_{\text{des}}\alpha_{\text{test}}$ due to the orthonormality of Q_{des}. Therefore, $\alpha_{\text{test}} = R_{\text{des}}^{-1}Q_{\text{des}}^T I_{\text{test}}$.

Having both B_{des} and α_{test} available, we are ready to generate the face image of the chosen subject at the same pose and under the same illumination condition as I_{test} using $I_{\text{des}} = B_{\text{des}}\alpha_{\text{test}}$. The only unknown to be estimated is the pose (θ, β) of I_{test}, which is needed in (3) and (4).

Estimating head pose from a single face image is an active research topic in computer vision. Either a generic 3D face model or several main facial features are utilized to estimate the head pose. Since we already have the feature points to build the correspondence across views, it is natural to use these feature points for pose estimation. In [27], five main facial feature points (four eye corners and the tip of the nose) are used to estimate the 3D head orientation. The approach employs the projective invariance of the cross-ratios of the eye corners and anthropometric statistics to determine the head yaw, roll and pitch angles. The focal length f has to be assumed known, which is not always available for the uncontrollable test image. We take the advantage that the facial features on the frontal view mean face are available, and show how to estimate the head pose without knowing f. All notations follow those in [27].

Let (u_2, u_1, v_1, v_2) be the image coordinates of the four eye corners, and D and D_1 denote the width of the eyes and half of the distance between the two inner eye corners, respectively. From the well known projective invariance of the cross ratios we have $J = (u_2 - u_1)(v_1 - v_2)/(u_2 - v_1)(u_1 - v_2) = D^2/(2D_1 + D)^2$ which yields $D_1 = DQ/2$, where $Q = 1/\sqrt{J} - 1$. In order to recover the yaw angle θ (around the Y-axis), it is easy to have, as shown in [27], that $\theta = \arctan(f/(S+1)u_1)$, where f is the focal length and S is the solution to the equation $\Delta u/\Delta v = -(S-1)(S-(1+2/Q))/(S+1)(S+1+2/Q))$, where $\Delta u = u_2 - u_1$ and $\Delta v = v_1 - v_2$. Assume that u_1^f is the inner corner of one of the eyes for the frontal view mean

Table 3: The mean and standard deviation (std) of the estimated pose for images from the Vetter's database.

Rotation angles	$(\theta = 30^\circ, \beta = 0^\circ)$	$(\theta = 30^\circ, \beta = -20^\circ)$	$(\theta = -30^\circ, \beta = 0^\circ)$	$(\theta = -30^\circ, \beta = 20^\circ)$
Mean of the estimated pose	$(\theta = 28^\circ, \beta = 2^\circ)$	$(\theta = 31^\circ, \beta = -23^\circ)$	$(\theta = -32^\circ, \beta = 1^\circ)$	$(\theta = -33^\circ, \beta = 22^\circ)$
std of the estimated pose	$(3.2^\circ, 3.1^\circ)$	$(3.9^\circ, 4.2^\circ)$	$(3.4^\circ, 2.7^\circ)$	$(4.2^\circ, 4.5^\circ)$

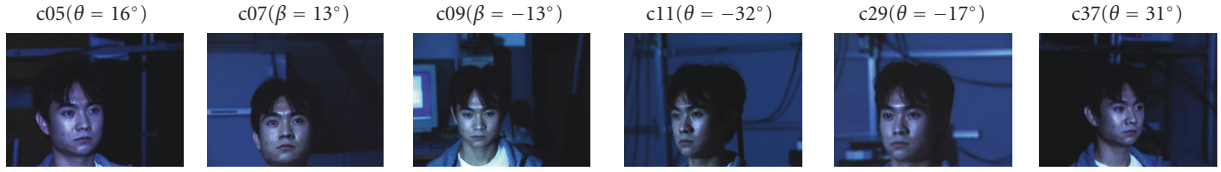


Figure 7: An illustration of the pose variation in part of the CMU-PIE database, with the ground truth of the pose shown beside each pose index. Four of the cameras (c05, c11, c29, and c37) sweep horizontally, and the other two are above (c09) and below (c07) the central camera, respectively.

face. With perspective projection, we have $u_1^f = fD_1/Z$ and $u_1 = fX_1/(Z + Z_1) = fD_1 \cos\theta/(Z + D_1 \sin\theta)$. Thus,

$$f = (S+1)u_1 \tan\theta. \tag{8}$$

Then we have $S = (u_1/u_1^f)((S+1)/\cos\theta)$, which gives

$$\theta = \arccos \frac{(S+1)}{S} \frac{u_1}{u_1^f}. \tag{9}$$

In [27], β (the rotation angle around the x-axis) is shown to be $\beta = \arcsin(E)$ with $E = (f/p_0(p_1^2 + f^2))[p_1^2 \pm \sqrt{(p_0^2 p_1^2 - f^2 p_1^2 + f^2 p_0^2)}]$, where p_0 denotes the projected length of the bridge of the nose when it is parallel to the image plane, and p_1 denotes the observed length of the bridge of the nose at the unknown pitch β. Anthropometric statistics is employed in [27] to get p_0. With the facial features on the mean face at the frontal view available, we do not need the anthropometric statistics. p_0 is just the length between the upper midpoint of the nose and the tip of the nose for the frontal view mean face. So we can directly use this value and the estimated focal length f in (8) to get the pitch angle β.

The head pose estimation algorithm is tested on both synthetic and real images. For synthetic images, we use Vetter's 3D face database. The 3D face model for each subject is rotated to the desired angle and project to the 2D image plane. Four eye corners and the tip of the nose are used to estimate the head pose. The mean and standard deviation of the estimated poses are listed in Table 3. For real images, we use the CMU-PIE database. The ground truth of the head pose can be obtained from the available 3D locations of the head and the cameras. The experiments are conducted for all 68 subjects in the CMU-PIE database at six different poses, illustrated in Figure 7 with the ground truth of the pose shown beside each pose index. The mean and standard deviation of the estimated poses are listed in Table 4. Overall the pose estimation results are satisfying and we believe that the relatively large standard deviation is due to the error in selecting the facial features. The mean and standard deviation (std) of the estimated pose for images from the Vetter's database.

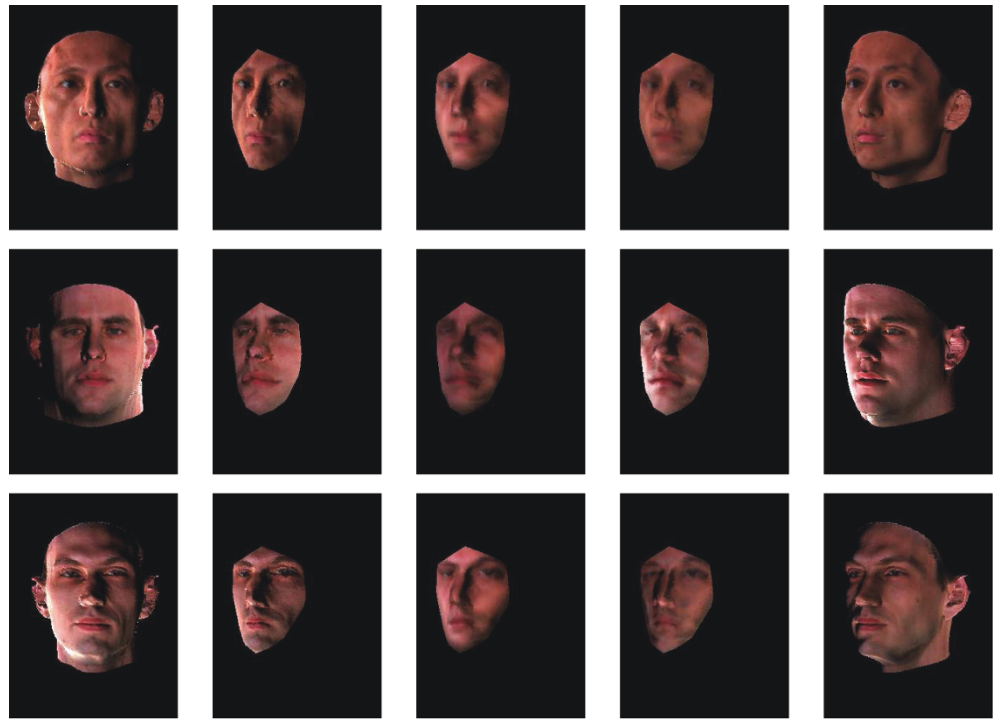
Having the head pose estimated, we can now perform the face synthesis. Figure 8 shows the comparison of the given test image I_{test} and some synthesized face images at the same pose as I_{test} from the chosen subject, where Figure 8(a) is for the synthetic images in Vetter's 3D database and Figure 8(b) is for the real images in the CMU-PIE database. Column one shows the training images. Column two shows the synthesized images at the same pose as I_{test} by direct warping. Column three shows the synthesized images using the basis images B_{des} from the chosen subject and the illumination coefficients α_{tr} of the training images. A noticeable difference between column two and three is the lighting change. By direct warping, we obtain the synthesized images by not only rotating the head pose, but also rotating the lighting direction at the same time. By using α_{tr}, we only rotate the head pose to get the synthesized images, while the lighting condition stays same as the training images. Column four shows the synthesized images using the basis images B_{des} from the chosen subject and the same illumination coefficients α_{test} of I_{test}. As a comparison, column five shows the given test image I_{test}. Overall, the columns from left to right in Figure 8 show the procedure migrating from the training images to the given test images.

5. RECOGNITION RESULTS

We first conducted recognition experiments on Vetter's 3D face model database. There are totally one hundred 3D face models in the database, from which fifty were used as the bootstrap set and the other fifty were used to generate training images. We synthesized the training images under a wide variety of illumination conditions using the 3D scans of the subjects. For each subject, only one frontal view image was stored as the training image and used to recover the basis images B using the algorithm in Section 4.1. We generated the test images at different poses by rotating the 3D scans and illuminated them with various lighting conditions (represented by the slant angle γ and tilt angle τ). Some examples are shown in Figures 9(a), 9(b), 9(c) and 9(d). For a test image I_{test} at an arbitrary pose, the frontal view image I_f was synthesized by warping I_{test}, as shown in Figures 9(e), 9(f), 9(g) and 9(h).

Table 4: The mean and standard deviation (std) of the estimated pose for images from the CMU-PIE database.

Pose index	c05	c07	c09	c11	c29	c37
Mean of the estimated pose	$\theta = 15^\circ$	$\beta = 11^\circ$	$\beta = -15^\circ$	$\theta = -36^\circ$	$\theta = -17^\circ$	$\theta = 35^\circ$
std of the estimated pose	4.1°	3.8°	4.0°	6.2°	3.3°	5.4°



(a)

(b)

Figure 8: View synthesis results with different lighting conditions for (a) synthetic images from Vetter's 3D database and (b) real images in the CMU-PIE database. Columns from left to right show the training images, the synthesized images at the same pose as the test images using direct warping (both the head pose and the lighting direction are rotated), the synthesized images at the same pose as the test images from B_{des} (the basis images of the chosen subject) and α_{tr} (the illumination coefficients of the training images), the synthesized images at the same pose as the test images from B_{des} and α_{test} (the illumination coefficients of the given test images), and the given test images I_{test}.

FIGURE 9: (a) shows the test images of a subject at azimuth $\theta = -30^\circ$ under different lighting conditions ($\gamma = 90^\circ, \tau = 10^\circ$; $\gamma = 30^\circ, \tau = 50^\circ$; $\gamma = 40^\circ, \tau = -10^\circ$; $\gamma = 20^\circ, \tau = 70^\circ$; $\gamma = 80^\circ, \tau = -20^\circ$; $\gamma = 50^\circ, \tau = 30^\circ$ from left to right). The test images of the same subject under some extreme lighting conditions ($\gamma = 20^\circ, \tau = -70^\circ$; $\gamma = 20^\circ, \tau = 70^\circ$; $\gamma = 120^\circ, \tau = -70^\circ$; $\gamma = 120^\circ, \tau = -70^\circ$ from left to right) are shown in (b). (c) and (d) show the generated frontal pose images from the test images in (a) and (b), respectively. The test images at another pose (with $\theta = -30^\circ$ and $\beta = 20^\circ$) of the same subject are shown in (e) and (f), with the generated frontal pose images shown in (g) and (h), respectively.

The recognition score was computed as $\|q_f q_f^T I_f - I_f\|$ where q_f is the orthonormal basis of the space spanned by b_f. As a benchmark, the first column (f2f) of Table 5 lists the recognition rates when both the testing images and the training images are from the frontal view. The correct recognition rates using the proposed method are listed in columns (r2f) of Table 5. As a comparison, we also conducted the recognition experiment on the same test images assuming that the training images at the same pose are available. By recovering the basis images B_{test} at that pose using the algorithm in Section 4.1 and computing $\|Q_{\text{test}} Q_{\text{test}}^T I_{\text{test}} - I_{\text{test}}\|$, we achieved the recognition rates as shown in columns (r2r) of Table 5. As we can see, the recognition rates using our approach (r2f) are comparable to those when the training images at the rotated pose are available (r2r). The last two rows of show the mean and standard deviation of the recognition rates for each pose under various illumination conditions. We believe that relatively larger standard deviation is due to the images under some extreme lighting conditions, as shown in Figures 9(b) and 9(f).

We also conducted experiments on real images from the CMU-PIE database. For testing, we used images at six different poses, as shown in the first and third rows in Figure 10, and under twenty one different illuminations. Examples of the generated frontal view images are shown in the second and fourth rows of Figure 10.

Similar to Table 5,Table 6 lists the correct recognition rates under all these poses and illumination conditions, where column (f2f) is the frontal view testing image against frontal view training images, columns (r2r) are the rotated testing image against the same pose training images, and columns (r2f) are the rotated testing image against the frontal view training images. The last two rows of Table 6 show the mean and standard deviation of the recognition rates for each pose under various illumination conditions. As we can see, the recognition rates using our approach are comparable to those when the training images at the rotated pose are available, even *slightly* better. The reason is that the training images of different subjects at the *same* rotated pose are actually at *slightly different* poses. Therefore, the 2D-3D registration of the training images and the bootstrap 3D face models are not perfect, producing *slightly worse* basis images recovery than the frontal pose case.

We have to mention that although colored basis images are recovered for visualization purpose, all the recognition experiments are performed on grayscale images for faster speed. We are taking the efforts to investigate how color information affects the recognition performance.

6. DISCUSSIONS AND CONCLUSION

We have presented an efficient face synthesis and recognition method to handle arbitrary pose and illumination from a single training image per subject using pose-encoded spherical harmonics. Using a prebuilt 3D face bootstrap set, we apply a statistical learning method to obtain the spherical

Table 5: The correct recognition rates at two rotated pose under various lighting conditions for synthetic images generated from Vetter's 3D face model database.

Lighting/pose	f2f	Pose $\theta = -30^\circ$ and $\beta = 0^\circ$		Pose $\theta = -30^\circ$ and $\beta = 20^\circ$	
		r2f	r2r	r2f	r2r
$(\gamma = 90^\circ, \tau = 10^\circ)$	100	100	96	84	80
$(\gamma = 30^\circ, \tau = 50^\circ)$	100	100	100	100	100
$(\gamma = 40^\circ, \tau = -10^\circ)$	100	100	100	100	100
$(\gamma = 70^\circ, \tau = 40^\circ)$	100	100	100	94	88
$(\gamma = 80^\circ, \tau = -20^\circ)$	100	100	98	88	84
$(\gamma = 50^\circ, \tau = 30^\circ)$	100	100	100	100	96
$(\gamma = 20^\circ, \tau = -70^\circ)$	94	86	64	80	68
$(\gamma = 20^\circ, \tau = 70^\circ)$	100	100	80	96	76
$(\gamma = 120^\circ, \tau = -70^\circ)$	92	84	74	74	64
$(\gamma = 120^\circ, \tau = 70^\circ)$	96	90	64	82	70
Mean	98	96	88	90	83
std	3	6.6	15	9.5	13

Figure 10: The first and third rows show the test images of two subjects in the CMU-PIE database at six different poses, with the pose numbers shown above each column. The second and fourth rows show the corresponding frontal view images generated by directly warping the given test images.

harmonic basis images from a single training image. For a test image at a different pose from the training images, we accomplish recognition by comparing the distance from a warped version of the test image to the space spanned by the basis images of each subject. The impact of some empirical factors (i.e., correspondence and interpolation) due to warping is embedded in a sparse transformation matrix, and we prove that the recognition performance is not significantly affected after warping the test image to the frontal view. Experimental results on both synthetic and real images show that high recognition rate can be achieved when the test image is at a different pose and under arbitrary illumination condition. Furthermore, the recognition results can be visually verified by easily generated face image of the chosen subject at the same pose as the test image.

In scenarios where only one training image is available, finding the cross-correspondence between the training images and the test image is inevitable. Automatic correspondence establishment is always a challenging problem. Recently, promising results have been shown by using the 4 planes, 4 transitions stereo matching algorithm described in [28]. The disparity map can be reliably built for a pair of images of the same person taken under the same lighting condition, even with some occlusions. We conducted some experiments using this technique on both synthetic

Table 6: The correct recognition rates at six rotated pose under various lighting conditions for 68 subjects in the CMU-PIE database.

Lighting/pose	f2f	c05		c07		c09		c11		c29		c37	
		(r2f)	(r2r)	(r2f)	(r2r)	(r2f)	(r2r)	(r2f)	(r2r)	(r2f)	(r2r)	(r2f)	(r2r)
$f02$	86	84	80	84	82	82	80	82	76	82	80	80	76
$f03$	95	94	90	95	92	94	92	92	84	92	88	90	84
$f04$	97	96	94	97	95	97	94	94	90	97	94	92	88
$f05$	98	98	94	98	96	96	96	94	90	96	94	92	90
$f06$	100	100	99	100	100	100	100	98	96	100	99	98	94
$f07$	98	98	96	100	100	100	98	94	94	97	95	92	92
$f08$	97	96	94	97	95	97	94	92	90	96	94	92	88
$f09$	100	100	98	100	99	100	98	100	96	100	98	99	96
$f10$	100	100	98	100	100	100	100	96	94	100	98	92	92
$f11$	100	100	100	100	100	100	100	98	96	100	100	98	96
$f12$	96	94	92	94	94	95	95	90	88	92	92	90	86
$f13$	98	96	92	96	94	94	94	92	88	94	92	90	88
$f14$	100	100	98	100	100	100	100	98	94	99	96	96	92
$f15$	100	100	100	100	100	100	100	100	97	100	98	98	96
$f16$	98	97	95	98	96	98	96	96	92	97	95	95	90
$f17$	95	94	92	95	95	95	95	92	88	94	90	90	86
$f18$	92	90	88	92	90	90	88	86	82	90	86	86	80
$f19$	96	95	90	94	92	92	92	90	86	94	90	84	82
$f20$	96	95	92	96	94	95	94	92	88	94	90	90	84
$f21$	97	97	97	97	96	97	95	94	92	95	95	94	90
$f22$	97	97	95	96	95	95	95	94	90	95	94	92	90
Mean	97	96	94	96	95	96	95	93	90	95	93	92	89
std	3.2	3.8	4.6	3.7	4.2	4.2	4.6	4.3	5.1	4.2	4.7	4.6	5.2

and real images. Reasonably good correspondence maps were achieved, even for cross-subject images. This technique has been used for 2D face recognition across pose [29]. However, like all the other stereo methods, the intensity-invariant condition is required, which does not hold if the images are taken under different lighting conditions. For our challenging face recognition application, the lighting condition of the test image is unconstraint. Therefore, currently this stereo method cannot be directly used to build the correspondence between I_{test} and I_f. Further investigations are being taken for dense stereo with illumination variations compensated.

APPENDIX

Assume that (n_x, n_y, n_z) and (n'_x, n'_y, n'_z) are the surface normals of point p at the frontal pose and the rotated view, respectively. (n'_x, n'_y, n'_z) is related to (n_x, n_y, n_z) as

$$\begin{bmatrix} n'_y \\ n'_y \\ n'_y \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}, \tag{A.1}$$

where $-\theta$ is the azimuth angle.

By replacing (n'_x, n'_y, n'_z) in (A.1) with $(n_z \sin\theta + n_x \cos\theta, n_y, n_z \cos\theta - n_x \sin\theta)$, and assuming that the correspondence between the rotated view and the frontal view has been built, we have

$$\begin{aligned}
b'_{00} &= \frac{1}{\sqrt{4\pi}}\lambda, \quad b'_{10} = \sqrt{\frac{3}{4\pi}}\lambda.*(n_z \cos\theta - n_x \sin\theta), \\
b'^{e}_{11} &= \sqrt{\frac{3}{4\pi}}\lambda.*(n_z \cos\theta - n_x \sin\theta), \quad b'^{o}_{11} = \sqrt{\frac{3}{4\pi}}\lambda.*n_y, \\
b'_{20} &= \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda.*(2(z\cos\theta - n_x \sin\theta)^2 \\
&\qquad - (n_z \sin\theta + n_x \cos\theta)^2 - n_y^2), \\
b'^{e}_{21} &= 3\sqrt{\frac{5}{12\pi}}\lambda.*(n_z \sin\theta + n_x \cos\theta)*(n_z \cos\theta - n_x \sin\theta), \\
b'^{o}_{21} &= 3\sqrt{\frac{5}{12\pi}}\lambda.*n_y(n_z \cos\theta - n_x \sin\theta), \\
b'^{e}_{22} &= \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda.*((n_z \sin\theta + n_x \cos\theta)^2 - n_y^2) \\
b'^{o}_{22} &= 3\sqrt{\frac{5}{12\pi}}\lambda.*(n_z \sin\theta + n_x \cos\theta)n_y.
\end{aligned} \tag{A.2}$$

Rearranging, we get

$$
\begin{aligned}
&b'_{00} = b_{00}, \qquad b'_{10} = b_{10}\cos\theta - b^e_{11}\sin\theta,\\
&b'^e_{11} = b^e_{11}\cos\theta + b_{10}\sin\theta, \qquad b'^o_{11} = b_{11},\\
&b'_{20} = b_{20} - \sqrt{3}\,\sin\theta\,\cos\theta b^e_{21} - \sqrt{\frac{5}{4\pi}}\frac{3}{2}\sin^2\theta(n_z^2 - n_x^2),\\
&b'^e_{21} = (\cos^2\theta - \sin^2\theta)b^e_{21} + 3\sqrt{\frac{5}{12\pi}}\,\sin\theta\,\cos\theta(n_z^2 - n_x^2),\\
&b'^o_{21} = b^o_{21}\cos\theta - b^o_{22}\sin\theta,\\
&b'^e_{22} = b^e_{22} + \cos\theta\,\sin\theta\, b^e_{21} + \sqrt{\frac{5}{12\pi}}\frac{3}{2}\sin^2\theta(n_z^2 - n_x^2),\\
&b'^o_{22} = b^o_{22}\cos\theta + b^o_{21}\sin\theta.
\end{aligned}
\tag{A.3}
$$

As shown in (A.3), $b'_{00}, b'_{10}, b'^e_{10}, b'^o_{11}, b'^o_{21}$ and b'^o_{22} are linear combinations of basis images at the frontal pose. For b'_{20}, b'^e_{21} and b'^e_{22}, we need to have $(n_z^2 - n_x^2)$ which is not known. From [4], we know that if the sphere is illuminated by a single directional source in a direction other than the z direction, the reflectance obtained would be identical to the kernel, but shifted in phase. Shifting the phase of a function distributes its energy between the harmonics of the same order n (varying m), but the overall energy in each order n is maintained. The quality of the approximation, therefore, remains the same. This can be verified by $b'^2_{10} + b'^{e2}_{11} + b'^{o2}_{11} = b^2_{10} + b^{e2}_{11} + b^{o2}_{11}$ for the order $n = 1$. Noticing that $b'^{o2}_{21} + b'^{o2}_{22} = b^{o2}_{21} + b^{o2}_{22}$, we still need $b'^2_{20} + b'^{e2}_{21} + b'^{e2}_{22} = b^2_{20} + b^{e2}_{21} + b^{e2}_{22}$ to preserve the energy for the order $n = 2$.

Let $G = 3\sqrt{5/12\pi}\sin^2\theta(n_z^2 - n_x^2)$ and also let $H = 3\sqrt{5/12\pi}\,\sin\theta\,\cos\theta(n_z^2 - n_x^2)$, we have

$$
\begin{aligned}
b'_{20} &= b_{20} - \sqrt{3}\,\sin\theta\,\cos\theta\, b^e_{21} - \frac{\sqrt{3}}{2}G,\\
b'^e_{21} &= (\cos^2\theta - \sin^2\theta)b^e_{21} + H,\\
b'^e_{22} &= b^e_{22} + \cos\theta\,\sin\theta\, b^e_{21} + \frac{1}{2}G.
\end{aligned}
\tag{A.4}
$$

Then

$$
\begin{aligned}
&b'^2_{20} + b'^{e2}_{21} + b'^{e2}_{22}\\
&= b^2_{20} + b^{e2}_{21} + b^{e2}_{22} + \frac{3G^2}{4} - 2\sqrt{3}\,\sin\theta\,\cos\theta\, b_{20}b^e_{21}\\
&\quad -\sqrt{3}\,b_{20}G + 3\sin\theta\,\cos\theta G + H^2 + 2(\cos^2\theta - \sin^2\theta)b^e_{21}H\\
&\quad +\frac{G^2}{4} + 2\sin\theta\,\cos\theta\, b^e_{22}b^e_{21} + b^e_{22}G + \sin\theta\,\cos\theta G\\
&= b^2_{20} + b^{e2}_{21} + b^{e2}_{22} + G^2 + 4\sin\theta\,\cos\theta\, b^e_{21}G\\
&\quad +(b^e_{22} - \sqrt{3}b_{20})(G + 2\sin\theta\,\cos\theta b^e_{21}) + H^2\\
&\quad +2(\cos^2\theta - \sin^2\theta)b^e_{21}H.
\end{aligned}
\tag{A.5}
$$

Having $b'^2_{20} + b'^{e2}_{21} + b'^{e2}_{22} = b^2_{20} + b^{e2}_{21} + b^{e2}_{22}$ and $H = G(\cos\theta/\sin\theta)$, we get

$$
\begin{aligned}
&G^2 + 2\sin\theta\,\cos\theta\, b^e_{21}G\\
&+(b^e_{22} - \sqrt{3}\,b_{20})(G\sin^2\theta + 2\sin\theta\,\cos\theta\, b^e_{21}) = 0,
\end{aligned}
\tag{A.6}
$$

and then $(G + 2\sin\theta\cos\theta b^e_{21})(G + \sin^2\theta(b^e_{22} - \sqrt{3}b_{20})) = 0$.

Two possible roots of the polynomial are $G = -2\sin\theta\cos\theta b^e_{21}$ or $G = -\sin^2\theta(b^e_{22} - \sqrt{3}b_{20})$. Substituting $G = -2\sin\theta\cos\theta b^e_{21}$ into (A.4) gives $b'_{20} = b_{20}$, $b'^e_{21} = -b^e_{21}$, $b'^e_{22} = b^e_{22}$, which is apparently incorrect. Therefore, we have $G = -\sin^2\theta(b^e_{22} - \sqrt{3}b_{20})$ and $H = -\cos\theta\sin\theta(b^e_{22} - \sqrt{3}b_{20})$. Substituting them in (A.4), we get

$$
\begin{aligned}
b'_{20} &= b_{20} - \sqrt{3}\,\sin\theta\,\cos\theta\, b^e_{21} + \frac{\sqrt{3}}{2}\sin^2\theta(b^e_{22} - \sqrt{3}b_{20}),\\
b'^e_{21} &= (\cos^2\theta - \sin^2\theta)b^e_{21} - \cos\theta\,\sin\theta(b^e_{22} - \sqrt{3}b_{20}),\\
b'^e_{22} &= b^e_{22} + \cos\theta\,\sin\theta\, b^e_{21} - \frac{1}{2}\sin^2\theta(b^e_{22} - \sqrt{3}b_{20}).
\end{aligned}
\tag{A.7}
$$

Using (A.3) and (A.7), we can write the basis images at the rotated pose in the matrix form of the basis images at the frontal pose, as shown in (3).

Assuming that there is an elevation angle $-\beta$ after the azimuth angle $-\theta$ and denoting by (n''_x, n''_y, n''_z) the surface normal for the new rotated view, we have

$$
\begin{bmatrix} n''_x \\ n''_y \\ n''_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} n'_x \\ n'_y \\ n'_z \end{bmatrix}. \tag{A.8}
$$

Repeating the above derivation easily leads to the linear equations in (4) which relates the basis images at the new rotated pose to the basis images at the old rotated pose.

Next, we show that the proved proposition is consistent with the general rotation matrix of spherical harmonics. If we use a ZYZ formulation for the general rotation, we have $R_{\theta,\omega,\beta} = R_z(\omega)R_y(\theta)R_z(\beta)$, the dependence of D_l on ω and β is simple, $D^l_{m,m'}(\theta,\omega,\beta) = d^l_{m,m'}(\theta)e^{im\omega}e^{im'\beta}$ where d^l is a matrix that defines how a spherical harmonic transforms under rotation about the Y-axis. We can further decompose it into a rotation of 90° about the X-axis, a general rotation θ about the Z-axis followed finally by a rotation of $-90°$ about the X-axis [30]. Since

$$
X_{\mp 90} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \mp 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1/2 & 0 & -\sqrt{3}/2 \\
0 & 0 & 0 & 0 & \mp 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -\sqrt{3}/2 & 0 & 1/2
\end{bmatrix}, \tag{A.9}
$$

$$
Z_\theta = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \cos\theta & 0 & \sin\theta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\sin\theta & 0 & \cos\theta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \cos 2\theta & 0 & 0 & 0 & \sin 2\theta \\
0 & 0 & 0 & 0 & 0 & \cos\theta & 0 & \sin\theta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -\sin\theta & 0 & \cos\theta & 0 \\
0 & 0 & 0 & 0 & -\sin 2\theta & 0 & 0 & 0 & \cos 2\theta
\end{bmatrix}, \tag{A.10}
$$

it is easy to show that $R_Y(\theta)$ is exactly the same as shown in (3) by taking the above equations into $R_Y(\theta) = X_{-90}Z_\theta X_{+90}$ and reorganizing the order of the spherical harmonics $Y_{l,m}$. Since (4) is derived similarly as (3), the rotation around the x-axis can be proved to be the same as (4). This can also be verified by taking the rotation angle $\beta = \mp 90^\circ$ into (4) which gives the same $X_{\mp 90^\circ}$ as shown above.

ACKNOWLEDGMENT

This work is partially supported by a contract from UNISYS.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[3] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 351–363, 2006.

[4] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[5] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 1322–1333, 2002.

[6] Y. Tanabe, T. Inui, and Y. Onodera, *Group Theory and Its Applications in Physics*, Springer, Berlin, Germany, 1990.

[7] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for reflection," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 4, pp. 1004–1042, 2004.

[8] Z. Yue, W. Zhao, and R. Chellappa, "Pose-encoded spherical harmonics for robust: face recognition using a single image," in *Proceedings of the 2nd International Workshop on Analysis and Modelling of Faces and Gestures (AMFG '05)*, vol. 3723, pp. 229–243, Beijing, China, October 2005.

[9] L. Zhang and D. Samaras, "Face recognition under variable lighting using harmonic image exemplars," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. 19–25, Madison, Wis, USA, June 2003.

[10] "3dfs-100 3 dimensional face space library (2002 3rd version)," University of Freiburg, Germany.

[11] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (AFGR '02)*, pp. 46–51, Washington, DC, USA, May 2002.

[12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[13] T. Sim and T. Kanade, "Illuminating the face," Tech. Rep. CMU-RI-TR-01-31, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pa, USA, 2001.

[14] B. Beyme, "Face recognition under varying pose," Tech. Rep. 1461, MIT AI Lab, Cambridge, Mass, USA, 1993.

[15] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 84–91, Seattle, Wash, USA, June 1994.

[16] W. T. Freeman and 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 (CVPR '97)*, pp. 554–560, San Juan, Puerto Rico, USA, June 1997.

[17] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "Illumination-based image synthesis: creating novel images of humanfaces under differing pose and lighting," in *Proceedings of the IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes (MVIEW '99)*, pp. 47–54, Fort Collins, Colo, USA, June 1999.

[18] W. Zhao and R. Chellappa, "Symmetric shape-from-shading using self-ratio image," *International Journal of Computer Vision*, vol. 45, no. 1, pp. 55–75, 2001.

[19] R. Dovgard and R. Basri, "Statistical symmetric shape from shading for 3D structure recovery of faces," in *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, pp. 99–113, Prague, Czech Republic, May 2004.

[20] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (AFGR '00)*, pp. 285–292, Grenoble, France, March 2000.

[21] Z. Yue and R. Chellappa, "Pose-normailzed view synthesis of a symmetric object using a single image," in *Proceedings of the 6th Asian Conference on Computer Vision (ACCV '04)*, pp. 915–920, Jeju City, Korea, January 2004.

[22] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs, "Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 230–245, 2007.

[23] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[24] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2D+3D active appearance models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 535–542, Washington, DC, USA, June-July 2004.

[25] S. Romdhani, J. Ho, T. Vetter, and D. J. Kriegman, "Face recognition using 3-D models: pose and illumination," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1977–1999, 2006.

[26] P. Henrici, "Barycentric formulas for interpolating trigonometric polynomials and their conjugates," *Numerische Mathematik*, vol. 33, no. 2, pp. 225–234, 1979.

[27] T. Horprasert, Y. Yacoob, and L. S. Davis, "Computing 3-D head orientation from a monocular image sequence," in *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (AFGR '96)*, pp. 242–247, Killington, Vt, USA, October 1996.

[28] A. Criminisi, J. Shotton, A. Blake, C. Rother, and P. H. S. Torr, "Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming," *International Journal of Computer Vision*, vol. 71, no. 1, pp. 89–110, 2007.

[29] C. Castillo and D. Jacobs, "Using stereo matching for 2-D face recognition across pose," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, Minneapolis, Minn, USA, June 2007.

[30] R. Green, "Spherical harmonic lighting: the gritty details," in *Proceedings of the Game Developers' Conference (GDC '03)*, San Jose, Calif, USA, March 2003.