Pose-Encoded Spherical Harmonics for Robust Face Recognition 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

Abstract. Face recognition under varying pose is a challenging problem, especially when illumination variations are also present. Under Lambertian model, spherical harmonics representation has proved to be effective in modelling illumination variations for a given pose. In this paper, we extend the spherical harmonics representation to encode pose information. More specifically, we show that 2D harmonic basis images at different poses are related by close-form linear combinations. This enables an analytic method for generating new basis images at a different pose which are typically required to handle illumination variations at that particular pose. Furthermore, the orthonormality of the linear combinations is utilized to propose an efficient method for robust face recognition where only one set of front-view basis images per subject is stored. In the method, we directly project a rotated testing image onto the space of front-view basis images after establishing the image correspondence. Very good recognition results have been demonstrated using this method.

1 Introduction

Face recognition is one of the most successful applications of image analysis and understanding. In spite of recent advances, robust face recognition under variable lighting and pose remains to be a challenging problem. This is due to the fact that we need to compensate for both significant pose and illumination change at the same time. It becomes even more difficult when only one training image per subject is available. Recently, methods have been proposed to handle the illumination problem when only one training image is available, for example, a statistical learning method [13] based on spherical harmonics representation [1, 9]. In this paper, we propose to extend the harmonics representation to encode pose information. That is, all the harmonic basis images of a subject at various poses are related to the front-view basis images via close-form linear combinations. Moreover, these linear combinations are orthonormal. This suggests that recognition methods based on projection onto the harmonic basis images [1] for rotated testing images can be made very efficient. We do not need to generate a new set of basis images at the same pose as that of the testing images. In stead, we can directly use the existing front-view basis images without changing the matching score defined in [1].

We propose an efficient face recognition method that needs only one set of basis images per subject for robust recognition of faces under variable illuminations and poses. The flow chart of our face recognition system is shown in Fig. 1. We have a single training image at the frontal pose for each subject in the training set. The basis images

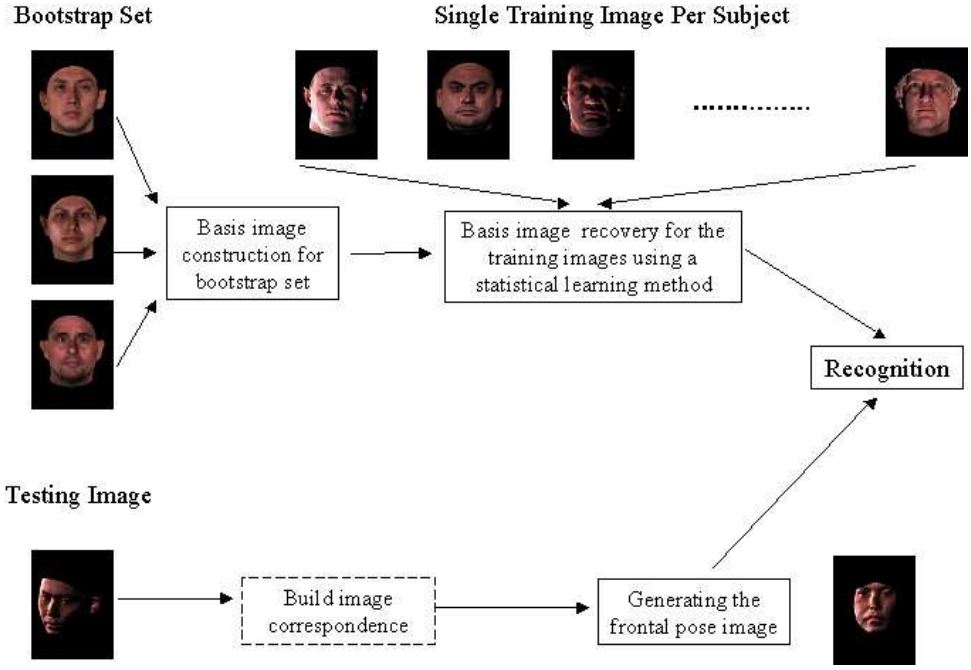


Fig. 1. The flow chart of the proposed face recognition system.

for each training subject are recovered using a statistical learning algorithm [13] with the aid of a bootstrap set consisting of 3D face scans. For a testing image at a rotated pose and under an arbitrary illumination condition, we first establish the image correspondence between the testing image and the training images. The frontal pose image is then warped from the testing image. Finally, a face is identified for which there exists a linear reconstruction based on basis images that is the closest to the testing image.

The remainder of the paper is organized as follows: Section 2 introduces the related work. The pose-encoded spherical harmonic representation is presented in Section 3 where we prove that the basis images at a rotated pose is a linear combination of the basis images at the frontal pose. Section 4 presents the complete face recognition 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 the system performance is demonstrated in Section 4.3. We conclude our paper in Section 5.

2 Related Work

Either pose variations or illumination variations may cause serious performance degradation for existing face recognition systems. [17] examined these two problems and reviewed some approaches to solving them. The early effort to handle illumination vari-

ations was to discard the first few principal components, which packs most of the energy caused by illumination variations [2]. In this method, the testing image must have the same pose as the training images. In [3], a template matching scheme was proposed to handle pose variations. It needs many different views per person and no lighting variations are allowed. Approaches to face recognition under pose variations [8][6] avoid the correspondence problem by storing multiple images at different poses for each person. View-based eigenface methods [8] explicitly code the pose information by constructing an individual eigenface for each pose. [6] treats face recognition across poses as a bilinear problem and disentangles the face identity and the head pose.

Few methods consider both pose and illumination variations at the same time. The synthesis method in [7] can handle both illumination and pose variations by reconstructing the face surface using the illumination cone method under fixed pose and rotating it to the desired pose. A set of training images are required for each subject to construct the illumination cone. [16] presented a symmetric shape-from-shading (SFS) approach to recover both shape and albedo for symmetric objects. This work was extended in [5] to recover the 3D shape of a human face using a single image. In [15], a unified approach was proposed to solving the pose and illumination problem. A generic 3D model was used to establish the correspondence and estimate the pose and illumination direction. [12] presents a pose-normalized face synthesis method under varying illuminations using the bilateral symmetry of the human face. A Lambertian model was assumed and single light source was considered. [18] extends the photometric stereo algorithms to recover albedos and surface normals from one image under unknown single distant illumination conditions.

Recent work on spherical harmonics representation has been independently conducted by by Basri et al. [1] and Ramamoorthi [9]. It has been shown that the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a low dimensional linear subspace. The basis images spanning the illumination space for each face can be rendered from a 3D scan of the face [1] or estimated by applying PCA to a number of images of the same subject under different illuminations [9]. Following the statistical learning scheme in [10], Zhang et al. [13] showed that the basis images spanning this space can be recovered from just one image taken under arbitrary illumination conditions when the pose is fixed.

To handle both pose and illumination variations, 3D morphable face model has been proposed. By far the most impressive face synthesis results were reported in [4] followed with very high recognition rates, 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. In order to handle illumination more effectively, a recent work [14] incorporates spherical harmonics into the morphable model framework. Most of the 3D morphable model approaches are computationally intense because of the large number of parameters that need to be optimized.

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 [1]. It can be shown that the

irradiance can be approximated by the combination of the first nine spherical harmonics for Lambertian surfaces. Any image of an object under certain illumination conditions is a linear combination of a series of basis images $\{b_{mn}\}$. In order to generate the basis images for the object, 3D information is required as shown in Appendix A.

For a fixed pose, spherical harmonics representation has proved to be effective in modelling illumination variations, even in the case when a bootstrap set of 3D models and only one training image per subject are available [13]. In the presence of both illumination and pose variance, two possible approaches can be taken. One is to use 3D morphoable model to reconstruct the 3D model from a single training image and then build spherical harmonic basis images at the pose of the testing image for recognition [14]. 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 3D morphoable model method could be computationally expensive. As for efficient recognition of a rotated testing image, a natural question to ask is: can we represent the basis images at different poses using one set of basis images at a given pose, say, the front-view pose? In this section, we address this question by showing that 2D harmonic basis images at different poses are related by close-form linear combinations. This enables an analytic method for generating new basis images at different poses from basis images at one pose.

Assuming that the testing image is at a different pose (rotated view) as the training images (usually frontal view), we aim to derive the basis images at the rotated pose from the basis images at the frontal pose, assuming that the correspondence between the rotated view and the frontal view has been built. The general rotation can be decomposed into three concatenated rotations around the X, Y and Z axis, namely elevation, azimuth and roll, respectively. Roll is an in-plane rotation that can be handled much easily and will not be discussed here. The following theorem states that the basis images at the rotated pose is a linear combination of the basis images at the frontal pose, and the transformation matrix is a function of the rotation angles only.

Theorem 1 Assume a rotated view is obtained by rotating a front-view head with an azimuth angle $-\theta$. With 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 in the following linear form:

$$\begin{cases} b'_{00} = b_{00} \\ \begin{bmatrix} b'_{10} \\ b'^{e}_{11} \\ b'^{o}_{11} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{10} \\ b^{e}_{11} \\ b^{o}_{11} \end{bmatrix} \\ \begin{bmatrix} b'_{20} \\ b'^{e}_{21} \\ b'^{o}_{21} \\ b'^{e}_{22} \\ b'^{o}_{22} \end{bmatrix} = \begin{bmatrix} 1-\frac{3}{2}\sin^2\theta & -\sqrt{3}\sin\theta\cos\theta & 0 & \frac{\sqrt{3}}{2}\sin^2\theta & 0 \\ \sqrt{3}\sin\theta\cos\theta & \cos^2\theta-\sin^2\theta & 0 & -\cos\theta\sin\theta & 0 \\ 0 & 0 & \cos\theta & 0 & -\sin\theta \\ \frac{\sqrt{3}}{2}\sin^2\theta & \cos\theta\sin\theta & 0 & 1-\frac{1}{2}\sin^2\theta & 0 \\ 0 & 0 & \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} b_{20} \\ b^{e}_{21} \\ b^{o}_{21} \\ b^{e}_{22} \\ b^{o}_{22} \end{bmatrix} \end{cases} \tag{1}$$

If there is an elevation angle $-\beta$ other than the azimuth angle $-\theta$, the basis images B'' for the newly rotated view are related to B' in the following linear form:

$$
\begin{cases}
b''_{00} = b'_{00} \\
\begin{bmatrix} b''_{10} \\ b''^{e}_{11} \\ b''^{o}_{11} \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} b'_{10} \\ b'^{e}_{11} \\ b'^{o}_{11} \end{bmatrix} \\
\begin{bmatrix} b''_{20} \\ b''^{e}_{21} \\ b''^{o}_{21} \\ b''^{e}_{22} \\ b''^{o}_{22} \end{bmatrix} = \begin{bmatrix} 1-\frac{3}{2}\sin^2\beta & 0 & \sqrt{3}\sin\beta\cos\beta & \frac{-\sqrt{3}}{2}\sin^2\beta & 0 \\ 0 & \cos\beta & 0 & 0 & \sin\beta \\ -\sqrt{3}\sin\beta\cos\beta & 0 & \cos^2\beta-\sin^2\beta & -\cos\beta\sin\beta & 0 \\ \frac{-\sqrt{3}}{2}\sin^2\beta & 0 & \cos\beta\sin\beta & 1-\frac{1}{2}\sin^2\beta & 0 \\ 0 & -\sin\beta & 0 & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} b'_{20} \\ b'^{e}_{21} \\ b'^{o}_{21} \\ b'^{e}_{22} \\ b'^{o}_{22} \end{bmatrix}
\end{cases}
\tag{2}
$$

For proof of this theorem, please see Appendix B.

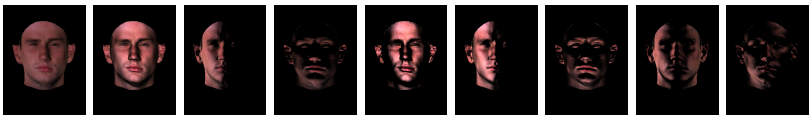
The basis images at various poses can be generated from a set of basis images at the frontal pose using the linear relationship in (1) and (2). Although in theory new basis images can be generated from a rotated 3D model inferred by existing basis images since basis images actually capture the albedo (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, let alone the computational cost. Now we have proved that the procedure of first rotating objects and then recomputing basis images at a desired pose can be *totally* avoided.

It is easy to see that the coefficient matrices in (1) and (2) are block diagonal, thus preserving the energy on each band $n = 0, 1, 2$. Moreover, the orthonormality of the coefficient matrices helps to further simplify the computation required for recognition of the rotated testing image as shown in Section 4.2.

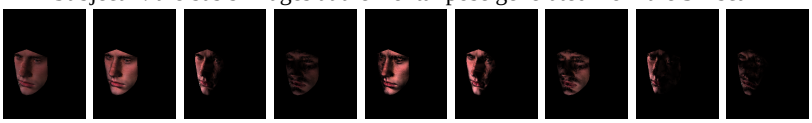
We synthesized the basis images at arbitrary rotated poses from those at the frontal pose using (1) and (2), and compared them with the ground truth in Fig. 2. The first row through the third row are the results for subject 1, with the first row showing the basis images at the frontal pose generated from the 3D scan, the second row showing the synthesized basis images at the rotated pose (azimuth angle $\theta = -30^o$, elevation angle $\beta = 20^o$), and the third row showing the ground truth of the basis images at the rotated pose. Rows four through six are 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 synthesized basis images for another rotated view (azimuth angle $\theta = -30^o$, elevation angle $\beta = -20^o$), and the last row showing the ground truth of the basis images at the rotated pose. As we can see from Fig. 2, the synthesized basis images at the rotated poses have no noticeable difference with the ground truth.

4 Face Recognition Using Pose-Encoded Spherical Harmonics

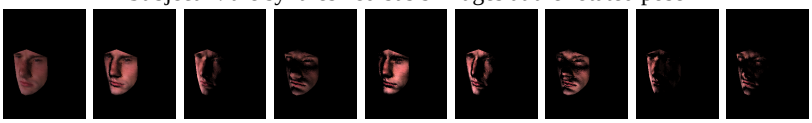
In this section we present an efficient face recognition method using pose-encoded spherical harmonics. Only one training image is needed per subject and high recognition performance is achieved even when the testing image is at a different pose from the training image and under an arbitrary illumination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



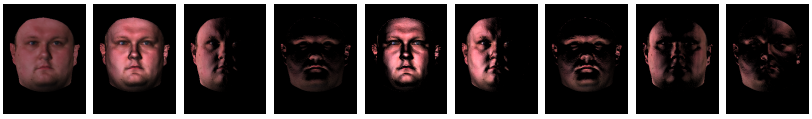
Subject 1: the basis images at the frontal pose generated from the 3D scan



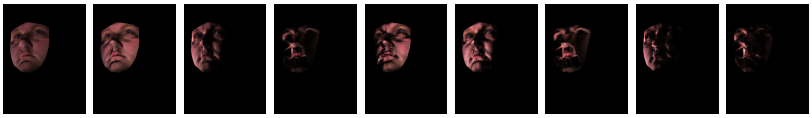
Subject 1: the synthesized basis images at the rotated pose



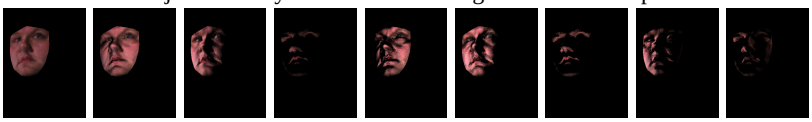
Subject 1: the ground truth of the basis images at the rotated pose



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

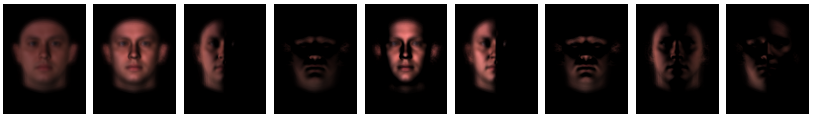


Subject 2: the synthesized basis images at the rotated pose



Subject 2: the ground truth of the basis images at the rotated pose

Fig. 2. Results of the synthesized basis images at the rotated pose. The first row through the third row are 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 synthesized basis images at the rotated pose (with the azimuth angle $\theta = -30^o$ and the elevation angle $\beta = 20^o$), and the third row showing the ground truth of the basis images at the rotated pose. Rows four through six are 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 synthesized basis images at another rotated pose (with the azimuth angle $\theta = -30^o$ and the elevation angle $\beta = -20^o$) and the last row showing the ground truth of the basis images at the rotated pose.



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

Fig. 3. The sample mean basis images estimated from the bootstrap set.

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 [13].

We build a bootstrap set with 50 3D face scans and the texture information from Vetter's 3D face database [19], and generate 9 basis images for each face model. For a novel d-dimensional vectorized image I, let B be the $d \times 9$ matrix of basis images, α a 9 dimensional vector and E a d-dimensional error term, we have $I = B\alpha + E$. It is assumed that the pdf's of B are Gaussian distributions and the sample mean vectors $\mu_b(x)$ and the sample covariance matrixes $C_b(x)$ are estimated from the basis images in the bootstrap set. Fig. 3 shows the sample mean of the basis images estimated from the bootstrap set.

The problem of estimating the basis images B and the illumination coefficients α is a coupled estimation problem because of its bilinear form. It is simplified by estimating α in a prior step with kernel regression and using it consistently across all pixels to recover B. K bootstrap images $\{J_k\}_{k=1}^K$ with known coefficients $\{\alpha_k\}_{k=1}^K$ are generated from the 3D face scans in the bootstrap set. Given a new image i_{tra}, the coefficients α_{tra} can be estimated as

$$\alpha_{tra} = \frac{\sum_{k=1}^K w_k \alpha_k}{\sum_{k=1}^K \alpha_k} \tag{3}$$

where $w_k = exp[-\frac{1}{2}(D(i, J_k)/\sigma_k)^2]$ and $D(i, J_k) = \|i - J_k\|_2$, σ_k is the width of the k-th Gaussian kernel which controls the influence of J_k on the estimation of α_{tra}. All $\{\sigma_k\}_{k=1}^K$ are pre-computed in a way such that ten percent of the bootstrap images are within $1 \times \sigma_k$ at each σ_k. The sample mean $\mu_e(x, \alpha)$ and the sample variance $\sigma_e^2(x, \alpha)$ of the error term $E(\alpha)$ are also estimated using kernel regression, similar to (3).

Given a novel face image $i(x)$, with the estimated coefficients α, the corresponding basis images $b(x)$ at each pixel x are recovered by computing the maximum a posteriori (MAP) estimate, $b_{MAP}(x) = argmax_{b(x)}(P(b(x)|i(x)))$. Using Bayes rule:

$$\begin{aligned} b_{MAP}(x) &= argmax_{b(x)} P(i(x)|b(x))P(b(x)) \\ &= argmax_{b(x)} \left\{\mathcal{N}\left(b(x)^T\alpha + \mu_e, \sigma_e^2\right) \times \mathcal{N}\left(\mu_b(x), C_b(x)\right)\right\} \end{aligned} \tag{4}$$

Taking logarithm, and setting the derivatives of the right hand side of (4) (w.r.t $b(x)$) to 0, we get $A * b_{MAP} = T$, where $A = \frac{1}{\sigma_e^2}\alpha\alpha^T + C_b^{-1}$ and $T = \frac{(i-\mu_e)}{\sigma_e^2}\alpha + C_b^{-1}\mu_b$. By solving this linear equation, $b(x)$ of the subject can be recovered.

Combining Section 3 and Eq. (4), we illustrate in Fig. 4 the procedure of generating the basis images at a rotated pose (azimuth angle $\theta = -30^o$) from a single training image at the frontal pose. In the first part of Fig. 4, rows one though three show the

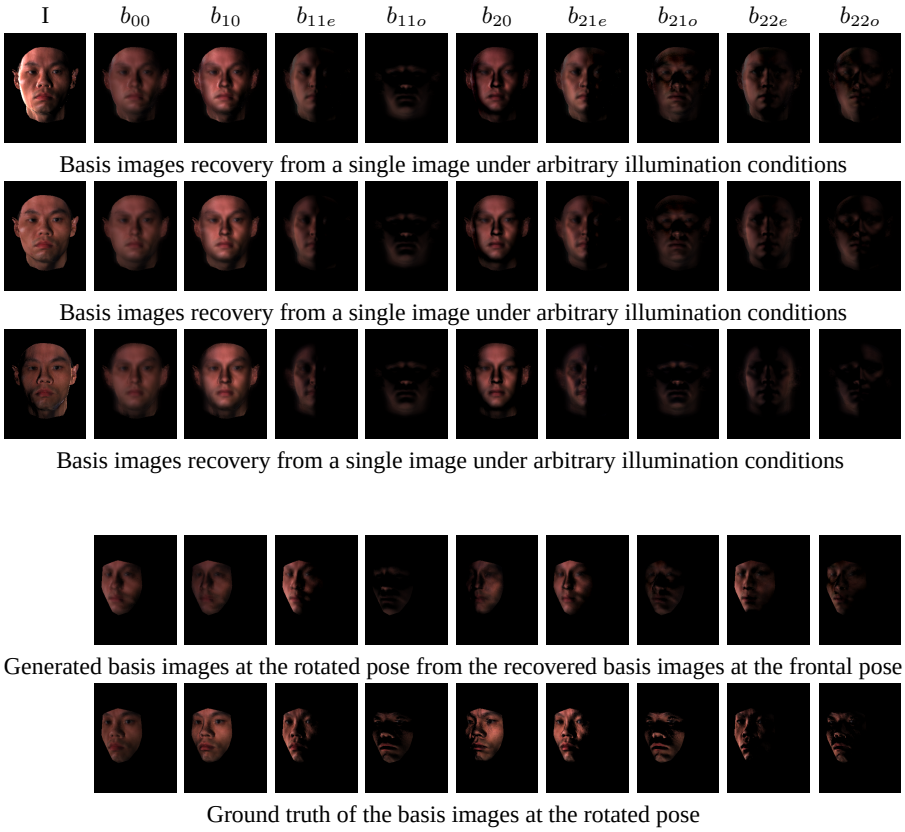


Fig. 4. Rows one though three show the basis images recovery from a single training image, with the first column showing different training images I under arbitrary illumination conditions for the same subject and the rest 9 columns showing the reconstructed basis images. Row four shows the generated basis images at the rotated pose from the recovered basis images at the frontal pose, and the fifth row show the ground truth of the basis images at the rotated pose.

basis images recovery 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 9 columns showing the reconstructed basis images. In the second part of Fig. 4, row four shows the generated basis images at the rotated pose from the recovered basis images at the frontal pose, and the fifth row shows the ground truth of the basis images at the rotated pose. As we can see from the plots, the basis images recovered from different training images of the same subject look very similar, although not perfect.

4.2 Recognition

For recognition, we follow a simple yet effective algorithm given in [1]. A face is identified for which there exists a weighted combination of basis images that is the closest to the testing image. Let B be the set of basis images at the frontal pose, with size $d \times r$,

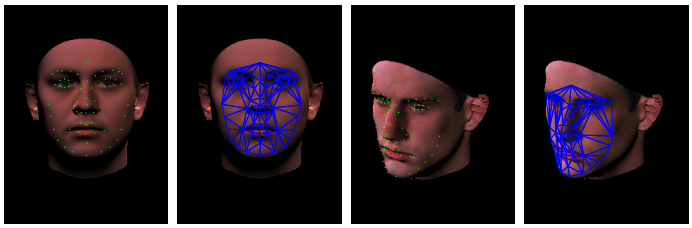


Fig. 5. Building dense correspondence between the rotated view and the frontal view using sparse features. The first and second image show the sparse features and the constructed meshes on the mean face at the frontal pose. The third and fourth image show the picked features and the constructed meshes on the given testing image at the rotated pose.

where d is the number of pixels in the image and r the number of basis images used. We use $r = 9$ as it is a natural choice capturing 98 percent of the energy of all the model's images [1]. 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, a $d \times r$ matrix with orthonormal columns, such that $B = QR$ where R is an $r \times r$ upper triangular matrix.

For a testing image I_{test} at a rotated pose, we can efficiently generate the set of basis images B' at that pose for each training subject from Section 3. The orthonormal basis Q' of the space spanned by B' can be computed by QR decomposition. The distance from the testing image I_{test} to the space spanned by B' is computed as $d_{match} = \|Q'Q'^T I_{test} - I_{test}\|$. However, this algorithm is not efficient overall because the set of basis images B', or the orthonormal basis Q', has to be generated for each training subject at the pose of an arbitrarily rotated testing image. The question is that can we have an overall efficient recognition method. The answer is yes based on the following lemma:

Lemma 2 The matching distance d_{match} of a rotated testing image I_{test} based on the basis images B' at that pose is the same as the matching distance of a geometrically synthesized front-view image I_f based on the basis images B.

Let C be the transpose of the combined coefficient matrices in (1) and (2), we have $B' = BC = QRC$ by QR decomposition. Applying QR decomposition again to RC, we have $RC = qr_{RC}$ where $q_{r\times r}$ is an orthonormal matrix. We now have $B' = Qqr_{RC} = Q_q r_{RC}$ by assuming $Q_q = Qq$. Since Q_q is the product of two orthonormal matrices, it forms a valid orthnormal basis for B'. Hence the matching distance is $\|Q_q Q_q^T I_{test} - I_{test}\|$. Now $Q_q Q_q^T = Qqq^T Q^T = QQ^T$ since q is orthonormal. Hence the final matching distance is $\|QQ^T I_{test} - I_{test}\|$. Recall this implies that the cross-pose correspondence between Q (B) and I_{test} has been established. To make this explicit, we use I_f, a geometrically warped front-view version of I_{test}, in the equation.

In brief summary, 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.

The remaining problem is that how the frontal pose image I_f is warped from I_{test}. Apparently the correspondence between the frontal pose and the rotated pose has to be established for the testing image. Finding correspondence is always challenging. Most

Table 1. The correct recognition rates at two rotated pose under various lighting conditions.

Illumination condition	Correct recognition rate at the pose $\theta = -30^o$ using our approach	Correct recognition rate with the training images at the same pose available	Correct recognition rate at the pose $\theta = -30^o, \beta = 20^o$ using our approach	Correct recognition rate with the training images at the same pose available
$(\gamma = 90^o, \tau = 10^o)$	94%	94%	94%	94%
$(\gamma = 30^o, \tau = 50^o)$	92%	100%	96%	100%
$(\gamma = 40^o, \tau = -10^o)$	90%	100%	92%	100%
$(\gamma = 70^o, \tau = 40^o)$	94%	100%	100%	100%
$(\gamma = 80^o, \tau = -20^o)$	80%	96%	86%	94%
$(\gamma = 50^o, \tau = 30^o)$	94%	100%	100%	100%
$(\gamma = 20^o, \tau = -70^o)$	86%	96%	94%	100%
$(\gamma = 20^o, \tau = 70^o)$	86%	92%	96%	96%
$(\gamma = 120^o, \tau = -70^o)$	42%	76%	62%	78%
$(\gamma = 120^o, \tau = 70^o)$	58%	84%	84%	86%

of the approaches to handle pose variations utilized manually picked sparse features to build the dense cross-pose or cross-subject correspondence. For I_{test} at an arbitrary pose, 63 designed feature points (eyebrows, eyes, nose, mouth and the face contour) were picked. A mean face from the 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. The number of feature points needed in our approach is comparable to the 56 manually picked feature points in [14] to deform the 3D model. Fig. 5 shows the feature points and the meshes on the mean face at the frontal pose and on a testing image at a rotated pose.

4.3 Recognition Results

We conducted the recognition experiments on Vetter's 3D face model database [19] for the sake of controllability and the convenience of comparison. There are totally 100 3D face models in the database, wherein 50 of them were used as the bootstrap set and the other 50 were used to generate training images. We synthesized the training images under a wide variety of illumination conditions with the 3D scans of the subjects. For each subject, only one frontal view image was stored as training image and used to recover the basis images B using the algorithm in Section 4.1. The orthonormal basis Q of the space spanned by B was obtained by applying QR decomposition to B. For a testing image I_{test} at an arbitrary pose, the frontal pose image I_f was synthesized by warping I_{test}, and the recognition score was computed as $\|QQ^T I_f - I_f\|$.

We generated the testing images at different poses from the training images by rotating the 3D scans and illuminated them with various lighting conditions (represented by slant angle γ and tilt angle τ). Fig. 6 (a) shows the testing images of a subject at the pose with the azimuth angle $\theta = -30^o$ and under 6 different lighting conditions.

We also did experiments under some extreme lighting conditions as shown in Fig. 6 (b). The corresponding frontal pose images were synthesized as shown in Fig. 6 (c) and (d) respectively. The correct recognition rates obtained by using $\|QQ^T I_f - I_f\|$ for all these illumination conditions are listed in column 2 of Table 1. The testing images at another pose (with $\theta = -30^o$ and $\beta = 20^o$) of the same subject are shown in Fig. 6 (e) and (f), with the generated frontal pose images shown in Fig. 6 (g) and (h) respectively and the correct recognition rates listed in column 4 of Table 1.

As an comparison, we also conducted the recognition experiment on the same testing images assuming that the training images at the same pose are available, as most of the approaches suggested. By recovering the basis images B at that pose using the algorithm in Section 4.1 and computing $\|QQ^T I_{test} - I_{test}\|$, we achieved the correct recognition rates as shown in column 3 and column 5 of Table 1 respectively, in correspondence with the two poses mentioned above. As we can see, the recognition rates using our approach are comparable to those when the training images at the rotated pose are available.

We have to point out that if the the testing image has a large pose variation from the frontal pose, it is inevitable that part of the face is self-occluded (Fig. 6). To have good recognition result, only the visible part of the face is used for recognition. Accordingly, only the visible parts of the basis images at the frontal pose are used as well.

5 Discussions and Conclusion

We have presented an efficient face recognition method to handle arbitrary pose and illumination from a single training image per subject using pose-encoded spherical harmonics. With a pre-built 3D face bootstrap set, we use a statistical learning method to obtain the spherical harmonic basis images from a single training image. We then show that the basis images at a rotated pose is a linear combination of the basis images at the frontal pose. For a testing image at a different pose from the training images, recognition is accomplished by comparing the distance from a warped version of the testing image to the space spanned by the basis images of each model. Experimental results show that high recognition rate can be achieved when the testing image is at a different pose and under an arbitrary illumination condition. We are planning to conduct experiments using the proposed approach on larger databases such as the CMU-PIE [11] database.

In the proposed method and existing methods where only one training image is available, finding the cross-correspondence between the training images and the testing image is inevitable. If the testing image is at a pose around the Y-axis only, a simpler method can be used to find the self-correspondence of the testing image by exploiting the bilateral symmetry of the human face. As a result, we do not need to build the cross-subject correspondence between the testing image and the training images. Unfortunately, automatic computation of these correspondences is not a trivial task and manual operation is required in existing methods. We are looking into possible solutions to address this issue.

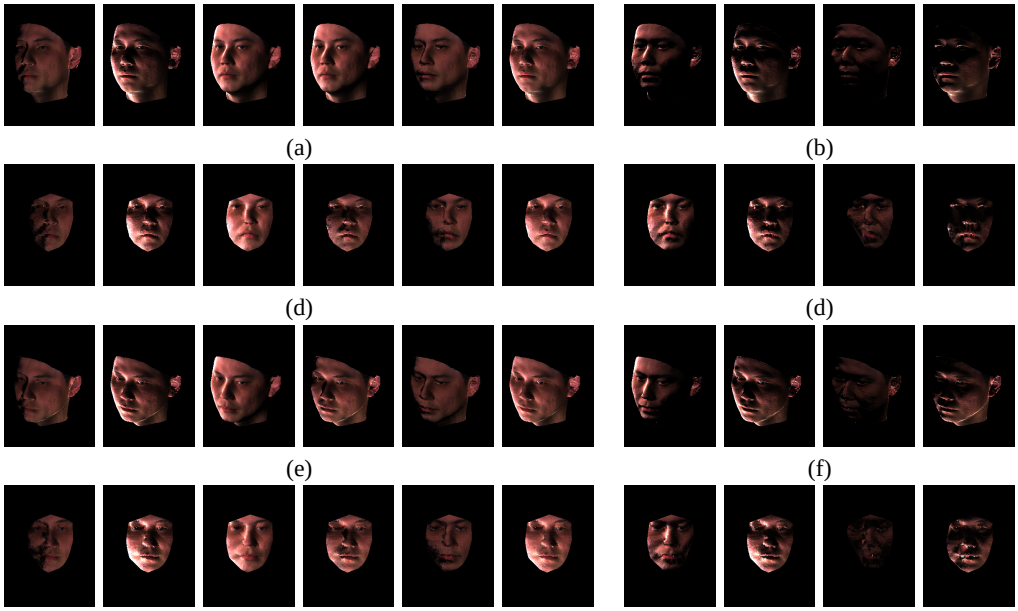


Fig. 6. (a) shows the testing images of a subject at the pose with the azimuth $\theta = -30^o$ under different lighting conditions (($\gamma = 90^o, \tau = 10^o$), ($\gamma = 30^o, \tau = 50^o$), ($\gamma = 40^o, \tau = -10$), ($\gamma = 20^o, \tau = 70^o$), ($\gamma = 80^o, \tau = -20^o$) and ($\gamma = 50^o, \tau = 30^o$) from left to right). The testing images of the same subject under some extreme lighting conditions (($\gamma = 20^o, \tau = -70^o$), ($\gamma = 20^o, \tau = 70^o$), ($\gamma = 120^o, \tau = -70^o$) and ($\gamma = 120^o, \tau = -70^o$) from left to right) are shown in (b). (c) and (d) show the generated frontal pose images from the testing images in (a) and (b) respectively. The testing images at another pose (with $\theta = -30^o$ and $\beta = 20^o$) of the same subject are shown in (e) and (f), with the generated frontal pose images shown in (g) and (h) respectively.

References

1. R. Barsi and D. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. PAMI*, Vol. 25(2), pp. 218–233, Feb. 2003.
2. P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. PAMI*, Vol. 19, pp. 711–720, July, 1997.
3. B. Beyme, "Face Recognition Under Varying Pose," *Tech. Report 1461, MIT AI Lab*, 1993.
4. V. Blanz and T. Vetter, "Face Recognition based on Fitting a 3D Morphable Model," *IEEE Trans. PAMI*, Vol. 25(9), pp. 1063–1074, Sept. 2003.
5. R. Dovgard and R. Basri, "Statistical Symmetric Shape from Shading for 3D Structure Recovery of Faces," *ECCV*, 2004.
6. W. Freeman and J. Tenenbaum, "Learning Bilinear Models for Two-Factor Problems in Vision," *Proceedings, IEEE Conference on CVPR*, Puerto Rico, pp. 554–560, June, 1997.
7. A. Geoghiades, P. Belhumeur, D. Kriegman, "Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose and Lighting," *Proceedings, Workshop on Multi-View Modeling and Analysis of Visual Scenes*, pp. 47–54, 1999.

8. A. Pentland, B. Moghaddam, and T. Starner, "View-based and Modular Eigenspaces for Face Recognition," *Proceedings, IEEE Conf. on CVPR*, pp. 84–91, June, 1994.
9. R. Ramamoorthi, "Analytic PCA Construction for theoretical Analysis of Lighting Variability in Images of A Lambertian Object," *IEEE. PAMI*, Vol. 24(10), pp. 1322–1333, Oct. 2002.
10. T. Sim and T. Kanade, "Illuminating the Face," *Tech. Report CMU-RI-TR-01-31*, Robotics Institute, CMU, 2001.
11. T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces," *AFGR*, pp. 46–51, 2002.
12. Z. Yue and R. Chellappa,"Pose-Normailzed View Synthesis of a Symmetric Object Using a Single Image," *ACCV*, 2004.
13. L. Zhang and D. Samaras, "Face Recognition Under Variable Lighting Using Harmonic Image Examplars," *CVPR*, Vol. I, pp. 19–25, 2003.
14. L. Zhang, S. Wang and D. Samaras, "Face Synthesis and Recognition from a Single Image under Arbitrary Unknown Lighting using a Spherical Harmonic Basis Morphable Model," *CVPR*, to appear, 2005.
15. W. Zhao and R. Chellappa, "SFS Based View Synthesis for Robust Face Recognition," *Int. Conf. on Automatic Face and Gesture Recognition*, 2000.
16. W. Zhao and R. Chellappa, "Symmetric Shape-from-Shading Using Self-ratio Image," *Int. Journal Computer Vision*, Vol. 45, pp. 55–75, 2001.
17. W. Zhao, R. Chellappa, J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, Dec. 2003.
18. S. Zhou, R. Chellappa and D. Jacobs, "Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints," *ECCV*, 2004.
19. "3DFS-100 3 Dimensional Face Space Library (2002 3rd version)", University of Freiburg.

Appendix A: Harmonic Basis Images

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 (5), 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 component-wise product of λ with any vector t.

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

Appendix B: Proof of Theorem 1

Assume that (n_x, n_y, n_z) and (n'_x, n'_y, n'_z) are the 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

$$
\begin{bmatrix} n'_x \\ n'_y \\ n'_z \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{6}
$$

where $-\theta$ is the azimuth angle.

By replacing (n'_x, n'_y, n'_z) in (5) with $(n_z \sin\theta + n_x \cos\theta, n_y, n_z \cos\theta - n_x \sin\theta)$, and assuming the correspondence between the rotated view and the frontal view has been built, we have

$$
\begin{aligned}
&b'_{00} = \frac{1}{\sqrt{4\pi}}\lambda,\ b'_{10} = \sqrt{\frac{3}{4\pi}}\lambda. * (n_z \cos\theta - n_x \sin\theta),\\
&b^{'e}_{11} = \sqrt{\frac{3}{4\pi}}\lambda. * (n_z \sin\theta + n_x \cos\theta),\ b^{'o}_{11} = \sqrt{\frac{3}{4\pi}}\lambda. * n_y,\\
&b'_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda. * (2(z\cos\theta - n_x \sin\theta)^2 - (n_z \sin\theta + n_x \cos\theta)^2 - n_y^2),\\
&b^{'e}_{21} = 3\sqrt{\frac{5}{12\pi}}\lambda. * (n_z \sin\theta + n_x \cos\theta) * (n_z \cos\theta - n_x \sin\theta),\\
&b^{'o}_{21} = 3\sqrt{\frac{5}{12\pi}}\lambda. * n_y (n_z \cos\theta - n_x \sin\theta),\\
&b^{'e}_{22} = \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda. * ((n_z \sin\theta + n_x \cos\theta)^2 - n_y^2),\\
&b^{'o}_{22} = 3\sqrt{\frac{5}{12\pi}}\lambda. * (n_z \sin\theta + n_x \cos\theta) n_y
\end{aligned}
\tag{7}
$$

Rearranging, we get

$$
\begin{aligned}
&b'_{00} = b_{00},\ b'_{10} = b_{10}\cos\theta - b^e_{11}\sin\theta,\ b^{'e}_{11} = b^e_{11}\cos\theta + b_{10}\sin\theta,\ b^{'o}_{11} = b_{11},\\
&b'_{20} = b_{20} - \sqrt{3}\sin\theta\cos\theta b^e_{21} - \sqrt{\frac{5}{4\pi}}\frac{3}{2}\sin^2\theta(n_z^2 - n_x^2),\\
&b^{'e}_{21} = (\cos^2\theta - \sin^2\theta) b^e_{21} + 3\sqrt{\frac{5}{12\pi}}\sin\theta\cos\theta(n_z^2 - n_x^2),\\
&b^{'o}_{21} = b^o_{21}\cos\theta - b^o_{22}\sin\theta,\\
&b^{'e}_{22} = b^e_{22} + \cos\theta\sin\theta b^e_{21} + \sqrt{\frac{5}{12\pi}}\frac{3}{2}\sin^2\theta(n_z^2 - n_x^2),\\
&b^{'o}_{22} = b^o_{22}\cos\theta + b^o_{21}\sin\theta.
\end{aligned}
\tag{8}
$$

As shown in (8), $b'_{00}, b'_{10}, b^{'e}_{10}, b^{'o}_{11}, b^{'o}_{21}$ and $b^{'o}_{22}$ are in the form of linear combination of the 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 [1], 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 $P = 3\sqrt{\frac{5}{12\pi}}\sin^2\theta(n_z^2 - n_x^2)$ and $Q = 3\sqrt{\frac{5}{12\pi}}\sin\theta\cos\theta(n_z^2 - n_x^2)$, we have

$$
b'_{20} = b_{20} - \sqrt{3}\sin\theta\cos\theta b^e_{21} - \frac{\sqrt{3}}{2}P,
$$

$$
\begin{aligned}
b_{21}^{'e} &= (\cos^2\theta - \sin^2\theta) b_{21}^e + Q, \\
b_{22}^{'e} &= b_{22}^e + \cos\theta \sin\theta b_{21}^e + \frac{1}{2} P.
\end{aligned} \tag{9}
$$

Then

$$
\begin{aligned}
& b_{20}^{'2} + b_{21}^{'e2} + b_{22}^{'e2} \\
&= b_{20}^2 + b_{21}^{e2} + b_{22}^{e2} + \frac{3P^2}{4} - 2\sqrt{3}\sin\theta\cos\theta b_{20} b_{21}^e - \sqrt{3} b_{20} P + 3\sin\theta\cos\theta P + Q^2 \\
&\quad + 2(\cos^2\theta - \sin^2\theta) b_{21}^e Q + \frac{P^2}{4} + 2\sin\theta\cos\theta b_{22}^e b_{21}^e + b_{22}^e P + \sin\theta\cos\theta P \\
&= b_{20}^2 + b_{21}^{e2} + b_{22}^{e2} + P^2 + 4\sin\theta\cos\theta b_{21}^e P + (b_{22}^e - \sqrt{3} b_{20})(P + 2\sin\theta\cos\theta b_{21}^e) \\
&\quad + Q^2 + 2(\cos^2\theta - \sin^2\theta) b_{21}^e Q
\end{aligned}
$$

Having $b_{20}^{'2} + b_{21}^{'e2} + b_{22}^{'e2} = b_{20}^2 + b_{21}^{e2} + b_{22}^{e2}$ and $Q = P\frac{\cos\theta}{\sin\theta}$, we get

$$
P^2 + 2\sin\theta\cos\theta b_{21}^e P + (b_{22}^e - \sqrt{3} b_{20})(P\sin^2\theta + 2\sin\theta\cos\theta b_{21}^e) = 0
$$

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

The two possible roots of the polynomial gives $P = -2\sin\theta\cos\theta b_{21}^e$ or $P = -\sin^2\theta(b_{22}^e - \sqrt{3} b_{20})$. Taking $P = -2\sin\theta\cos\theta b_{21}^e$ into (9) gives $b'_{20} = b_{20}$, $b_{21}^{'e} = -b_{21}^e$, $b_{22}^{'e} = b_{22}^e$, which is apparently incorrect. Therefore, we have $P = -\sin^2\theta(b_{22}^e - \sqrt{3} b_{20})$ and $Q = -\cos\theta\sin\theta(b_{22}^e - \sqrt{3} b_{20})$. Substituting them in (9) we get

$$
\begin{aligned}
b'_{20} &= b_{20} - \sqrt{3}\sin\theta\cos\theta b_{21}^e + \frac{\sqrt{3}}{2}\sin^2\theta(b_{22}^e - \sqrt{3} b_{20}), \\
b_{21}^{'e} &= (\cos^2\theta - \sin^2\theta) b_{21}^e - \cos\theta\sin\theta(b_{22}^e - \sqrt{3} b_{20}), \\
b_{22}^{'e} &= b_{22}^e + \cos\theta\sin\theta b_{21}^e - \frac{1}{2}\sin^2\theta(b_{22}^e - \sqrt{3} b_{20}).
\end{aligned} \tag{10}
$$

Using (8) and (10), 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 (1).

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

$$
\begin{bmatrix} n''_x \\ n''_y \\ n''_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} n'_x \\ n'_y \\ n'_z \end{bmatrix} \tag{11}
$$

Repeating the above derivation easily leads to the linear equations in (2) which relates the basis images at the new rotated pose to the basis images at the old rotated pose.