Face Recognition from a Single Training Image under Arbitrary Unknown Lighting Using Spherical Harmonics

Lei Zhang, *Student Member, IEEE*, and Dimitris Samaras, *Member, IEEE*

Abstract—In this paper, we propose two novel methods for face recognition under arbitrary unknown lighting by using spherical harmonics illumination representation, which require only one training image per subject and no 3D shape information. Our methods are based on the recent result which demonstrated 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. We provide two methods to estimate the spherical harmonic basis images spanning this space from just one image. Our first method builds the statistical model based on a collection of 2D basis images. We demonstrate that, by using the learned statistics, we can estimate the spherical harmonic basis images from just one image taken under arbitrary illumination conditions if there is no pose variation. Compared to the first method, the second method builds the statistical models directly in 3D spaces by combining the spherical harmonic illumination representation and a 3D morphable model of human faces to recover basis images from images across both poses and illuminations. After estimating the basis images, we use the same recognition scheme for both methods: we recognize the face for which there exists a weighted combination of basis images that is the closest to the test face image. We provide a series of experiments that achieve high recognition rates, under a wide range of illumination conditions, including multiple sources of illumination. Our methods achieve comparable levels of accuracy with methods that have much more onerous training data requirements. Comparison of the two methods is also provided.

Index Terms—Face recognition, spherical harmonics illumination representation, 3D face morphable models, illumination invariance.

◆

1 INTRODUCTION

FACE recognition has recently received extensive attention as one of the most significant applications of image understanding [1], [2], [3]. Although rapid progress has been made in this area during the last few years, the general task of recognition remains unsolved. In general, face appearance does not depend solely on identity. It is also influenced by illumination and viewpoint. Changes in pose and illumination will cause large changes in the appearance of a face. In this paper, we propose two novel methods for face recognition under arbitrary unknown illumination by using spherical harmonics illumination representation.

In the past few years, there have been attempts to address image variation produced by changing in illumination and pose [1], [3], [4], [5]. In general, appearance-based methods like Eigenfaces [6] and SLAM [7] need a number of training images for each subject, in order to cope with pose and illumination variabilities. Georghiades et al. [8] presented a method which exploited the fact that the set of images of an object in fixed pose, is a convex cone in the space of images. By using a small number of training images per subject taken with different lighting directions, the generative model of the face can be reconstructed. For each pose, the corresponding illumination cone was approximated by a low-dimensional linear subspace and the basis vectors can be estimated using the reconstructed model. This method required a set of training images for each subject to construct the illumination cone. Zhou et al. [9], [10] extended photometric stereo algorithms to handle the appearances of the class of human faces and proposed a method of recovering albedos and surface normals from one images under unknown illumination conditions. Both their methods and ours use a bootstrap step to encapsulate texture and shape information of the class of human faces. The major difference is that, methods in [9], [10] assume that the human face is lit by a distant illumination, thus, are bound to images taken under single directional illuminant while in our methods, the spherical harmonic illumination representation removes this limitation.

Blanz and Vetter proposed a 3D face Morphable Model in [11] where each face can be represented by linear combinations of a set 3D of face exemplars. By fitting the Morphable model to the input image, the shape and texture parameters of a 3D Morphable Model can be recovered in an analysis-by-synthesis fashion. With the recovered shape and texture parameters, a face recognition scheme with high recognition rates was proposed in [12] and impressive face synthesis results were reported in [11], [13]. In our second method, we integrate a more general illumination representation into the Morphable Model approach. The method in [14], [12] applies Phong illumination model, thus is bound to images taken under directional illuminations and requires the knowledge of number of lights for images taken under multiple illuminants, which is difficult to know in most cases. In our method, the illumination variations are captured by the spherical harmonic basis, thus there is no illumination limitation on the input images.

Previous research has suggested that illumination variability in face images is low-dimensional, e.g., [15], [16], [17],

- *The authors are with the Computer Science Department, 2429 Computer Science, State University of New York, Stony Brook, NY 11794-4400. E-mail: {lzhang, samaras}@cs.sunysb.edu.*

[18], [19], [20], [5], [21]. Using spherical harmonics and signal-processing techniques, Basri and Jacobs [15] and Ramamoorthi [16] have independently 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 nine-dimensional linear subspace. Furthermore, a simple scheme for face recognition with excellent results was described in [15]. However, to use this recognition scheme, the basis images spanning the illumination space for each face are required. These images can be rendered from a 3D scan of the face or can be estimated by applying PCA to a number of images of the same subject under different illuminations [16]. Lee et al. proposed an effective approximation of this basis by nine single light source images of a face and Wang et al. [22] proposed a illumination modeling and normalization method for face recognition. All the above mentioned methods need a number of training images and/or 3D scans of the subjects in the database, requiring specialized equipment and procedures for the capture of the training set, thus limiting their applicability. A promising earlier attempt by Zhao and Chellappa [23] used symmetric shape from shading but suffered from the drawbacks of SFS.

In this paper, we introduce and compare two methods [24], [25] which perform face recognition for images under arbitrary illumination conditions. Our methods require only one training image for each subject without any illumination limitations. Our methods are based on the recent result [15], [16] which demonstrated 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. In our first method, we show that we can recover the basis images spanning this space from just one image taken under arbitrary illumination conditions when the pose is fixed. This method consists of three steps: statistical model computation (bootstrap), training, and testing. Initially, we use a *bootstrap* set consisting of 3D face models. We compute a statistical model for each of the spherical harmonics. Our statistical models aim to learn the probability density function (*pdf*) for the basis images. The heaviest data capture requirements for our method are placed in the construction of the bootstrap set. This set and the resulting pdfs for the basis images need be constructed only once. No further modifications are necessary when new subjects are added to the database, allowing for increased scalability of the recognition system. During the second step, *training*, when presented with an image of a novel face, we use the statistical model learned during the bootstrap phase to recover the corresponding basis images. Given a face, we first estimate the weight coefficients for the basis images. Then, we recover the basis images by computing the *maximum a posteriori* (MAP) estimate. We show that the set of images we recover is the set of basis images with maximum a posterior probability. For each subject in the training set, we recover the basis images that span the illumination space for this subject. The faces in the training set need not belong to the bootstrap set. The only requirement is that the bootstrap data capture the statistical characteristics of the training set. Finally, during the *testing* step, we use the recognition scheme proposed in [15]. We return the face from the training set for which there exists a weighted combination of basis images that is the closest to the test face image.

In the first method, the statistical model is based on a collection of 2D basis images which implicitly include the 3D information. Thus, for images taken across both illumination and pose variations, we need to build a statistical model for each sample pose, which is inefficient. Hence, we propose the second method which combines a 3D morphable model and spherical harmonic illumination representation to perform face recognition for both pose and illumination variations. This method also consists of three steps: construction of the 3D morphable model (bootstrap), face reconstruction and basis image rendering (training), and recognition (testing). Initially, we use a *bootstrap* set consisting of 3D face models. We construct a 3D face morphable model similar to [12]. During the second step, *training*, when presented with an image of a novel face, we compute the shape parameters of a morphable model from a shape error estimated by the displacements of a set of feature points and estimate the illumination coefficients and texture information using the spherical harmonics illumination representation. The reconstructed face models then serve as generative models that can be used to synthesize sets of basis images under novel poses and spanning the illumination field. The recognition (testing) step is similar to that of the first method: We return the face from the training set for which there exists a weighted combination of basis images that is the closest to the test face image.

For our methods, we use a collection of 3D face scans supplied by USF [11] as the bootstrap set and we use the Yale Face Database B [26] and the CMU-PIE database [28] for training and testing. We provide a series of experiments which demonstrate that our method has high recognition accuracy even though none of the subjects in the training set belongs to the bootstrap set. Our methods achieve comparable levels of accuracy with methods that have much more onerous training data requirements. In our experiments, as predicted by the theory, the basis images recovered during training are noticeably insensitive to the particular illumination of the training image, which indicates that our method should perform well on larger databases.

This paper is organized as follows: In Section 2, we explain the Spherical Harmonics and how to acquire basis images from 3D face models. In Section 3, the basis image recovery of the first method is introduced and in Section 4, the basis image recovery of the second method is introduced. In Section 5, we explain the recognition process based on the recovered basis images. In Section 6, we describe our experiments and their results and the final Section presents the conclusions and future work directions.

2 SPHERICAL HARMONICS ILLUMINATION REPRESENTATION

In this section, we will briefly explain the spherical harmonic illumination representation and how we render basis images from 3D face models using the results of [15]. Let L denote the distant lighting distribution. By neglecting the cast shadows and near-field illumination, the irradiance E is then a function of the surface normal n only and is given by an integral over the upper hemisphere Ω_n [16]: $E(n) = \int L(\omega)(n \cdot \omega)d\omega$. We then scale E by the surface albedo λ to find the radiosity I, which corresponds to the image intensity directly:

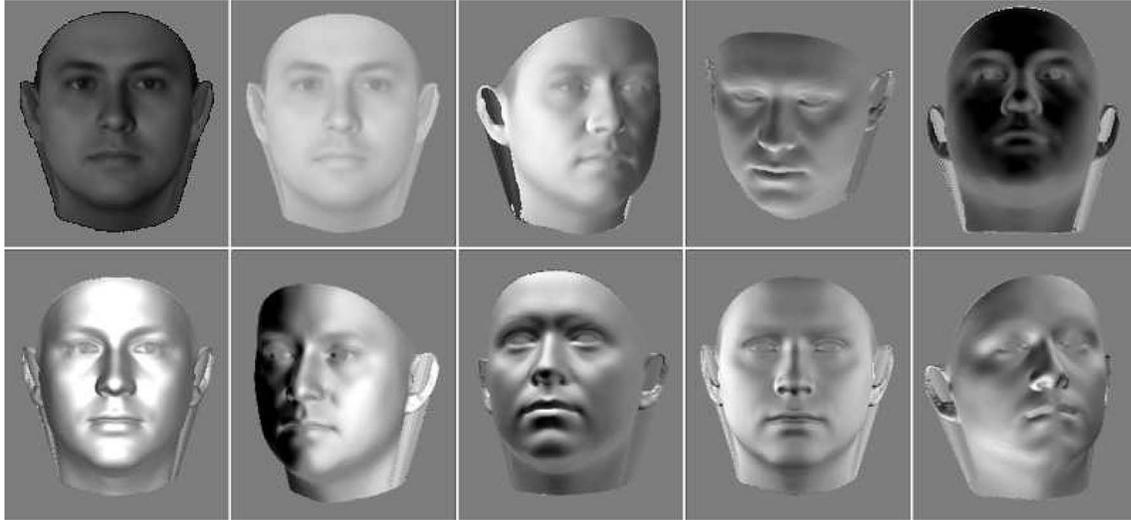$$I(p, n) = \lambda(p)E(n). \tag{1}$$

Fig. 1. The first image is the face model and the following nine images are the basis images under various view-points, represented by spherical harmonics. Lighter gray represents positive values and darker gray represents negative values.

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. They are analogue on the sphere to the Fourier basis on the line or circles. Basri and Jacobs [15] and Ramamoorthi [16] have independently shown that E can be approximated by the combination of the first nine spherical harmonics $H(x, y, z)$ for Lambertian surfaces:

$$
\begin{aligned}
&h_{00} = \frac{1}{\sqrt{4\pi}}, \quad h_{11}^o = \sqrt{\frac{3}{4\pi}}y, \quad h_{21}^o = 3\sqrt{\frac{5}{12\pi}}yz \\
&h_{10} = \sqrt{\frac{3}{4\pi}}z, \quad h_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}(2z^2 - x^2 - y^2), \\
&h_{22}^e = \frac{3}{2}\sqrt{\frac{5}{12\pi}}(x^2 - y^2), \quad h_{11}^e = \sqrt{\frac{3}{4\pi}}x, \\
&h_{21}^e = 3\sqrt{\frac{5}{12\pi}}xz, \quad h_{22}^o = 3\sqrt{\frac{5}{12\pi}}xy,
\end{aligned}
\tag{2}
$$

where the superscripts e and o denote the even and the odd components of the harmonics, respectively, and x, y, z denote the Cartesian components. Then, the image intensity of a point p with surface normal $n = (n_x, n_y, n_z)$, and albedo λ can be computed according to (1) by replacing x, y, z with n_x, n_y, n_z, as shown in (3), where we use n_{x^2} to denote a vector such that $n_{x^2,i} = n_{x,i}n_{x,i}$ for the ith voxel and define $n_{y^2}, n_{z^2}, n_{xz}, n_{yz}, n_{xy}$ similarly. We use $\lambda . * \mathbf{v}$ to denote the component-wise product of λ with any vector $\mathbf{v}$. Fig. 1 gives an example of a face model under a spherical harmonics representation.

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

3 RECOVERY OF BASIS IMAGES FROM 2D STATISTICS FOR FIXED POSE

In this section, we will explain the basis image recovery of our first method by using statistics inference in 2D image space. We will show that we can recover basis images spanning the illumination space from just one image taken under arbitrary unknown illumination conditions. First, using a bootstrap set consisting of 3D face models, we compute a statistical model for each basis image. During training, given a novel face image under arbitrary illumination, we recover a set of images for this face. We prove that these images are the set of basis images with maximum probability.

3.1 Statistical Models of Basis Images

The key equation of this method is:

$$
i(x) = b(x)^T \alpha + e(x, \alpha),
\tag{4}
$$

which states that at pixel position x, the pixel intensity $i(x)$, is the weighted combination of the basis images $b(x)$ plus an error term $e(x)$. α is the set of illumination coefficients. More precisely, let I be a d-dimensional image (in our experiments, $d = 75,972$) and let B be the $9 * d$ matrix of basis images, the columns of which are the vectors $\{b(x)\}_{x=1}^d$. Also, let α be a nine-dimensional vector denoting the coefficients of $b(x)$ and let E be a d-dimensional vector denoting the error term. Thus, we get: $I = B^T \alpha + E$. Similar to [27], our statistical models aim to learn the probability density function (pdf) for B and E. For B, we assume that the pdfs are Gaussian distributions of unknown means and covariances which we can estimate from the basis images rendered by 3D scans. In our experiments, we had 50 3D face scans with texture information in our bootstrap set. We render nine basis images per face model using equation set (3). From these basis images, we compute the sample mean vectors $\mu_b(x)$ and the sample covariance matrixes $C_b(x)$. For the ith ($i = 1..9$) basis, the component of the covariance matrix $C_{m,n}^i = cov(b_m^i, b_n^i)$ with b_j^i representing the jth ($j = 1..d$) pixel vector consisting of the jth pixels of the 50 bootstrap faces. Fig. 2 displays the mean of the statistical models of the basis images computed from our bootstrap set.
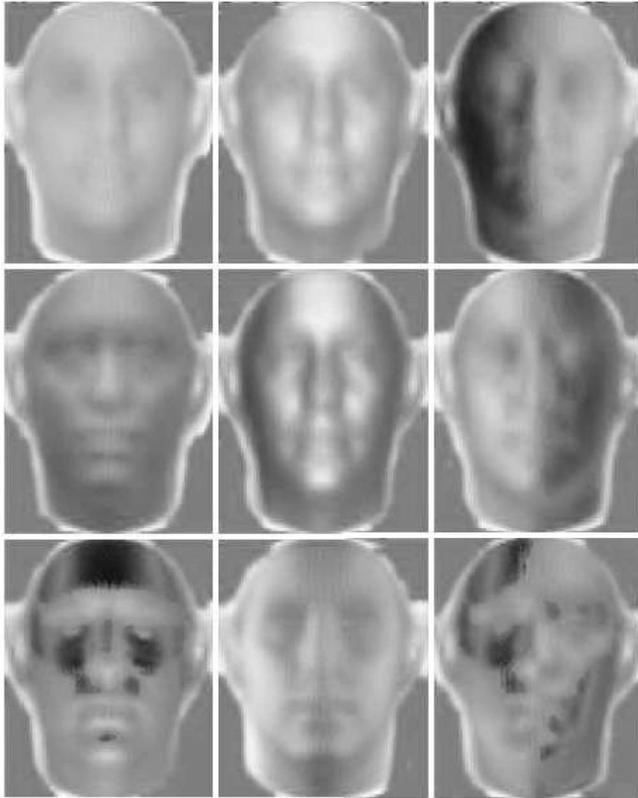
Fig. 2. The mean images of the statistical models computed from the rendered basis images.

To estimate the statistics of the error term E, we need real images with known harmonic coefficients. Unfortunately, in our current implementation, we do not have both 3D scans and real images for the subjects in our data set. Instead, we synthesized images and estimated their coefficients. Since we have 3D scans with texture information, we illuminated synthetically each 3D model, obtaining face images under a wide variety of illumination conditions. Then, we applied a least-squares method to estimate the coefficients α. Hence, we can compute the sample mean $\mu_e(x, \alpha)$ and sample variance $\sigma_e^2(x, \alpha)$ of the error term by $E = I - B^T \alpha$. This error term models the deviations from Lambertian reflectance and the errors of the low-dimensional approximation which cause the estimates of the coefficients to diverge from the true values. Since the sample mean and variance are functions of α, it is impossible to learn the statistics of error term for each possible α from a finite number of samples. Given a new image with illumination α_{tra}, the mean and variance of the error term at α_{tra} need to be interpolated using kernel regression as described in Section 3.2.1.

Our statistical models are learned for every pixel and we assume the spherical harmonics at different pixels are independent of one other, or more precisely, we recover the spherical harmonics at each pixel as if it is independent of those at other pixels. This is clearly not ideal, but it greatly makes the problem tractable.

3.2 Recovering the Basis Images

Given a training image of a novel face (possibly a different subject from those in the bootstrap) under arbitrary illumination, first we need to estimate the coefficients α in (4) and to update the error term according to the coefficients. Then, we

recover $b(x)$, the corresponding basis images, at each x by computing the *maximum a posteriori* (MAP) estimate,

$$b_{MAP}(x) = argmax_{b(x)}(P(b(x)|i(x))).$$

This is explained in detail in the following section.

3.2.1 Estimating α and the Error Term

To estimate the unknown spherical harmonic coefficients α for the given novel face image, we use kernel regression. Notice that since in the bootstrap set we have images with known coefficients, we can regard coefficient estimation as a continuous-valued classification problem. The approach is similar to the recovery of unknown illumination which is a well-studied problem [29], [27]. We first store all the K bootstrap images, $\{J_k\}_{k=1}^K$ along with their spherical harmonic coefficients $\{\alpha_k\}_{k=1}^K$. Similar to [27], given a new image, i_{tra}, we can estimate the coefficients α_{tra} as follows:

$$\alpha_{tra} = \frac{\sum_{k=1}^K w_k \alpha_k}{\left(\sum_{k=1}^K w_k\right)}, \qquad (5)$$

where $w_k = \exp\left[-\frac{1}{2}(D(i, J_k)/\sigma_k)^2\right]$ and $D(i, J_k) = \|i - J_k\|_2$ the L_2 norm, σ_k is the width of the kth Gaussian kernel which controls the influence of J_k on the estimation of coefficient α_{tra}. We can precompute all $\{\sigma_k\}_{k=1}^K$ in a way such that 10 percent of the bootstrap images are within $1 \times \sigma_k$ at each σ_k.

We described how we learn the statistics of the error term E in Section 3.1. However, as we described in Section 3.1, both the sample mean $\mu_e(x, \alpha)$ and the sample variance $\sigma_e^2(x, \alpha)$ are functions of α. It is impossible to learn the statistics for every possible α beforehand using a finite number of samples. Thus, again, given a new image, i_{tra}, we estimate the statistics of the error term $E(\alpha)$ using kernel regression. The mean and the variance of the error estimation $e_{tra}(x, \alpha_{tra})$ at α_{tra} will be interpolated from the known mean and variance of the error term at coefficients $\{\alpha_k\}_{k=1}^K$ which have been calculated. The equation used here is similar to that in (5).

To check the accuracy of our estimation of the coefficients, we synthesized 200 images using the basis images with various sets of coefficients $\{\alpha_j\}_{j=1}^{200}$. From the synthesized images, we estimated coefficients $\{\alpha_j'\}_{j=1}^{200}$ using (5). Ideally, the two sets of α_j and α_j' should be identical. We measure each estimation error by: $\epsilon_j = D(\alpha_j^{i'} - \alpha_j^i)/|\alpha_j^i|$, where $D(x, y) = \|x - y\|_2$ the L_2 norm. The mean of the estimation error is 0.0734, which shows that the estimated values are very close to the actual values.

3.2.2 Computing the Basis Images

After estimating the coefficients α, we can recover the basis images for a novel face image. As we have shown already: $b_{MAP}(x) = argmax_{b(x)}(P(b(x)|i(x)))$. It is hard to calculate $argmax_{b(x)} P(b(x)|i(x))$ directly. Using Bayes' rule:

$$P(b(x)|i(x))P(i(x)) = P(i(x)|b(x))P(b(x)). \qquad (6)$$

Since we have already estimated the spherical harmonic coefficients α, according to (4), we know that given an image $i(x)$ and known coefficients α, $P(i(x))$ is a constant. Thus,

$$b_{MAP}(x) = argmax P_{b(x)}((i(x)|b(x))P(b(x))). \qquad (7)$$

Fig. 3. The recovered basis images (bottom) compared to the rendered ones (top) for an individual subject in the bootstrap set.

In this term, $P(b(x))$ is the Gaussian probability density function we have learned previously and $P(i(x)|b(x))$ can be computed from (4). Thus, $i(x)$ is a random variable with Gaussian pdf of mean $b(x)^T \alpha + \mu_e(x, \alpha)$ and variance $\sigma_e^2(x, \alpha)$. Thus,

$$b_{MAP}(x) = argmax_{b(x)}(Gauss(b(x)^T \alpha + \mu_e, \sigma_e^2) \times Gauss(\mu_{b(x)}, C_{b(x)})). \quad (8)$$

Using log probability, and ignoring constant terms, (8) becomes:

$$b_{MAP}(x) = argmax_{b(x)}\left(-\frac{1}{2}\left(\frac{i - b^T \alpha - \mu_e^2}{\sigma_e}\right)^2 - \frac{1}{2}(b - \mu_b)^T C_b^{-1}(b - \mu_b)\right). \quad (9)$$

Then, we set the derivatives of the right side of (9) (with regards to $b(x)$) to 0 in order to get the maximum probability, we get:

$$-\frac{2}{\sigma_e^2}(i - b^T \alpha - \mu_e)\alpha + 2C_b^{-1}(b - \mu_b) = 0. \quad (10)$$

Rearranging, we can get the following linear equation:

$$A * b_{MAP} = T, \quad (11)$$

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 (11), the new set of basis images of the subject can be recovered. Fig. 3 shows the comparison of the rendered basis images and the recovered basis images. Again, green represents positive value and red for negative values. In our recovery process, we derive the nine basis images without explicit constraints to enforce consistency of a single 3D object. We will demonstrate in Section 4 that such constraints can be enforced by using a 3D face morphable model.

The problem of estimating the basis images B and the illumination coefficients α is a coupled estimation problem because it is in a bilinear form. In this method, we simplify the problem by estimating α in a prior step using kernel regression and using it consistently across all pixels to recover B. We will show in our second method (Section 4) that by acquiring the 3D information explicitly, this estimation problem can be solved iteratively.

3.2.3 Experiments on Basis Image Recovery

Since our recognition method is based on the basis images recovered during training, it is very important that we have an accurate method to recover the basis images.

We performed our experiments on Yale Database B. Following [8], we group our data set into four subsets (Table 1). Each subset contains images illuminated from a specific range of directions. Subset 4 contains the extreme illumination conditions. Please refer to [8] for more information about this grouping. Fig. 4 shows the basis images recovered from various face images under different illuminations for one subject belonging to the training set (but not the bootstrap). We found that, except for the basis images recovered from images under extreme illumination (i.e., images in Subset 4), the resulting basis images recovered from images under different illumination are very close. For each subject in the training set, we calculated the basis images using 20 images under different illumination (five of them belonging to Subset 4). The comparison of the intensity variances is reported in Table 2 where the per person mean variance of the 20 resulting sets of basis images, was 14.66 intensity levels per pixel for the full set of images and 5.34 for the images belonging to Subsets 1-3. The per person variance of the original training set images was 31.13 (24.77 for Subsets 1-3) while the cross-person mean variance of the 200 total image basis sets for all subjects was 31.47 for the full set (16.33 for Subsets 1-3). We see that the basis images, we recover exhibit much greater invariance to illumination effects.

It is interesting to study the performance of our method on images taken under multiple directional illumination sources. For comparison between the performance on images taken under a single directional illumination source and that on images taken under multiple directional illumination sources, we synthesized images by combining face images of the same subject in our data set. We randomly selected 2-4 images of the same subject from the training data set and combined them together with random weights to simulate face images under multiple directional illumination sources (12 synthetic images per person). Fig. 5 shows the recovered results of basis images from the synthesized images compared to the recovered results from a real image under a single illuminant.

3.2.4 Pose Variance

So far, we have provided experiments on the images with frontal view images. We applied our method to two sets of

TABLE 1
The Separation of the Yale Database

Subsets	1	2	3	4
Number of Images	70	120	120	140
Illumination	0-12	13-25	26-50	51-77

Fig. 4. The basis images recovered from images of the same subject under various single directional illuminants. The first column is the images we used for the recovery followed by the set of basis images. The resulting basis images recovered from images under different illumination are very close.

images with poses -12 degrees and 24 degrees azimuth from the frontal view. We rotated the 3D scans in the bootstrap set by -12 degrees and 24 degrees, respectively, and rendered basis images in the same way described previously. For training and testing, from the Yale Face Database B we selected two sets of face images which are taken approximately -12 degrees and 24 degrees (8 images per person per angle) from the frontal view, respectively, and repeated the same steps described previously. Experimental results are similar to the experimental results of our method for frontal views. Our experiments imply a solution for face recognition in the presence of both illumination and pose variance. The basic idea would be to combine face pose estimation with our method; however, we need to calculate the statistics of the basis images for each pose in order to recover the new set of basis images, which is not efficient. Moreover, for each subject in the training, we need one training image per pose to recover the basis images for each pose, which is difficult. In the following section, we will show that, the combination of a morphable model and spherical harmonic illumination representation facilitates recognition for images with variations of both pose and illumination.

4 RECOVERY OF BASIS IMAGES BY USING A MORPHABLE MODEL AND SPHERICAL HARMONICS

In this section, we will explain how we recover the shape and texture information of a training subject by combining a morphable model and spherical harmonics illumination representation.

TABLE 2
The Comparison of the Intensity Variances of the Original
Images and Basis Images on the Yale Database

Subsets	subset 1-3	full set
IV orig	24.77	31.13
IV basis	5.34	14.66

The first row shows the per person mean variance of the original images and the second row shows the per person mean variance of the resulting sets of basis images. We see that the basis images we recover exhibit much greater invariance to illumination effects.

4.1 Morphable Model

The 3D Morphable Face Model [11], [12] is a 3D model of faces with separate shape and texture models that are learned from a set of exemplar faces. Morphing between faces requires complete sets of correspondences between all of the faces. When building a 3D morphable model, we transform the shape and texture spaces into vector spaces, so that any convex combination of exemplar shapes and textures represents a realistic human face. We present the geometry of a face with a shape-vector $S = (X_1, Y_1, Z_1, X_2, ..., Y_n, Z_n)^T \in \Re^{3n}$, which contains the X, Y, Z-coordinates of its n vertices. Similarly, the texture of a face can be represented by a texture-vector $T = (R_1, G_1, B_1, R_2, ..., G_n, B_n)^T \in \Re^{3n}$, where the R, G, B texture values are sampled at the same n points. A morphable model can be constructed using a data set of m exemplar faces: exemplar i is represented by the shape-vector S_i and texture-vector T_i. New shapes s and textures t can be generated by convex combinations of the shapes and textures of the m exemplar faces: $s = \sum_{i=1}^{m} a_i S_i, t = \sum_{i=1}^{m} b_i T_i, \sum_{i=1}^{m} a_i = \sum_{i=1}^{m} b_i = 1$. To reduce the dimensionality of the shape and texture spaces, Principal Component Analysis (PCA) is applied separately on the shape and texture spaces:

$$ s = \overline{s} + \sum_{i=1}^{m-1} a_i \sigma_{s,i} s_i, \quad t = \overline{t} + \sum_{i=1}^{m-1} b_i \sigma_{t,i} t_i. \qquad (12) $$

By setting the smallest eigenvalues to zero, (12) is reformulated as:

$$ s = \overline{s} + Sa, \quad t = \overline{t} + Tb. \qquad (13) $$

In (13), the columns of S and T are the most significant eigenvectors s_i and t_i rescaled by their standard deviation and the coefficients a and b constitute a pose and illumination invariant low-dimensional coding of a face [14]. PCA also provides an estimate of the probability densities of the shapes and textures, under a Gaussian assumption: $p(s) \sim e^{-\frac{1}{2}\|a\|^2}, p(t) \sim e^{-\frac{1}{2}\|b\|^2}$.

4.2 Forward and Inverse Face Rendering

We can generate photo-realistic face images by using the morphable model, we described in Section 4.1. Here, we
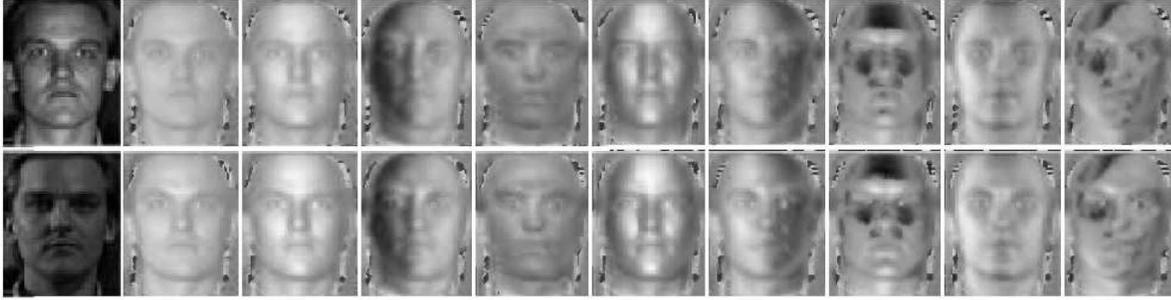
Fig. 5. The basis images recovered from an image taken under multiple illuminants (bottom) compared to those recovered from an image under single illumination (top). The first column shows the input images. We see that our method performs well on images under multiple illuminants.

describe how we synthesize a new face image from the face shape and texture vectors s and t, thus the inversion process of the synthesis is how we recover shape and texture information from the image.

Shape: Similar to [14], a realistic face shape can be generated by:

$$s_{2d} = fPR(\bar{s} + Sa + t_{3d}) + t_{2d}, \qquad (14)$$

where f is a scale parameter, P an orthographic projection matrix and R a rotation matrix with ϕ, γ, and θ the three rotation angles for the three axes. t_{3d} and t_{2d} are translation vectors in 3D and 2D, respectively. Equation (14) relates the vector of 2D image coordinates s_{2d} and the shape parameters a. For rendering, a visibility test must still be performed by using a z-buffer method [30].

For a training image, inverting the rendering process, the shape parameters can be recovered from the shape error: If f, ϕ, γ, and θ are kept constant, the relation between the shape s_{2d} and a is linear according to Equation (14):

$$\frac{\partial s_{2d}}{\partial a} = fPRS. \qquad (15)$$

Thus, updating a from a shape error δs_{2d} requires only the solution of a linear system of equations. In our method, the shape error is estimated by the displacements of a set of manually picked feature points s_f [31] corresponding to image coordinates s_f^{img}. The shape reconstruction goes through the following steps.

Model Initialization: All the parameters are initialized in this step. Shape parameter a is set to 0 and pose parameters f, ϕ, γ, θ, and t_{2d} are initialized manually. We do not need to know the illumination conditions of the training image, unlike [14].

Feature Correspondence: For the set of prepicked feature points in the morphable model, we find the correspondence s_f^{img} in the training image semiautomatically. The set of feature points contains major and secondary features, see Fig. 6. After the correspondences of major features are manually set, the secondary features are updated automatically.

Rotation, Translation, and Scale Parameters Update: The parameters f, ϕ, γ, and θ can be recovered by using a Levenberg-Marquardt optimization [32] to minimize the error between s_f^{img} and the model feature points [33]:

$$argmin_{f,\phi,\gamma,\theta,t_{2d}} \| s_f^{img} $$
$$- (fPR(\bar{s}_f + S_f a + t_{3d}) + t_{2d}) \|^2 = (\tilde{f}, \tilde{\phi}, \tilde{\gamma}, \tilde{\theta}, \tilde{t_{2d}}), \qquad (16)$$

where $\bar{s}_f$ and S_f are the corresponding shape information of the feature points in the morphable model in (13).

Shape Parameter Update: The shape error of the feature points, δs_f^{2d}, is defined as the difference between s_f^{img} and the new shape information of feature points in the model that was rendered by recovered parameters $\tilde{f}, \tilde{\phi}, \tilde{\gamma}, \tilde{\theta}$, and $\tilde{t}_{2d}$. Thus, the vector of shape parameters a can be updated by solving a linear system of equations with constraint $\sum_{i=1}^m a_i = 1$:

$$\delta s_f^{2d} = fPRS_f \delta a. \qquad (17)$$

Texture: For texture information recovery, most of the previous methods [27], [8], [14] are applicable to images taken under single light source, which limits their applicability. Here, we propose a method which performs texture fitting to a training image and has no limitation in the image illumination conditions.

According to (1) and (2), the texture of a face can be generated by:

$$t = B * \alpha, \quad B = H(n_x, n_y, n_z) \cdot \lambda, \qquad (18)$$

where H is the spherical harmonics representation of the reflectance function (2) and α is the vector of illumination coefficients. Hence, if we know the illumination coefficients, the texture information is only dependent on image intensity t and surface normal n, which can be computed from the 3D shape we recovered during the shape fitting step. Thus, the estimation of illumination coefficients is crucial. Given input image t_{tra}, we compute the texture and illumination coefficients iteratively as following:

1. Step 0: Initialize the texture parameter b_0 as 0 and define the initial albedo $\lambda_0 = \bar{t} + Tb_0 = \bar{t}$. With the recovered shape information, we first compute the surface normal n for each vertex. Then, the first nine basis images B and spherical harmonics $H(n)$ for reflectance function can be computed according to (18) and (2), respectively. Set step index i=1 and set $\eta = 0.5$.

2. For each Step i, estimate the set of illumination coefficients by solving a linear equation:

$$t_{tra} = B_{cur}\alpha_i. \qquad (19)$$

3. As described above, the new albedo λ_i can be directly computed by solving: $t_{tra} = H(n_x, n_y, n_z)\alpha_i \cdot \lambda_i$. However, the illumination coefficients are computed using the mean texture, thus not accurate. Since the texture is dependent on both current texture and illumination coefficients, we compute the new texture

Fig. 6. Recovery Results: Images in the first row are the input training images, those in the second row are the initial ttings, the third row shows images of the recovered 3D face model and the last row gives the illuminated rotated face models. In the first column, the black points are prepicked major features, the white points are the corresponding features and the points lying in the white line are secondary features.

parameters b_i by solving $\lambda'_i = \bar{t} + Tb_i$ with constraint $\sum_{i=1}^{m} b_i = 1$, where $\lambda'_i = (1 - \eta)\lambda_{i-1} + \eta(t_{tra}/(H(n_x, n_y, n_z)\alpha_i))$. Then, we update the albedo $\lambda_i = \bar{t} + Tb_i$.

4. Perform Steps 2 and 3 iteratively until convergency for the current η.

5. Increase η by 0.1, set $b_0 = b_i$, $i = 0$, and $\lambda_0 = \bar{t} + Tb_0$ and perform steps 2) to 4) iteratively until η reaches 1. The final albedo is then directly computed using the estimated α: $\lambda = (t_{tra}/(H(n_x, n_y, n_z)\alpha))$.

In the above algorithm, the weight η is used to trade off the prior probability and the fitting quality. Instead of using texture parameters [14], we estimate the albedo value for each vertex, which will be used for basis image rendering and recognition. For occluded vertices, texture information is estimated through facial symmetry. Fig. 6 shows the results of our method.

4.3 Basis Images Rendering

For each training subject, we recover a 3D face model using the algorithm described in the previous section. The recovered face models serve as generative models to render basis images. In this section, for each subject, a set of basis images across poses are generated, to be used during recognition. We sample the pose variance for each 5 degrees in both azimuth and altitude axes. In our experiments, the range of azimuth is [-70, 70] and the range of altitude is [-10, 10]. Fig. 7 shows a subset of the basis images for one subject.

4.4 Experiments on Basis Image Recovery

Similar to our first method, we perform experiments to evaluate the illumination invariance of the basis image recovery. Fig. 8 shows four sets of rendered basis images recovered from various face images under different illuminations and poses for one subject. The resulting basis images rendered from images under different illumination are very close. For each subject, we calculated 10 sets of basis using 10 training images under different illumination. The per person mean variance of the 10 resulting sets of basis images was 3.32. For comparison, per person variance of the original training images was 20.25. That means the rendered basis images have much greater invariance to illumination effects than original images.

Fig. 7. A subset of the rendered basis images across poses.

5 RECOGNITION EXPERIMENTS AND RESULTS

We perform recognition based on [15]. During testing, we recognize the face for which there exists a weighted combination of basis images that is the closest to the test face image. Here, B is the set of basis images with size $r * d$, d is the number of points in the image, and r the number of basis images used (nine is a natural choice as discussed in [15]). Every row of B contains one spherical harmonic image and the rows of B form a basis for the linear subspace. We can simply apply QR decomposition to B to obtain an orthonormal basis. Thus, we compute the distance from the test image I and the space spanned by B as $\|QQ^T I - I\|$.

We performed our experiments on Yale Face Database B [26] and CMU-PIE database [28]. The Yale Database contains images of 10 people at nine poses and 64 illuminations per pose. We used 45*10 face images for 10 subjects in a single pose with each subject having 45 face images taken under different directional light sources. The CMU-PIE database

contains 68 individuals. We performed experiments on a set of 4,488 images which contains 68 subjects, three poses for each subject, and 22 different illuminations for each pose. None of Yale Database B and CMU-PIE databse is also in the USF set used to compute the statistical model and the morphable model.

5.1 Experiments on Frontal Images

For training and testing, we first used the Yale Face Database B which contains faces that do not belong to our bootstrap set. Despite its relatively small size, this database provides images that sample sufficiently the whole illumination space and has therefore become a testing standard for variable illumination recognition methods.

For the first method, we report two sets of experimental results in Table 3. For the first set, we use the basis images recovered from randomly selected training images, excluding extreme illumination conditions. For the second set, we perform recognition using the basis images recovered from
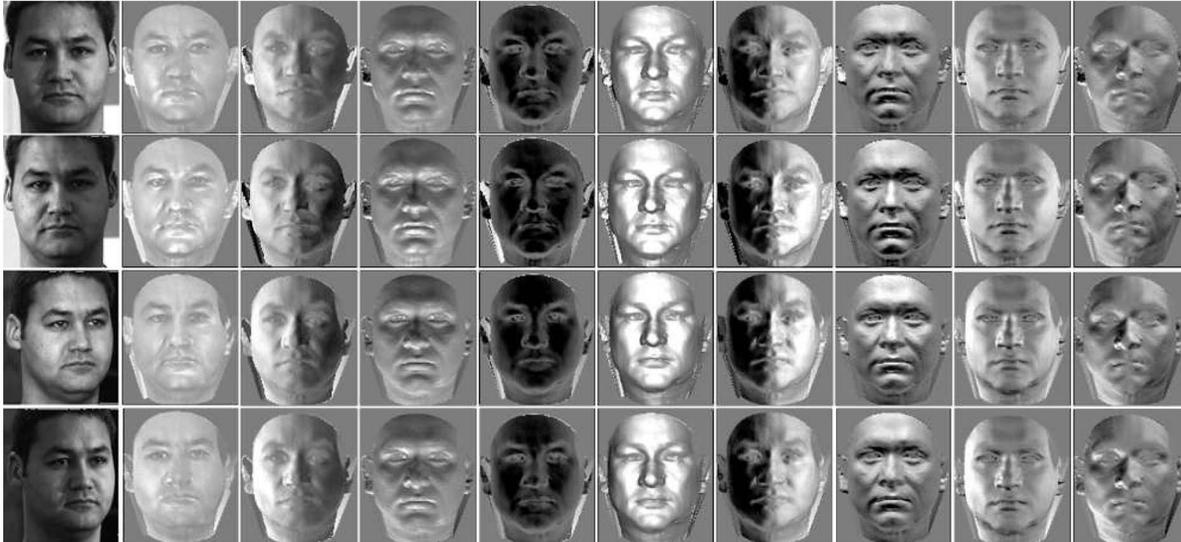
Fig. 8. Rendered basis images (wrapped to frontal view) from training images taken under different illumination conditions and poses. The first column shows the training images. We see that the basis images we recover exhibit much greater invariance to illumination effects.

the training images captured only under extreme illumination. We tested 400 face images and the error rates are shown in Table 3.

For the first method, we also performed recognition experiments based on the basis images we recovered from training images under multiple illuminants. Error rates were very close to the single illuminant training set, e.g., 0.3 for test images under multiple illumination, 3.3 for test images from Subset 4.

The comparison between our methods and other methods [34] on Yale Database B is shown in Table 4. In order to compare with other methods, we randomly select face images to recover the corresponding spherical harmonic images and perform recognition on 400 images. We repeated the process 10 times and calculated the average recognition rates. All the methods in Table 4 which have an offline training process require all the images of Subset 1 and 2 for training, while we require only one image per subject for training. The method of Nine Points of Light (9PL) [34] does not have offline training, but still needs nine images per subject for training. With our methods, training on one image randomly selected from Subsets 1 and 2 and testing on images from Subsets 1-3, we obtain perfect results and even when we test on images from Subset 4, we still achieve high accuracy. As can be seen from Table 4, the results from our methods are comparable with methods that require extensive training data per subject even though our method requires only one training image.

5.2 Experiments on Images of Arbitrary Pose and Illumination

In our experiments, we used the CMU-PIE database which provides images of both pose and illumination variation. We used only one image per subject to recover the 3D face model. We used the front and side galleries for training and all three pose galleries for testing. Notice that training images can have very different illumination conditions (unlike [14]). We performed recognition by using both all the nine basis images and the first four basis images. We report our experimental results and comparison to LiST [14] in Table 5. From the experimental results, we find that our method gives good recognition rates. When the poses of training and testing images are very different, our method is not as good as [14] because we only used a set of feature points to recover the shape information, which is rather rough and can be improved using flow computations.

We also performed our first method on a subset of CMU-PIE database. We used a subset of 1,496 images which contains 68 subjects and 22 different illuminations for each pose. The results and the comparison of two methods are also reported in Table 5. We can see that for images taken under

TABLE 3
The First Row of "Non-Ext" Is the Error Rates Achieved by Using the Set of Basis Images Recovered from Randomly Selected Training Images (Eight per Person) from Subsets 1, 2, and 3, and the Second Row ("Extreme") Is the Error Rates by Using Basis Images Recovered from Randomly Selected Training Images (Six per Person) from Subset 4

	Subset1	Subset2	Subset3	Subset4
Non-Ext	0.0	0.0	0.3	3.1
Extreme	8.0	7.5	7.5	6.8

The columns denote the testing sets.

TABLE 4
Error Rates of Recognition Using
Various Previous Methods and Our Method

Methods	Subset1,2	Subset3	Subset4
Correlation	0.0	23.3	73.6
Eigenfaces	0.0	25.8	75.7
Linear Subspace	0.0	0.0	15.0
Cones-attached	0.0	0.0	8.6
9PL	0.0	0.0	2.8
Cones-cast	0.0	0.0	0.0
Gradient Angle	0.0	0.0	1.4
Our method 1	0.0	0.3	3.1
Our method 2	0.0	0.0	2.8

Except for our method and the method of Gradient Angle, the data were taken from [34]. The result of Gradient Angle method was taken from [35] where only images under frontal illumination was used as a training image.

TABLE 5
Recognition Results and Comparison: The First Column Lists the Light Numbers and the
Following Two Columns List the Recognition Rate for Each Pose

Light	Front Gallery Using all 9 basis			Front Gallery Using first 4 basis			Side Gallery Using all 9 basis			Side Gallery Using first 4 basis		
	Front	Side	Profile	Front	Side	Profile	Front	Side	Profile	Front	Side	Profile
1	95	89	51	89	81	49	91	92	52	79	78	52
2	89	81	34	79	73	31	80	83	34	67	67	33
3	97	88	44	89	79	42	92	96	50	83	86	48
4	98	91	52	89	83	50	94	96	62	85	88	55
5	99	89	57	89	84	52	94	97	64	87	90	59
6	100	92	55	91	86	50	100	100	64	89	89	59
7	99	95	54	88	83	51	92	96	60	86	89	58
8	99	93	62	94	89	54	94	99	70	84	85	62
9	100	96	61	90	88	55	95	100	71	83	92	64
10	100	97	60	92	87	56	98	100	69	89	88	64
11	100	98	58	88	89	50	97	95	63	90	90	63
12	98	99	61	90	88	57	94	98	72	84	86	69
13	99	93	55	89	88	50	98	99	70	87	89	63
14	100	94	53	91	86	49	95	100	62	91	92	58
15	100	93	54	91	87	49	98	99	61	89	89	52
16	99	91	53	91	82	49	95	97	60	87	89	59
17	98	92	55	91	80	50	96	99	65	89	91	61
18	95	88	52	90	78	47	92	94	62	82	87	58
19	98	90	56	89	81	51	92	95	61	82	91	55
20	99	94	58	88	80	51	93	97	63	86	88	57
21	99	96	51	90	81	50	96	96	55	84	89	54
22	99	95	62	89	81	56	94	99	65	79	90	60
mean	98.2	92.4	54.5	89.4	83.4	50.0	94.4	96.7	61.6	84.6	87.4	57.4
LiST mean	97	91	60				93	96	71			
Method1	94.6			87.1								

The recognition rates of the LiST algorithm are taken from [14].

TABLE 6
Experimental Results of Images under Multiple Directional Illumination and We Can See
Our Method Also Perform Well under Multiple Sources of Arbitrary Direction

	Train:s; Test:s	Train:m; Test:s	Train:s; Test:m	Train:m; Test:m
Train:F; Test:F	98.2	98.3	97.8	98.1
Train:F; Test:D	92.4	92.0	91.5	92.2
Train:D; Test:F	94.4	93.6	94.2	94.8
Train:D; Test:D	96.7	95.9	96.3	96.1

"s" denotes images under single directional lighting and "m" denotes synthesized images under multiple illumination. "F" denotes the front gallery and "D" denotes the side gallery.

fixed pose, our two methods both perform well and the first method is simpler and more automatic. However, the first method cannot perform recognition for images of arbitrary pose and illumination given one single training image per subject. For Yale database, the performances of both our methods are similar and for PIE database, the second method performs slightly better than the first method because that the frontal images in the PIE database are not well-aligned and our first method requires accurate alignments while the second method estimates the parameters related to pose.

Again, we study the performance of our method on images taken under multiple directional illumination sources to test our method under arbitrary illuminations. We synthesized images by combining face images in our data set and performed experiments on front and side galleries. For each subject, we randomly selected 2-6 images from the training data set and combined them together with random weights to simulate face images under multiple directional illumination sources (16 images per subject). We did experiments on the

synthesized images both during training step and testing step. Table 6 shows the experimental results and we can see that our method also performed equally well under multiple sources of arbitrary direction.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we propose two novel methods for face recognition under arbitrary illumination conditions. We have demonstrated that by using statistical models, we can recover spherical harmonic basis images spanning the illumination space from a single image taken under arbitrary unknown illumination conditions. Experimental results indicated that our methods' recognition rates were comparable to other methods for images under single illumination; moreover, our methods performed as well with multiple illuminants, which was not handled by most previous methods. During the training phase, we need only one image per subject taken under general illumination to recover the basis images. Thus,

the training set can be expanded easily with new subjects. Such scalability is desirable in Face Recognition Systems. In our experiments, as predicted by the theory, the basis images recovered during training are noticeably insensitive to the particular illumination of the training image, which indicates that our methods should perform well on much larger databases, than the ones available to us now.

In our first method, we compute the statistics of spherical images in 2D space and in the second method, the statistics of the spherical images is computed from the statistics of the face structures in 3D. From the results we have, for face recognition of images under certain fixed pose, both methods have similar performance while the second method requires more human interactivity and computation time. For the recognition task of images across both pose and illumination variations, the first method requires a large set of training images across different poses for each subject, while our second method performs well requiring only one single training images per subject.

In our experiments, we tested both images under single and multiple-directional illuminations. At this time, there exist relatively few publicly available sets of images of faces under arbitrary illumination conditions, so we plan to continue validation of our method with a database with more types of light sources, e.g., area sources. In our second method, there is human interactivity in the initialization of the model and the feature correspondences. We plan to integrate head pose estimation methods [36], [37], [38] for model initialization and optical flow algorithms for shape error estimation [39]. In the face recognition phase where our method currently needs to search the whole pose space, we expect great speed-up with a prefiltering process (again, using face pose estimation algorithms) to narrow the search space.
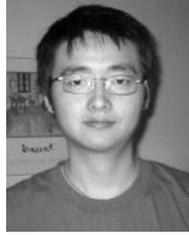
ACKNOWLEDGMENTS

REFERENCES

[1]   R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE,* vol. 83, no. 5, pp. 705-740, 1995.
[2]   A.P. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 107-119, Jan. 2000.
[3]   W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," Technical Report CAR-TR-948, Univ. of Maryland, CfAR, 2000.
[4]   R. Gross, S. Baker, I. Matthews, and T. Kanade, "Face Recognition across Pose and Illumination," *Handbook of Face Recognition,* Springer-Verlag, 2004.
[5]   A. Shashua and R.T. Riklin, "The Quotient Image: Class Based Re-Rendering and Recognition with Varying Illuminations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 129-139, Feb. 2001.
[6]   M. Turk and A.P. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-96, 1991.
[7]   H. Murase and S.K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance," *Int'l J. Computer Vision,* vol. 14, no. 1, pp. 5-24, 1995.
[8]   A.S. Georghiades and P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.
[9]   S.K. Zhou, R. Chellappa, and D.W. Jacobs, "Characterization of Human Faces under Illumination Variations Using Rank, Integrability, and Symmetry Constraints," *Proc. European Conf. Computer Vision,* 2004.
[10]   S. Zhou and R. Chellappa, "Illuminating Light Field: Image-Based Face Recognition across Illuminations and Poses," *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* 2004.
[11]   V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D-Faces," *Proc. SIGGRAPH,* 1999.
[12]   V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 9, pp. 1063-1074, Sept. 2003.
[13]   V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel, "Exchanging Faces in Images," *Proc. EuroGraphics,* 2004.
[14]   S. Romdhani, V. Blanz, and T. Vetter, "Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions," *Proc. European Conf. Computer Vision,* 2002.
[15]   R. Basri and D.W. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 218-233, Feb. 2003.
[16]   R. Ramamoorthi, "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 10, Oct. 2002.
[17]   P.W. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *Proc. Computer Vision and Pattern Recognition,* pp. 995-999, 1994.
[18]   P.N. Belhumeur and D.J. Kriegman, "What Is the Set of Images of an Object under All Possible Illumination Conditions," *Int'l J. Computer Vision,* vol. 28, no. 3, pp. 245-260, 1998.
[19]   Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 721-732, July 1997.
[20]   A. Shashua, "Illumination and View Position in 3D Visual Recognition," *Proc. Advances in Neural Information Processing Systems 4,* 1991.
[21]   S.Z. Li, J. Yan, X. Hou, Z. Li, and H. Zhang, "Learning Low Dimensional Invariant Signature of 3-D Object under Varying View and Illumination from 2-D Appearances," *Proc. Int'l Conf. Computer Vision,* pp. 635-640, 2001.
[22]   H. Wang, S.Z. Li, Y. Wang, and W. Zhang, "Illumination Modeling and Normalization for Face Recognition," *Proc. ICCV Workshop on Analysis and Modeling of Faces and Gestures,* 2003.
[23]   W.Y. Zhao and R. Chellappa, "Illumination-Insensitive Face Recognition Using Symmetric Shape-from-Shading," *Proc. Computer Vision and Pattern Recognition,* vol. 1, pp. 286-293, 2000.
[24]   L. Zhang and D. Samaras, "Face Recognition under Variable Lighting Using Harmonic Image Exemplars," *Proc. Computer Vision and Pattern Recognition,* vol. 1, pp. 19-25, 2003.
[25]   L. Zhang and D. Samaras, "Pose Invariant Face Recognition under Arbitrary Unknown Lighting Using Spherical Harmonics," *Biometric Authentication,* vol. 1, pp. 10-23, 2004.
[26]   P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class-Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.
[27]   T. Sim and T. Kanade, "Combining Models and Exemplars for Face Recognition: An Illuminating Example," *Proc. CVPR Workshop Models versus Exemplars in Computer Vision,* 2001.
[28]   T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 46-51, 2002.
[29]   R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah, "Shape from Shading: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 690-706, Aug. 1999.
[30]   J. Foley and A. van Dam, "Fundamentals of Interactive Computer Graphics," *The Systems Programming Series,* Addison-Wesley, 1984.

[31] B.W. Hwang, V. Blanz, T. Vetter, and S.W. Lee, "Face Reconstruction from a Small Number of Feature Points," *Proc. Int'l Conf. Pattern Recognition,* vol. 2, pp. 842-845, 2000.

[32] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge Univ. Press, 1992.

[33] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, 2000.

[34] K.C. Lee, J. Ho, and D.J. Kriegman, "Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting," *Proc. Computer Vision and Pattern Recognition,* vol. 1, pp. 519-526, 2001.

[35] H.F. Chen, P.N. Belhumeur, and D.W. Jacobs, "In Search of Illumination Invariants," *Proc. Computer Vision and Pattern Recognition,* pp. 1254-1261, 2000.

[36] H. Wu, T. Shimada, Q. Chen, K. Chihara, and T. Shioyama, "A Robust Algorithm for 3D Head Pose Estimation," *Proc. Int'l Conf. Pattern Recognition,* 1998.

[37] J. Sherrah and S. Gong, "Fusion of 2D Face Alignment and 3D Head Pose Estimation for Robust and Real-Time Performance," *Proc. IEEE Int'l Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems,* 1999.

[38] V. Kruger and G. Sommer, "Gabor Wavelet Networks for Efficient Head Pose Estimation," *Image and Vision Computing,* vol. 20, nos. 9-10, pp. 665-672, 2002.

[39] J.R. Bergen and R. Hingorani, "Hierarchical Motion-Based Frame Rate Conversion," technical report, David Sarnoff Research Center, Princeton, N.J., 1990.

Lei Zhang received the BSc degree in computer science from Nanjing University in 1999 and the MSc degree in computer science from the State University of New York at Stony Brook in 2003. Currently, he is a PhD candidate in the Department of Computer Science at the State University of New York at Stony Brook. He specializes in face synthesis and recognition under both illumination and pose conditions. He is also interested in biomedical image analysis. He is a student member of the IEEE.

Dimitris Samaras received a diploma degree in computer science and engineering from the University of Patras in 1992, the MSc degree in computer science from Northeastern University in 1994, and the PhD degree from the University of Pennsylvania in 2001. Since September 2000, he has been an assistant professor in the Department of Computer Science at the State University of New York at Stony Brook. He specializes in deformable model techniques for 3D shape estimation and motion analysis, illumination modeling and estimation for recognition and graphics, and biomedical image analysis. He is a member of the ACM and IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.