Face Relighting from a Single Image under Arbitrary Unknown Lighting Conditions

Yang Wang, *Member*, *IEEE*, Lei Zhang, Zicheng Liu, *Senior Member*, *IEEE*, Gang Hua, *Member*, *IEEE*, Zhen Wen, Zhengyou Zhang, *Fellow*, *IEEE*, and Dimitris Samaras, *Member*, *IEEE*

Abstract—In this paper, we present a new method to modify the appearance of a face image by manipulating the illumination condition, when the face geometry and albedo information is unknown. This problem is particularly difficult when there is only a single image of the subject available. Recent research demonstrates 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 using a spherical harmonic representation. Moreover, morphable models are statistical ensembles of facial properties such as shape and texture. In this paper, we integrate spherical harmonics into the morphable model framework by proposing a 3D spherical harmonic basis morphable model (**SHBMM**). The proposed method can represent a face under arbitrary unknown lighting and pose simply by three low-dimensional vectors, i.e., shape parameters, spherical harmonic basis parameters, and illumination coefficients, which are called the SHBMM parameters. However, when the image was taken under an extreme lighting condition, the approximation error can be large, thus making it difficult to recover albedo information. In order to address this problem, we propose a subregion-based framework that uses a Markov random field to model the statistical distribution and spatial coherence of face texture, which makes our approach not only robust to extreme lighting conditions, but also insensitive to partial occlusions. The performance of our framework is demonstrated through various experimental results, including the improved rates for face recognition under extreme lighting conditions.

Index Terms—Face synthesis and recognition, Markov random field, 3D spherical harmonic basis morphable model, vision for graphics.

1 Introduction

Recovering the geometry and texture of a human face from images remains a very important but challenging problem, with wide applications in both computer vision and computer graphics. One typical application is to generate photorealistic images of human faces under arbitrary lighting conditions [30], [9], [33], [12], [25], [24]. This problem is particularly difficult when there is only a single image of the subject available. Using spherical harmonic representation [2], [28], 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 by a low-dimensional linear subspace. In this paper, we propose a new framework to estimate lighting, shape, and albedo of a human face from a single image, which can even be taken under extreme lighting conditions and/or with partial occlusions. The proposed method includes two parts. The first part is the 3D spherical harmonic basis morphable model (SHBMM), an integration of spherical harmonics into the morphable model framework. As a result, any face under arbitrary pose and illumination conditions can be represented simply by three low-dimensional vectors: shape parameters, spherical harmonic basis parameters, and illumination coefficients, which are called the SHBMM parameters. Therefore, efficient methods can be developed for both face image synthesis and recognition. Experimental results on public databases, such as the Yale Face Database B [17] and the CMU-PIE Database [36], show that using only a single image of a face under unknown lighting, we can achieve high recognition rates and generate photorealistic images of the face under a wide range of illumination conditions.

However, when the images are taken under extreme lighting conditions, the approximation error can be large [2], which remains an unsolved problem for both face relighting and face recognition. Furthermore, this problem becomes even more challenging in the presence of cast shadows, saturated areas, and partial occlusions. Therefore, in the second part, we propose a subregion-based approach using Markov random fields to refine the lighting, shape, and albedo recovered in the first stage. Since lighting in smaller image regions is more homogeneous, if we divide the face image into smaller regions and use a different set of face model parameters for each region, we can expect the overall estimation error to be smaller than a single holistic approximation. There are, however, two main problems with such a region-based approach. First, if the majority of the pixels in a

- *Y. Wang is with Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540. E-mail: wangy@cs.cmu.edu.*
- *L. Zhang is with the Computer Science Department, Stony Brook University, 731 Lexington Ave., New York, NY 10022. E-mail: lzhang@cs.sunysb.edu.*
- *Z. Liu and Z. Zhang are with Microsoft Research, One Microsoft Way, Redmond, WA 98052. E-mail: {zliu, zhang}@microsoft.com.*
- *G. Hua is with Microsoft Live Labs Research, Microsoft Corporate, One Microsoft Way, Redmond, WA 98052. E-mail: ganghua@microsoft.com.*
- *Z. Wen is with IBM T.J. Watson Research Center, 19 Skyline Dr., Hawthorne, NY 10532. E-mail: zhenwen@us.ibm.com.*
- *D. Samaras is with the Department of Computer Science, Stony Brook University, 2429 Computer Science, Stony Brook, NY 11794-4400. E-mail: samaras@cs.sunysb.edu.*

Manuscript received 12 Oct. 2007; revised 2 May 2008; accepted 15 Sept. 2008; published online 3 Oct. 2008.
Recommended for acceptance by K. Kutulakos.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2007-10-0702.
Digital Object Identifier no. 10.1109/TPAMI.2008.244.

region are problematic (e.g., they are in cast shadows, saturated, or there are large lighting estimation errors), the albedo information in that region cannot be correctly recovered. Second, the estimated albedo may not be consistent across regions. To address both problems, we introduce neighboring coherence constraints to the albedo estimation, which also leads to a natural solution for partial occlusions. Basically, the estimation of the model parameters of each region depends not only on the observation data but also on the estimated model parameters of its neighbors. As is well known in other fields such as super-resolution and texture synthesis [16], [50], Markov random fields (MRFs) are effective to model the spatial dependence between neighboring pixels. Therefore, we propose an MRF-based energy minimization framework to jointly recover the lighting, the shape, and the albedo of the target face. Compared to previous methods, the contributions of our work include: 1) We divide an image into smaller regions and use an MRF-based framework to model the spatial dependence between neighboring regions and 2) we decouple the texture from the geometry and illumination models to enable a spatially varying texture representation thus being able to handle challenging areas such as cast shadows and saturated regions, and being robust to extreme lighting conditions and partial occlusions as well.

Empowered by our new approach, given a single photograph of a human face, we can recover the lighting, shape, and albedo even under extreme lighting conditions and/or partial occlusions. We can then use our relighting technique to generate face images under a novel lighting environment. The proposed face relighting technique can also be used to normalize the illumination effects in face recognition under varying illumination conditions, including multiple sources of illumination. The experimental results further demonstrate the superb performance of our approach.

The remainder of this paper is organized as follows: We describe the related work in Section 2 and briefly review two important approaches for the face shape and texture recovery in Section 3: the 3D morphable model [5] and the spherical harmonic illumination representation [2], [28]. After that, the 3D spherical harmonic basis morphable model is proposed in Section 4 by integrating spherical harmonics into the morphable model framework, whose performance on face relighting and recognition is demonstrated in Section 5. In order to handle extreme lighting conditions, an MRF-based framework is proposed in Section 6 to improve the shape, albedo, and illumination estimation from the SHBMM-based method. Experimental results, along with the implementation details, on face image synthesis and recognition are presented in Section 7. Furthermore, to clarify the differences between SHBMM- and MRF-based methods, a comparison is included in Section 8. Finally, we conclude our paper and discuss future work in Section 9.

2 Related Work

Inverse rendering is an active research area in both computer vision and computer graphics. Despite its difficulty, great progress has been made in generating photorealistic images of objects including human faces [12], [43], [13] and face recognition under different lighting conditions [1], [32], [49], [17], [21], [37]. Marschner et al. [25], [26] measured the geometry and reflectance field of faces from a large number of image samples in a controlled environment. Georghiades et al. [17] and Debevec et al. [12] used a linear combination of basis images to represent face reflectance. Ramamoorthi and Hanrahan [29] presented a signal processing framework for inverse rendering which provides analytical tools to handle general lighting conditions.

Furthermore, Sato et al. [34] and Loscos et al. [23] used the ratio of illumination to modify the input image for relighting. Interactive relighting was achieved in [23], [43] for certain point light source distributions. Given a face under two different lighting conditions, and another face under the first lighting condition, Riklin-Raviv and Shashua [30] used the color ratio (called the quotient image) to generate an image of the second face under the second lighting condition. Wang et al. [41] used self-quotient images to achieve good face recognition performance under varying lighting conditions. Stoschek [38] combined the quotient image with image morphing to generate relit faces under continuous changes of poses. Recently, Liu et al. [22] developed a ratio image technique to map one person's facial expression details to other people's faces. One essential property of the ratio image is that it can capture and transfer the texture details to preserve photorealistic quality.

Because illumination affects face appearance significantly, illumination modeling is important for face recognition under varying lighting. In recent years, there has been a lot of work in the face recognition community addressing face image variation due to illumination changes [48], [10]. Georghiades et al. [17] presented a method using the illumination cone. Sim and Kanade [37] proposed a model and exemplar-based approach for recognition. In both [17] and [37], there is a need to reconstruct 3D face information for each subject in the training set so that they can synthesize face images in various lighting to train the face recognizer. Blanz et al. [5] recovered the shape and texture parameters of a 3D morphable model in an analysis-by-synthesis fashion. These parameters were then used for face recognition [5], [31] and face image synthesis [7], [6]. The illumination effects are modeled by the Phong model [15].

Generally, in order to handle the illumination variability, appearance-based methods such as Eigenfaces [39] and AAM [11], [27] need a number of training images for each subject. Previous research suggests that the illumination variation in face images is low-dimensional, e.g., [1], [2], [4], [28], [14], [18]. Using the spherical harmonic representation of Lambertian reflection, Basri and Jacobs [2] and Ramamoorthi and Hanrahan [28] have obtained a theoretical derivation of the low-dimensional space. Furthermore, a simple scheme for face recognition with excellent results is presented in [2], and an effective approximation of these bases by nine single light source images of a face is reported in [21]. However, to use these recognition schemes, the basis images spanning the illumination space for each face are required. Zhao and Chellappa [49] used symmetric shape-from-shading. It suffers from the general drawbacks of shape-from-shading approach such as the assumption of point light sources. Zhang and Samaras [46] proposed to recover the nine spherical harmonic basis images from the input image. It requires a bootstrap step to estimate a statistical model of the spherical harmonic basis images. Another recent method proposed by Lee et al. [20]

used a bilinear illumination model to reconstruct a shape-specific illumination subspace. However, it requires a large data set collected in a well-controlled environment in order to capture the wide variation of the illumination conditions.

3 Face Shape and Texture Recovery

In this section, we will briefly describe the 3D morphable model [5] and the spherical harmonic illumination representation [2], [28]. In the following sections, we present a new framework to recover the shape, texture, and illumination from an input face image. The proposed framework includes two parts: 1) a 3D **SHBMM** by integrating spherical harmonics into the morphable model framework and 2) an energy minimization approach to handle extreme lighting conditions based on the theory of Markov random fields.

3.1 Face Morphable Models

The 3D face morphable model was proposed by Blanz et al. [7] to define a vector space of 3D shapes and colors (reflectances). More specifically, both the shape s_{model} and the texture ρ_{model} of a new face can be generated by a linear combination of the shapes and textures of the m exemplar 3D faces, i.e.,

$$s_{model} = \overline{s} + \sum_{i=1}^{m-1} \alpha_i s_i, \quad \rho_{model} = \overline{\rho} + \sum_{i=1}^{m-1} \beta_i t_i, \tag{1}$$

where $\overline{s}$ and $\overline{\rho}$ are the mean shape and texture, s_i and t_i are the eigenvectors of the shape and texture covariance matrix, and α and β are the weighting coefficients to be estimated, respectively.

Based on [31], a realistic face shape can be generated by

$$s^{2D} = fPR\left(\overline{s^{3D}} + \sum_{i=1}^{m-1} \alpha_i s_i^{3D}\right) + t^{2D}, \tag{2}$$

where f is a scale parameter, P is an orthographic projection matrix, and R is a rotation matrix with ϕ, γ, and θ being the three rotation angles for the three axes. The t^{2D} is the 2D translation vector. Given an input face image, the pose parameters f, ϕ, γ, and θ and the shape parameter α can be recovered by minimizing the error between the set of preselected feature points in the 3D morphable model and their correspondences $s(F)^{img}$ detected in the target image:

$$\arg \min_{f,\phi,\gamma,\theta,\alpha,t^{2D}} \left\| s(F)^{img} - \left(fPR\left(\overline{s(F)^{3D}} + \sum_{i=1}^{m-1} \alpha_i s_i(F)^{3D} \right) + t^{2D} \right) \right\|^2, \tag{3}$$

where $\overline{s(F)^{3D}}$ and $s_i(F)^{3D}$ are the shapes of the corresponding feature points in the morphable model in (1).

3.2 Spherical Harmonics Representation

In general, spherical harmonics are the sphere analog of the Fourier bases on the line or circle and they provide an effective way to describe reflectance and illumination. Furthermore, 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 using the first nine spherical harmonic bases [2], [28]:

$$I_{u,v} = \rho_{u,v} E(\vec{n}_{u,v}) \approx \rho_{u,v} \sum_{i=1}^{9} h_i(\vec{n}_{u,v}) \cdot l_i, \tag{4}$$

where I denotes the image intensity, (u, v) is the image pixel coordinate, $\vec{n}$ is the surface normal, ρ is the surface albedo, E is the irradiance, l_i is the illumination coefficient, and h_i is the spherical harmonic basis as follows:

$$\begin{aligned}
&h_1 = \frac{1}{\sqrt{4\pi}}, \quad h_2 = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} * n_z, \quad h_3 = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} * n_y, \\
&h_4 = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} * n_z, \quad h_5 = \frac{\pi}{8}\sqrt{\frac{5}{4\pi}} * \left(3n_z^2 - 1\right), \\
&h_6 = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} * n_y n_z, \quad h_7 = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} * n_x n_z, \\
&h_8 = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} * n_x n_y, \quad h_9 = \frac{3\pi}{8}\sqrt{\frac{5}{12\pi}} * \left(n_x^2 - n_y^2\right),
\end{aligned} \tag{5}$$

where n_x, n_y, n_z denote the x, y, and z components of the surface normal $\vec{n}$. Therefore, any image under general illumination conditions (i.e., without any specific illumination assumption such as a point light source) can be approximately represented by a linear combination of the above spherical harmonic illumination bases, which forms a linear equation system, i.e.,

$$I \approx [\rho_1 H_1, \rho_2 H_2, \ldots, \rho_n H_n]^T \cdot l, \tag{6}$$

where $I = [I(\vec{n}_1), I(\vec{n}_2), \ldots, I(\vec{n}_n)]^T$, $H_i = [h_1(\vec{n}_i), h_2(\vec{n}_i), \ldots, h_9(\vec{n}_i)]^T$, $l = [l_1, l_2, \ldots, l_9]^T$, and n is the number of sample points on the face image.

4 3D Spherical Harmonic Basis Morphable Model

Face morphable models [7] were successfully applied in both face recognition and face synthesis applications [8], [45], where a face was represented by a shape vector and a texture vector. Inspired by the idea of morphing, we propose a 3D SHBMM to estimate and change the illumination condition of face images based on the statistical ensembles of facial properties such as shape and texture. More specifically, we integrate morphable models and the Spherical harmonic illumination representation by modulating the texture component with the spherical harmonic bases. Thus, any face under arbitrary illumination conditions and poses can be represented simply by three low-dimensional vectors: shape parameters, spherical harmonic basis parameters, and illumination coefficients, which are called the SHBMM parameters. This low-dimensional representation greatly facilitates both face recognition and synthesis especially when only one input image under unknown lighting is provided.

4.1 Low-Dimensional Representation

A spherical harmonic basis morphable model is a 3D model of faces with separate shape and spherical harmonic basis models that are learned from a set of exemplar faces. Morphing between faces requires complete sets of correspondences between the faces. Similarly to [7], when building such a model, we transform the shape and

spherical harmonic basis spaces into vector spaces. We used the morphable model database supplied by USF [7] to construct our model. For each 3D face, we computed the surface normal $\vec{n}$ of each vertex in the 3D scan mesh, and then, the first nine harmonic images of the objects, $\vec{b^h}(\vec{n}) = [b_1^h(\vec{n}), b_2^h(\vec{n}), \ldots, b_9^h(\vec{n})]^T$, by multiplying the surface albedo ρ with the first nine harmonic bases $H(\vec{n})$ in (5), i.e.,

$$\vec{b^h}(\vec{n}) = \rho \cdot H(\vec{n}) = [\rho \cdot h_1(\vec{n}), \rho \cdot h_2(\vec{n}), \ldots, \rho \cdot h_9(\vec{n})]^T, \quad (7)$$

where $\vec{n}$ denotes the surface normal. With this set of basis images, (4) can be simplified as

$$I(\vec{n}_{u,v}) \approx \sum_{i=1}^{9} b_i^h(\vec{n}_{u,v}) \cdot l_i, \quad (8)$$

where (u, v) is the image coordinate. Consequently, any image under arbitrary illumination conditions can be approximately represented by the linear combination of the basis images.

We represent a face using a shape vector $s = [X_1, Y_1, Z_1, X_2, ..., Y_n, Z_n]^T \in \Re^{3n}$ and a spherical harmonic basis vector $b = [\vec{b^h}(\vec{n_1})^T, ..., \vec{b^h}(\vec{n_n})^T]^T \in \Re^{9n}$, where n is the number of the pixels in the target face area of the input 2D image. The spherical harmonic basis morphable model can be constructed using a data set of m exemplar faces such that $S = [s_1, \ldots, s_{m-1}]$ and $B = [b_1, \ldots, b_{m-1}]$ are the centered shape and spherical harmonic basis, respectively. Therefore, new shapes s and spherical harmonic bases b can be generated by a linear combination of the shapes and textures of the m exemplar faces as follows:

$$s = \overline{s} + S\alpha; \quad b = \overline{b} + B\gamma, \quad (9)$$

where $\overline{s}$ is the mean shape vector, $\overline{b}$ is the mean SHB vector, and $\alpha = [\alpha_1, \ldots, \alpha_{m-1}]$ and $\gamma = [\gamma_1, \ldots, \gamma_{m-1}]$ are the weighting coefficient vectors. Combining (8) and (9) shows that any face under arbitrary illumination conditions can be represented by three low-dimensional vectors (SHBMM parameters): $\{\alpha, \gamma, \ell\}$, i.e., the geometry parameters, the spherical harmonic basis parameters, and the illumination coefficients, respectively.

Furthermore, as discussed in Section 3.1, the geometry parameters α can be recovered based on the face feature points detected by an automatic face feature detection method [44]. The resulting shape parameters α establish the alignment between the input 2D image and the 3D face models, which allows us to recover the spherical harmonic basis parameters γ and the illumination coefficients ℓ. The estimation algorithm, including the initial shape estimation step, will be described in the following section.

4.2 Estimating Spherical Harmonic Basis Parameters and Illumination Coefficients

Based on (8) and (9), a new face image can be generated by

$$I = (\bar{b} + B\gamma)^T(\mathbb{I} \otimes \ell), \quad (10)$$

where $\bar{b} + B\gamma$ is the spherical harmonic basis component of the SHBMM, ℓ is the vector of illumination coefficients, $\mathbb{I}$ is the $n \times n$ identity matrix, and $\otimes$ is the Kronecker product. Given an input image I_{input} of a face, the spherical harmonic basis parameters γ and the illumination coefficients ℓ can be estimated by minimizing the difference between the input image and the rendered image from SHBMM:

$$\arg\min_{\gamma,\ell} \|(\bar{b} + B\gamma)^T(\mathbb{I} \otimes \ell) - I_{input}\|^2. \quad (11)$$

Equation (11) is solved iteratively as shown in Algorithm 1.

Algorithm 1. The outline of our estimation algorithm based on 3D SHBMM

1. Estimate the initial shape parameter α as described in Section 3.1. The face feature points $s(F)^{img}$ on the input image are detected by an automatic face feature detection method [44]. Then, based on the set of detected feature points and the corresponding preselected feature points in the 3D morphable model, the pose parameters f, ϕ, γ, and θ and the shape parameter α can be recovered using (3), i.e.,
$$\arg\min_{f,\phi,\gamma,\theta,\alpha,t^{2D}} \left\| s(F)^{img} - (fPR(\overline{s(F)^{3D}} + \sum_{i=1}^{m-1} \alpha_i s_i(F)^{3D}) + t^{2D}) \right\|^2,$$
where $\overline{s(F)^{3D}}$ and $s_i(F)^{3D}$ are the shape of the corresponding feature points in the morphable model in (1).
2. Initialize the spherical harmonic basis parameter $\gamma^{(0)}$ as 0 and set the step index $i = 1$.
3. For each step i, estimate the illumination coefficients ℓ_i by solving a linear system based on the input image I_{input} and the spherical harmonic basis parameter $\gamma^{(i-1)}$ from the previous step, i.e.,
$$b^{(i-1)^T}\left(\mathbb{I} \otimes \ell^{(i)}\right) = \left(\bar{b} + B\gamma^{(i-1)}\right)^T\left(\mathbb{I} \otimes \ell^{(i)}\right) = I_{input}.$$
4. Compute the residual error $\delta I^{(i)}$ between the input image I_{input} and the rendered image $I^{(i)} = b^{(i-1)}\ell^{(i)}$ based on the illumination coefficients $\ell^{(i)}$ estimated in step 3,
$$\delta I^{(i)} = I_{input} - b^{(i-1)^T}\left(\mathbb{I} \otimes \ell^{(i)}\right).$$
5. Compute the update of the spherical harmonic basis parameters $\delta\gamma$ by solving the following linear equation system,
$$(B\delta\gamma)^T\left(\mathbb{I} \otimes \ell^{(i)}\right) = \delta I^{(i)}.$$
6. Update the spherical harmonic basis parameters $\gamma^{(i)} = \gamma^{(i-1)} + \delta\gamma$ and increase the step index i by 1.
7. Perform steps 3 to 6 iteratively until $\|\delta I\| < \xi_I$ or $\|\delta\gamma\| < \xi_\gamma$, where ξ_I and ξ_γ are preselected constants.

The key part of this process is the minimization of the image error as shown in (11), where two variables γ, ℓ need to be recovered iteratively. Moreover, there are two methods to start the iteration: Initialize γ as 0 and compute ℓ afterward as we described above, or start with a random ℓ to achieve the global minimum. We chose the first method since our experiments on synthetic data showed that the illumination coefficients ℓ computed by using the mean spherical harmonic basis $\bar{b}$ were close to the actual values, which made the recovery process fast and accurate. Another important

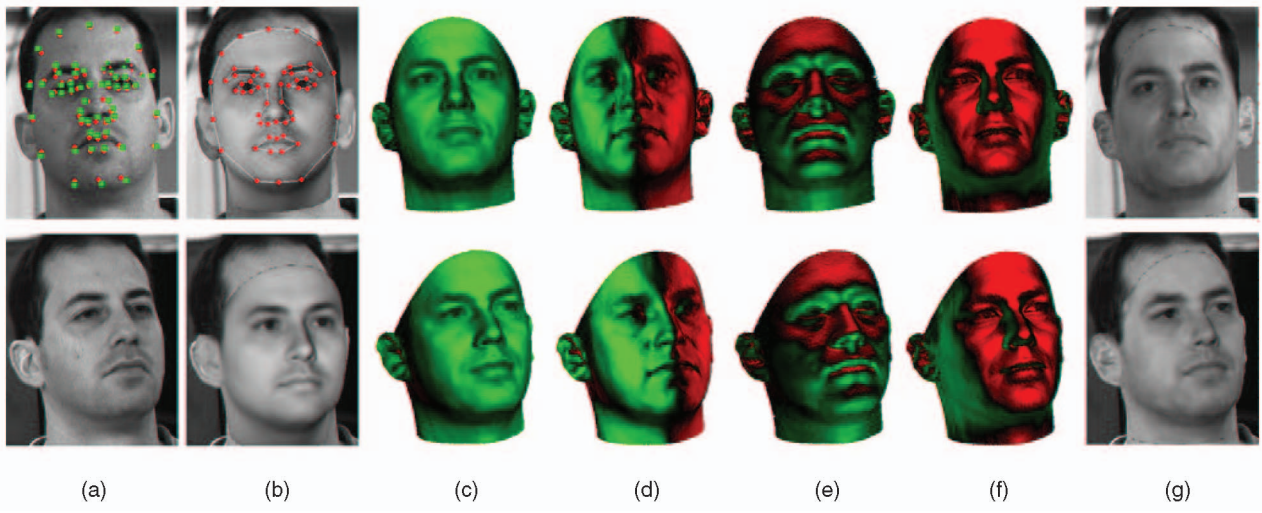


Fig. 1. Fitting 3D SHBMM to images. (a) The input image. (b) The initial result of fitting the 3D face model to the input image. Red points are selected major feature points and green points are the corresponding points on the face mesh model. (c) The recovered first order spherical harmonic basis. (d)-(f) The recovered second order spherical harmonic bases, where the red color means positive values and the green color means negative values. (g) The rendered image using the recovered parameters.

point is that the spherical harmonic basis cannot capture specularities and cast shadows. Thus, for better recovery results, we employed two thresholds to avoid using the image pixels in the regions of strong specular reflection or cast shadow. Fig. 1 shows the fitting process and results, where the first image is the input image followed by initial fitting and recovered spherical harmonic basis, and the last image is the rendered image using the recovered parameters. Red points are selected major feature points and green points are the corresponding points on the face mesh model.

5 SHBMM for Synthesis and Recognition

In this section, we will demonstrate how to apply our spherical harmonic basis morphable model to face synthesis and face recognition. Section 5.1 will explain how to combine our SHBMM with the ratio image technique for photorealistic face synthesis. In Section 5.2, we will propose two face recognition methods based on the recovered SHBMM parameters and delit images, respectively.

5.1 Image Synthesis Using SHBMM

The face synthesis problem we will discuss can be stated as following: given a single image under unknown lighting, can we remove the effects of illumination from the image ("delighting") and generate images of the object consistent with the illumination conditions of the target images ("relighting")? The input image and target images can be acquired under different unknown lighting conditions and poses. Based on the set of SHBMM parameters $\{\alpha_s, \gamma_s, \ell_s\}$ from an input face I_s, we combine our spherical harmonic basis morphable model and a concept similar to ratio images [30], [43] to generate photorealistic face images. In particular, we can render a face $I_s^{'}$ using the recovered parameters to approximate I_s: $I_s^{'} = (\bar{b} + B\gamma_s)^T(\mathbb{I} \otimes \ell_s)$. Thus, the face texture (delit face) can be directly computed from the estimated spherical harmonic basis, and face relighting can be performed by setting different values to the illumination parameters ℓ similar to [2]. Furthermore, ignoring cast shadows and specularities, we notice that:

$$\frac{I_{s_i}}{I_{d_i}} = \frac{H(\vec{n}_{g_i})\rho_{g_i}\ell}{\rho_{g_i}} \approx \frac{H(\vec{n}_{e_i})\rho_{e_i}\ell}{\rho_{e_i}} = \frac{I_{s_i}^{'}}{\rho_{e_i}}, \tag{12}$$

where I_s is the input image, I_d is the delit image, i is the index of the sample points, $H(n)\ell$ is the spherical harmonic basis, n_g and n_e are the actual and estimated surface normals, and ρ_g and ρ_e are the actual and estimated face albedo, respectively. Equation (12) states that the intensity ratio of the input image to the delit image should be approximately equal to that of the rendered face and the corresponding face texture (albedo). The face texture (albedo) of the rendered face can be easily computed based on (5) and (7), i.e., $\rho = \sqrt{4\pi}b_1$. Therefore, the delit image can be computed by rewriting (12):

$$I_{d_i} = \frac{I_{s_i} \cdot \sqrt{4\pi}b_{1_{(i)}}}{(\bar{b}_{(i)} + B_{(i)}\gamma_s)^T\ell_s}, \tag{13}$$

where $b_{1_{(i)}}$, $\bar{b}_{(i)}$, and $B_{(i)}$ are the ith nine elements of the vector b_1, the ith nine elements of the vector $\bar{b}$, and the ith nine rows of the matrix B, respectively.

Based on (12) and (13), given an input image I_s and the recovered SHBMM parameters $\{\alpha_s, \gamma_s, \ell_s\}$, we can obtain the relation between the original image I_s and the delit image I_d as the following equation:

$$\frac{I_{s_i}}{I_{d_i}} = \frac{(\bar{b}_{(i)} + B_{(i)}\gamma_s)^T\ell_s}{\sqrt{4\pi}b^s_{1_{(i)}}}, \tag{14}$$

where b_1^s is the estimated SHB vector b_1 from the original image I_s. Furthermore, if another input image I_t and its recovered illumination coefficients ℓ_t are provided, we can also obtain the relation between the relit image I_r and the delit image I_d similar to (14):

$$\frac{I_{r_i}}{I_{d_i}} = \frac{(\bar{b}_{(i)} + B_{(i)}\gamma_s)^T\ell_t}{\sqrt{4\pi}b^s_{1_{(i)}}}. \tag{15}$$

Combining (14) and (15), the relit image can be recovered directly without computing the delit image explicitly:

Fig. 2. Comparison of relit images from the same input image "driven" by target images of different subjects under similar illumination conditions. The illumination information is preserved to a large extent, across different subjects.

$$I_{r_i} = \frac{(\bar{b}_{(i)} + B_{(i)}\gamma_s)^T \ell_t}{(\bar{b}_{(i)} + B_{(i)}\gamma_s)^T \ell_s} I_{s_i}. \quad (16)$$

In order to evaluate the performance of our method, we used the CMU-PIE data set [36] which includes images taken under varying pose and illumination conditions. The CMU-PIE Database contains 68 individuals, which are not included in the USF data set used to compute the spherical harmonic basis morphable model. Fig. 2 shows relit images of the same input image "driven" by target images of three different subjects. The results suggest that our SHBMM-based method extracts and preserves illumination information consistently across different subjects and poses.

More examples are included in Fig. 3 which shows a series of face relighting experiments. The top row shows the images with target illumination conditions. The input images from two difference subjects are listed in the leftmost column. The input image and target images can be acquired under different unknown lighting conditions and poses. The results show that high-quality images can be synthesized even if only a single-input image under arbitrary unknown lighting is available.

5.2 Face Recognition

Recognition based on SHBMM parameters. We divide an image data set into a gallery set and a test set. The gallery set includes the prior images of people to be recognized. For each image I_i in the gallery set and a testing image I_t, we recover SHBMM parameters $\{\alpha_i, \gamma_i, \ell_i\}$ and $\{\alpha_t, \gamma_t, \ell_t\}$. Since the identity of a face is represented by $\{\alpha, \gamma\}$, recognition is done by selecting the face of a subject i whose recovered parameters $\{\alpha_i, \gamma_i\}$ are the closest to $\{\alpha_t, \gamma_t\}$. In this method, the gallery image and the testing image can be acquired under different arbitrary illumination conditions. In our implementation, the shape recovery was based on an automatic face feature detection method [44]. For images of one face under the same pose, the shape parameters recovered were almost the same; thus, the shape parameters α might encode strong subject identity information which can support recognition very well. Since the focus of this paper is on illumination and texture estimation, to examine our recognition method unbiasedly,

Fig. 3. Face relighting results. First column: The input images from different subjects. Top row: The images with desired illumination conditions. Images with good quality are synthesized even if only one input image is available.

Fig. 4. Example images from the CMU-PIE Database. (a)-(f) The lighting conditions are 2, 6, 8, 12, 16, and 20, respectively. The details about flash light positions can be found in [36].

we performed experiments by just using the spherical harmonic basis parameters γ. In a complete application, shape would be recovered using dense shape reconstruction methods [8], [13], so both shape and texture parameters would be used for recognition.

Recognition based on delit images. For each image I_i in the gallery set and a testing image I_t, we compute delit images I_d^i and I_d^t. The recognition is done by selecting the face of a subject whose delit image is closest to the delit image of the test subject. In this method, delit images should be aligned before they are compared against each other.

Experimental results. In order to evaluate the performance of the above two methods, i.e., recognition based on SHBMM parameters and delit images, we examined the recognition rates on the CMU-PIE Database [36]. There are 68 subjects included in the CMU-PIE Database with 13 poses and 21 directional illumination directions. Since the capability of handling illumination variations is of the central focus of this paper, face recognition results are reported on the frontal pose images only (recognition results on a subset of the PIE database under varying pose are reported in [47]). For each subject, out of the 21 directional illumination conditions, the image with certain illumination direction was included in the training set (which is referred to as the *training illumination condition*) and the remaining images were used for testing. The details about flash light positions can be found in [36]. In our experiment, we selected the following six representative illumination conditions: 1) frontal lighting: flash 08; 2) near-frontal lighting (between 22.5-45 degrees): flash 06, 12, and 20; and 3) side lighting (with the largest illumination angles in the PIE database): flash 02 and 16. The image examples under the selected six illumination conditions are shown in Fig. 4. Fig. 5 reports the recognition results of six representative illumination conditions with each selected as the training illumination condition. The results in Fig. 5 show that the SHBMM-based methods, using both SHBMM parameters and delit images, can achieve high recognition rates for images under regular illumination conditions. However, their performance decreases significantly in extreme illumination cases, such as light positions 2 and 16.

Light Positions	SHBMM-Based Method	
	Using SHB Parameters	Using De-lit Images
2	68.23%	70.74%
6	91.14%	97.65%
8	93.31%	98.93%
12	92.66%	99.12%
16	67.92%	69.52%
20	89.62%	98.04%

Fig. 5. Face recognition under different illumination conditions: We evaluate and compare the recognition performance based on the SHBMM parameters and delit images. The CMU-PIE Database [36] is used in this experiment, which includes 68 subjects and 21 directional illumination conditions. For each subject, the images with the illumination directions, listed in the left column, are included in the training set and the remaining images are used for testing. The details about flash light positions listed in the left column can be found in [36] and the image examples are shown in Fig. 4. The results show that the SHBMM-based method can achieve high recognition rates for images under a wide range of illumination conditions. However, its performance decreases significantly in extreme illumination conditions, such as light positions 2 and 16.

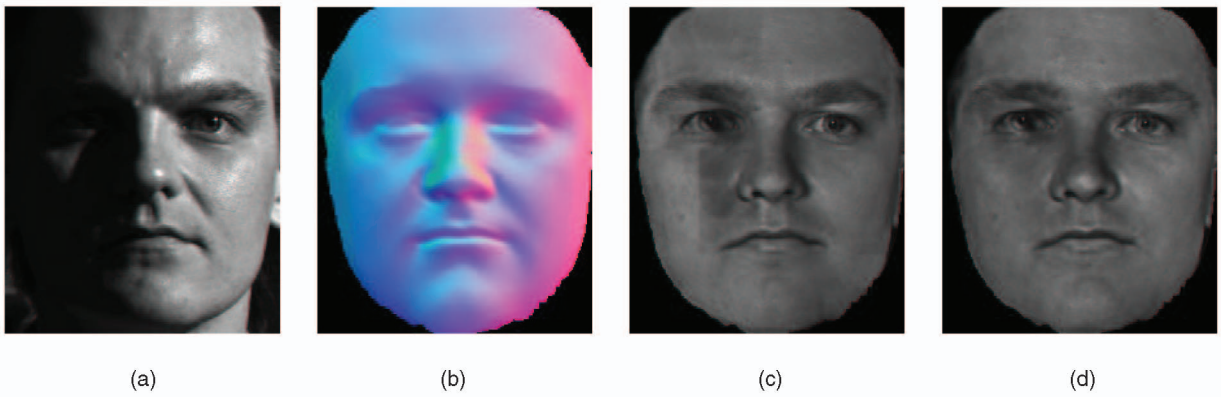


Fig. 6. Example result. (a) The original image taken under an extreme lighting condition. (b) The recovered surface normals from our method (where R,G,B color values represent the x, y, z components of the normal), and the recovered albedo from our method is shown in (c) without the spatial coherence term and (d) with the spatial coherence term. As is evident, the region inconsistency artifacts in (c) are significantly reduced in (d).

6 Energy Minimization Framework

As discussed in Section 5, although the proposed SHBMM-based method can achieve good performance on face relighting and recognition, the performance decreases significantly under extreme lighting conditions. This is because the representation power of the 3D SHBMM model is inherently limited by the coupling of texture and illumination bases. For an image taken under extreme lighting conditions, the lighting approximation errors vary significantly across image regions. Such spatially varying lighting approximation errors are difficult to handle with a single set of SHBMM coefficients over the entire image region. To address this problem, we propose a spatially varying texture morphable model by decoupling the texture from shape and illumination and dividing the image into multiple regions. Facilitated by the theory of MRFs, we propose a novel energy minimization framework to jointly recover the lighting, the geometry (including the surface normal), and the albedo of the target face. We show that our technique is able to handle challenging areas, such as cast shadows and saturated regions, and is robust to extreme lighting conditions and partial occlusions as well.

6.1 Subregion-Based Scheme

Since illumination effects in smaller image regions are more homogeneous, we subdivide a face into smaller regions to better fit the image under an extreme lighting condition. The idea of subdivision was also used by Blanz and Vetter [7], where a face is subdivided along feature boundaries (such as eyes, nose, mouth, etc.) to increase the expressiveness of the morphable models. They estimated morphable model parameters independently over each region and performed smoothing along region boundaries to avoid visual discontinuity. However, this approach cannot be applied to images under extreme lighting conditions because of the inconsistency of the estimated textures in different regions (e.g., Fig. 6c). Furthermore, if most pixels in a region are in cast shadows or saturated areas, there might not be enough information to recover the texture within the region itself. To address these problems, we introduce the spatial coherence constraints to the texture model between neighboring regions.

Instead of subdividing a face along feature boundaries as in [7], for simplicity, we divide a face into regular regions with a typical size of 50×50 pixels.[1] For each region, we represent its face texture by using a PCA texture model similar to (1):

$$\rho^q = \overline{\rho}^q + \sum_{k=1}^{m-1} \beta_k^q t_k^q, \quad q = 1, \ldots, Q, \tag{17}$$

where Q is the total number of regions, $\overline{\rho}^q$ is the mean albedo of the qth region, and t_k^q are the albedo eigenvectors of the qth region, which are computed from the exemplar faces in the morphable model database by dividing them into the same regions as the target face. Then, we pose the coherence constraints on the PCA coefficients β_k^q between neighboring regions: Given two neighboring regions q_i and q_j, for each PCA coefficient $k = 1, \ldots, m-1$, we model $\beta_k^{q_i} - \beta_k^{q_j}$ as a random variable of Gaussian distribution with mean 0 and variance $(\sigma_k^{q_i q_j})^2$. We also obtain the spatial coherence between the two neighboring regions by maximizing $\Pi_{k=1}^{m-1} Pr(\beta_k^{q_i} - \beta_k^{q_j})$, which is equivalent to minimizing

$$\sum_{k=1}^{m-1} \left(\frac{\beta_k^{q_i} - \beta_k^{q_j}}{\sigma_k^{q_i q_j}} \right)^2. \tag{18}$$

It is worth pointing out that the spatial coherence constraints are posed over texture PCA coefficients, not on pixel values directly. The main advantage is that even if the PCA coefficients are the same between two regions, the pixel values can be completely different.

We could potentially use a similar idea for the shape model representation. But, since we are not trying to recover detailed geometry, a single-shape model is sufficient. This agrees with [29] and the perception literature (such as Land's retinex theory [19]), where on Lambertian surfaces, high-frequency variation is due to texture and low-frequency variation is probably associated with illumination, which is determined by the surface geometry and the environment lighting. Given that we are mainly interested in surface normals, we directly model the surface normal as

1. The subdivision is done in the image space and projected back to the 3D face model since the relationship between the 2D input image and the 3D face model is recovered by (3).

$$\vec{n}^M_{u,v} = \left(\overline{\vec{n}_{u,v}} + \sum_{j=1}^{m-1} \lambda_j \vec{n}^j_{u,v}\right) \Bigg/ \left\| \overline{\vec{n}_{u,v}} + \sum_{j=1}^{m-1} \lambda_j \vec{n}^j_{u,v} \right\|, \quad (19)$$

where λ is the weighting coefficient to be estimated.

6.2 MRF-Based Framework

Following the discussion in Section 3.2, the illumination model in (4) can be added as another constraint to fit the image I. Note that, for pixels which are saturated or in cast shadows, (4), in general, does not hold. Therefore, for each pixel (u, v), we assign a weight $W_{u,v}$ to indicate the contribution of the illumination model in (4). $W_{u,v}$ is set to a small value if the pixel is in the cast shadow or the saturated area.

Finally, all the constraints can be integrated into an energy minimization problem as follows:

$$\begin{aligned} \arg\min_{\rho,\lambda,\beta,l} & \sum_{q=1}^{Q} \sum_{(u,v)\in\Omega_q} \left\{ W_{u,v} \left(I_{u,v} - \rho_{u,v} \sum_{i=1}^{9} h_i(\vec{n}^M_{u,v}) l_i \right)^2 \right. \\ & \left. + W_{MM} \left(\rho_{u,v} - \rho^q_{u,v}\right)^2 \right\} \\ & + W_{SM} N_{sr} \sum_{(i,j)\in\mathcal{N}} \sum_{k=1}^{m-1} \left(\frac{\beta^i_k - \beta^j_k}{\sigma^{ij}_k} \right)^2, \end{aligned} \quad (20)$$

where ρ is the output albedo, (u, v) is the pixel index, Ω_q denotes the qth region, $\mathcal{N} = \{(i, j) | \Omega_i \; and \; \Omega_j \; are \; neighbors\}$ is the set of all pairs of neighboring regions, $\vec{n}^M$ is constrained by the shape subspace defined in (19), ρ^q is constrained by the texture subspace defined in (17), and W_{MM} and W_{SM} are the weighting coefficients of the texture morphable model term and the coherence constraint term, respectively. N_{sr} is the average number of pixels in a region and $(\sigma^{ij}_k)^2$ is estimated from the exemplar texture data in the morphable models [7].

The objective function in (20) is an energy function of a Markov random field. The first two terms in (20) are the first order potentials corresponding to the likelihood of the observation data given the model parameters, and the third term is the second order potential which models the spatial dependence between neighboring regions. Therefore, we have formulated the problem of jointly recovering the shape, texture, and lighting of an input face image as an MRF-based energy minimization (or maximum a posteriori) problem. Furthermore, this framework can be extended to handle different poses by replacing the normal constraint in (19) to the shape constraint in (3).

In our implementation, we determine whether a pixel is in a cast shadow or saturated region by simple thresholding. Typically, in our experiments on a 0-255 gray-scale face image, the threshold values are 15 for the cast shadows and 240 for the saturated pixels.[2] $W_{u,v}$ is set to 0 for the pixels in the shadow and saturated areas and 1 for the pixels in other regular areas, and $W_{MM} = 4$ and $W_{SM} = 500$ for all regions. Because the typical size of a regular region is 50×50 pixels, the average pixel number N_{sr} is 2,500. Due to the nonlinearity of the objective function (20), the overall optimization problem is solved in an iterative fashion. First, by fixing the albedo ρ and the surface normal $\vec{n}$, we solve for the global lighting l. Then, by fixing the lighting l, we solve for the albedo ρ and the surface normal $\vec{n}$. Because gradients of (19) and (20) can be derived analytically (for the details, refer to the Appendix), the standard conjugate gradient method is used for the optimization.

To solve the objective function (20), initial albedo values ρ are required for the nonlinear optimization. Since the linear equation system (6) is underconstrained as the surface albedo ρ varies from point to point, it is impossible to obtain the initial lighting l_{init} directly without any prior knowledge. One solution is to approximate the face albedo ρ by a constant value ρ_{00} and estimate the initial lighting l_{init} by solving an overconstrained linear system [42]. However, since the initial lighting l_{init} can be estimated by the previous SHBMM-based method as described in Section 4.2, we are able to obtain the initial albedo values based on the spherical harmonics representation in (4). More specifically, we can compute ρ_{init} as follows:

$$\rho_{init_{u,v}} = \frac{I_{u,v}}{\sum_{i=1}^{9} h_i(\vec{n}_{u,v}) \cdot l_{init_i}}, \quad (21)$$

where I denotes the image intensity, (u, v) is the image pixel coordinate, $\vec{n}$ is the surface normal, l_{init_i} is the initial lighting coefficient, and h_i is the spherical harmonic basis. In particular, given the initial shape and the associated surface normal $\vec{n}$ recovered by the SHBMM-based method as described in Algorithm 1 in Section 4.2, the spherical harmonic basis h_i can be computed by (5).

For clarity purposes, the outline of the optimization algorithm is presented in Algorithm 2. An example result is shown in Fig. 6, where Fig. 6a is the original image taken under an extreme lighting condition, Fig. 6b shows the recovered surface normal from our method, and the recovered albedo from our method is shown in Fig. 6c without the spatial coherence term and Fig. 6d with the spatial coherence term. As we can see, the region inconsistency artifacts in Fig. 6c are significantly reduced in Fig. 6d.

Algorithm 2. The outline of our MRF-based estimation algorithm

1. **Initial Estimation, Illumination and Albedo Estimation**: Obtain the initial values of the shape parameter α and the lighting coefficient l_{init} by the SHBMM-based method as described in Algorithm 1 in Section 4.2. Compute the initial albedo value ρ_{init} by (21), i.e.,
$$\rho_{init_{u,v}} = \frac{I_{u,v}}{\sum_{i=1}^{9} h_i(\vec{n}_{u,v}) \cdot l_{init_i}}.$$
2. **Image Segmentation**: Segment the input face image into the following parts: regular shaded regions, saturated regions, and shadow regions, by thresholding the image intensity values, and further divide the image into regular subregions. Typically, in our experiments on a $0 - 255$ gray scale face image, the threshold values are 15 for the cast shadow and 240 for the saturated pixels, and the size of a subregion is 50×50 pixels.
3. **Iterative Minimization**: Solve the objective function (20), i.e.,

2. In the presence of multiple illuminants, since the cast shadows become less dominant, a higher threshold value, typically 55 in our experiments, might be needed to detect strong cast shadows. A more sophisticated method, such as in [47], could also be employed to detect cast shadows automatically.

$$\arg\min_{\rho,\lambda,\beta,l} \sum_{q=1}^{Q} \sum_{(u,v)\in\Omega_q} \left\{ W_{u,v}\left(I_{u,v} - \rho_{u,v} \sum_{i=1}^{9} h_i(\vec{n}^M_{u,v}) l_i \right)^2 + W_{MM}\left(\rho_{u,v} - \rho^q_{u,v}\right)^2 \right\} + W_{SM} N_{sr} \sum_{(i,j)\in\mathcal{N}} \sum_{k=1}^{m-1} \left(\frac{\beta^i_k - \beta^j_k}{\sigma^{ij}_k} \right)^2$$

in an iterative fashion. As shown in the Appendix, the gradients of (20) can be derived analytically. Therefore the standard conjugate gradient method is used for the optimization. Typically, only two iterations were needed in our experiments to generate photorealistic results.

- Fixing the lighting l, solve for the albedo ρ, the texture PCA coefficients β, and the shape PCA coefficients λ for the surface normal $\vec{n}$.
- Fixing the albedo ρ and the surface normal $\vec{n}$, solve for the global lighting l by (4), i.e.,

$$I_{u,v} = \rho_{u,v} \sum_{i=1}^{9} h_i(\vec{n}_{u,v}) \cdot l_i.$$

7 Image Synthesis and Recognition

Using the approach proposed in Section 6, we can recover the albedo ρ, the surface normal $\vec{n}$, and the illumination parameter l from an input face image I. In this section, we will show how to perform face relighting for image synthesis and delighting for face recognition based on the recovered parameters. Compared to the methods proposed in [43], [46], our proposed framework can also handle images with cast shadows, saturated areas, and partial occlusions and is robust to extreme lighting conditions.

7.1 Image Relighting and Delighting

Based on the recovered albedo ρ, the surface normal $\vec{n}$, and the illumination parameter l, we can render a face I' using the recovered parameters by setting different values to the illumination parameter l' [2], [43]:

$$I'_{u,v} = \rho_{u,v} \sum_{i=1}^{9} h^i(\vec{n}_{u,v}) \cdot l'_i, \tag{22}$$

where (u, v) is the image pixel coordinate. However, because certain texture details might be lost in the estimated face albedo ρ, we also use the ratio image technique to preserve photorealistic quality. The ratio image technique used in [43], which is based on the spherical harmonic illumination representation, has generated promising results under regular lighting conditions. However, it cannot be adopted in our framework because of the large approximation error for extreme lighting conditions. Instead, we smooth the original image using a Gaussian filter, and then, compute the pixelwise ratio between the original image and its smoothed version. This pixelwise ratio is then applied to the relit image computed by (22) to capture the details of the original face texture. Typically, for a 640×480 image, the size of the Gaussian kernel is 11×11 with $\sigma = 2$. Note that we treat the dark regions in the same way as regular bright regions. Since it is possible that there are fewer texture details in dark regions than in other regions, the relit dark regions might not have the same quality as the relit bright regions.

In order to evaluate the performance of our framework, we conducted the experiments on two publicly available face data sets: the Yale Face Database B [17] and the CMU-PIE Database [36]. The face images in both databases contain challenging examples for relighting. For example, there are many images with strong cast shadows, saturated or extremely low-intensity pixel values. More specifically, in Yale Face Database B, the images are divided into five subsets according to the angles of the light source direction from the camera optical axis, i.e.,

1. less than 12 degree;
2. between 12 and 25 degree;
3. between 25 and 50 degree;
4. between 50 and 77 degree;
5. larger than 77 degree.

Fig. 7a shows one sample image per group of Yale Face Database B. The corresponding relit results from our method are shown in Fig. 7d. Compared to the results from Wen et al.'s method [43] and the previous SHBMM-based method, which are shown in Figs. 7b and 7c, respectively, the results generated by our MRF-based method, as shown in Fig. 7d, have much higher quality especially under extreme lighting conditions such as the images in groups (4-5). Fig. 8 shows more face relighting results on both Yale Face Database B [17] and CMU-PIE Database [36]. Despite the different extreme lighting conditions in the input images (Fig. 8a), our method can still generate high-quality relit results, as shown in Fig. 8b. Readers are also encouraged to download the accompanying video of this paper from http://www.cs.cmu.edu/~wangy/paper/pami09.mov, which includes more relit results demonstrating the performance of our method.

7.2 Face Recognition

In this section, we show that our framework on face relighting from a single image can be used for face recognition. In order to normalize the illumination effects for face recognition, we relight all face images into a canonical lighting condition, i.e., the frontal lighting condition, using (22). Once the illumination effects in images are normalized, any existing face recognition algorithms, such as Eigenfaces (PCA) [39] and Fisherfaces (LDA) [3], can be used on the relit face images for face recognition. In order to evaluate the face recognition performance of our proposed method, we tested our MRF-based method using the Yale Face Database B [17], which includes images taken under different lighting conditions, and compared our recognition results with other existing methods in the literature. The experimental results are reported in Fig. 9.

In Yale Face Database B, there are 5,760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). In our current experiment, we consider only illumination variations so that we choose to perform face recognition for the 640 frontal pose images. We choose the simplest image correlation as the similarity measure between two images, and nearest neighbor as the classifier. For the 10 subjects in the database, we take only one frontal image per person as the gallery image. The remaining 630 images are used as testing images.

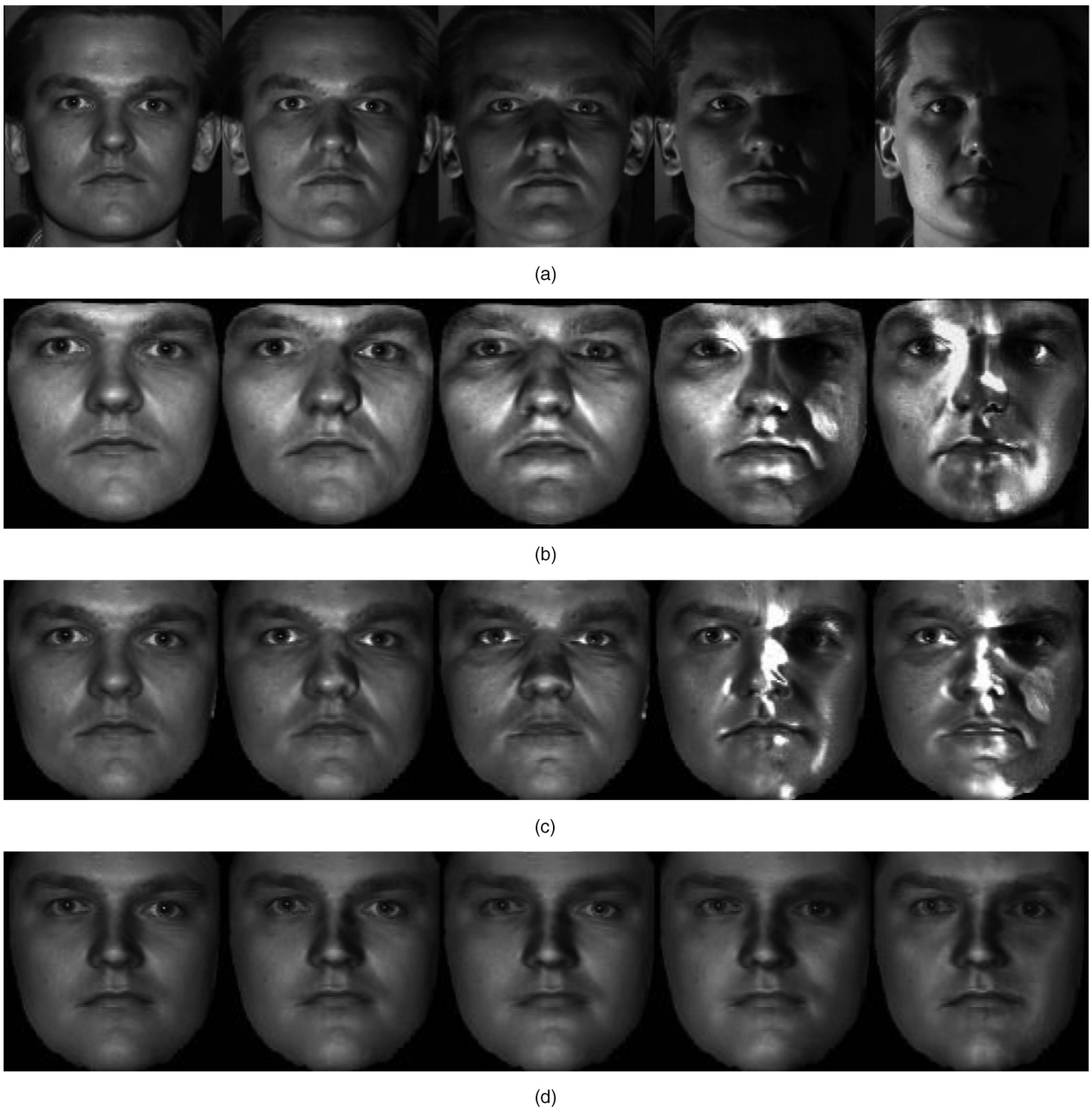


Fig. 7. Face relighting experiment on Yale Face Database B [17]. (a) Example input images from group 1 to group 5. (b) The corresponding results under frontal lighting using the method proposed by Wen et al. [43]. (c) The relit results from our SHBMM-based method. (d) The relit results from our MRF-based method. Compared to the methods by Wen et al. [43] and the SHBMM-based method, our MRF-based method preserves photorealistic quality, especially under extreme lighting conditions such as the images in rightmost two columns, i.e., in groups (4-5).

As shown in Fig. 9, our method has a very low recognition error rate, compared to all the existing recognition methods in the literature, and maintains almost the same performance even when the lighting angles become large. When the lighting direction of the test image is further away from the lighting direction of the training image, the respective illumination effects exhibit larger differences, which will cause a larger recognition error rate. As is evident, our relighting technique significantly reduces error rates, even in extreme lighting conditions (e.g., lighting angles > 50 degree).

8 Comparison between the SHBMM and MRF-Based Methods

In Sections 4 and 6, we have proposed two methods to estimate and modify the illumination conditions of a single image, namely, the 3D-SHBMM and the MRF-based method. To better understand the difference and relationship between two methods, we compare them in terms of computation complexity, face synthesis, and face recognition performance.

8.1 Computational Complexity

As explained in Section 4, the proposed SHBMM-based method includes only three low-dimensional vectors: shape

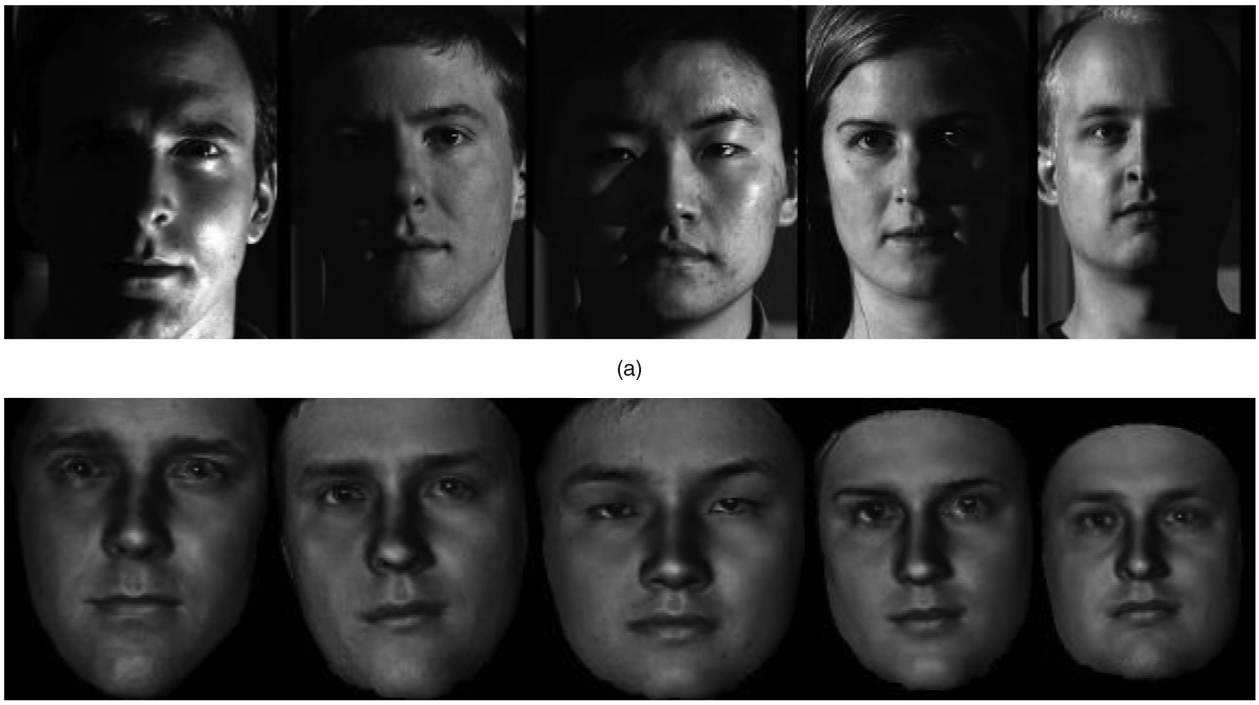


Fig. 8. Face relighting experiment on subjects in both the Yale Database B [17] and the CMU-PIE Database [36]. (a) Example input images taken under different extreme lighting conditions. (b) The synthesized frontal lighting results generated by our MRF-based method with high quality.

parameters, spherical harmonic basis parameters, and illumination coefficients, which are called the SHBMM parameters. However, the MRF-based method proposed in Section 6 involves a more complicated optimization process to solve for a large number of shape, albedo, and illumination parameters. In particular, compared to (20) in the MRF-based method, the objective function of (11) in the SHBMM-based method is much simpler. The reduction of computational complexity mainly comes from two factors:[3]

1. Differently from the MRF-based method, the SHBMM-based method does not require subdividing the input face image into small regions.
2. The objective function itself involves a much smaller number of variables to be optimized.

More specifically, given an input image, we assume that the face area has N pixels, which is divided into Q subregions, and the size of 3D face database is M. Then, the number of variables to be optimized in (20) is

$$\begin{aligned} &N + Q \times (M-1) + (M-1) + 9 \\ &\quad = N + (Q+1) \times (M-1) + 9, \end{aligned}$$

while the number in (11) is only

$$(M-1) + 9 = M + 8,$$

where, typically, N is much larger than M and Q is larger than 10. Therefore, the optimization of (11) is much easier and less expensive than (20).

3. Both the SHBMM and MRF-based methods share the same M-1 shape parameter α as described in the initial shape estimation step in Algorithm 1.

8.2 Face Image Synthesis

In the previous sections, such as Sections 4.2 and 5.1, we showed that the simplified approach based on 3D SHBMM can achieve good performance on delighting and relighting images. However, the representation power of the 3D SHBMM model is limited by the coupling of texture and illumination bases. Therefore, it might fail in extreme lighting conditions, e.g., in the presence of saturated areas.

Methods	Error Rate (%) in Subsets		
	(1, 2)	(3)	(4)
Correlation	0.0	23.3	73.6
Eigenfaces	0.0	25.8	75.7
Linear Subspace	0.0	0.0	15.0
Illum. Cones - Attached	0.0	0.0	8.6
9 Points of Light (9PL)	0.0	0.0	2.8
Illum. Cones - Cast	0.0	0.0	0.0
Zhang & Samaras[46]	0.0	0.3	3.1
BIM (30 Bases)[20]	0.0	0.0	0.7
Wen *et al.* [43]	0.0	1.7	30.7
Our MRF-Based Method	0.0	0.0	0.1

Fig. 9. Recognition results on the Yale Face Database using various previous methods in the literature and our proposed method. Except for Wen et al.'s method [43] and our method, the data were summarized from [20].

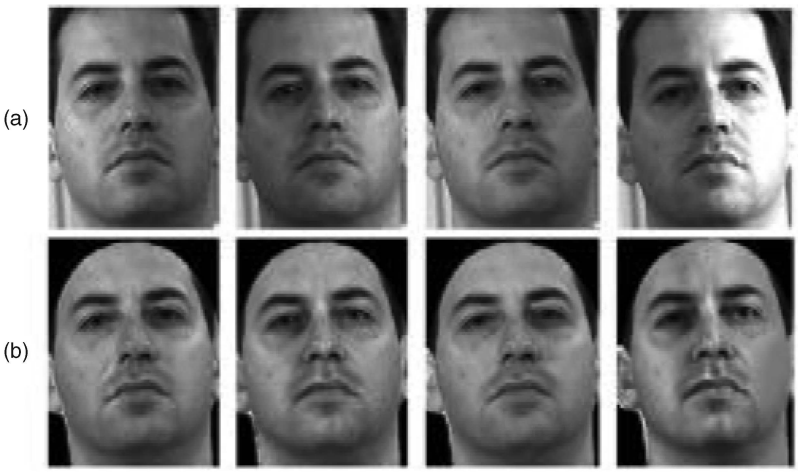


Fig. 10. Example delit faces using our SHBMM-based method. (a) The input images under different illumination conditions. (b) The corresponding delit images. The rightmost column shows a failure example where the input image is saturated.

Fig. 10 shows an example of the face delighting experiment using the SHBMM-based method on the CMU-PIE Database, where the four images in the first row are the input images under different unknown illuminations and the images in the second row are the corresponding delit images. For regular conditions, such as the ones in the left three columns, we found that the delit images we computed exhibit much better invariance to illumination effects than the original images. Quantitatively, for one subject, we computed the delit images of 40 images under different illumination conditions. The variance of the 40 delit images was 6.73 intensity levels per pixel, while the variance of the original images was 26.32. However, for an extreme lighting condition, such as the one in the rightmost column of Fig. 10, where the input image is taken under an extreme illumination condition and part of the face is saturated, the delit result could not recover the saturated area faithfully.

Our MRF-based method, however, decouples the estimation of the illumination and albedo and can handle this situation successfully. An example of high-quality synthesized results in the saturated area is demonstrated in Fig. 11c. For comparison purposes, we also include the delit result using the SHBMM-based method in Fig. 11b. The close-up views in Figs. 11d, 11e, and 11f demonstrate the high quality of the images synthesized by our method even in the presence of saturated areas.

Furthermore, because our MRF-based framework models spatial dependence, it can handle image occlusions as well. This is, in spirit, similar to superresolution and texture synthesis [16], [50], but we are able to recover missing information and remove lighting effects simultaneously. Fig. 12 shows two examples of the face delighting experiment on images under occlusions. Figs. 12a and 12c are the original images under different occlusions and Figs. 12b and 12d are the recovered albedo from our method. The results demonstrate that our method can generate high-quality delit images for the occluded areas as well.

8.3 Face Recognition

In order to compare our SHBMM-based method to the MRF-method in Sections 4 and 6, we examined the recognition performance on all 68 subjects in the CMU-PIE Database [36] using the same setup as in Fig. 5, i.e., using images taken under six representative illumination conditions. The results are reported in Fig. 13. The details about flash light positions can be found in [36]. Fig. 4 shows some image examples. The results in Fig. 13 show that both the SHBMM and MRF-based methods can achieve high recognition rates for images under regular illumination conditions. However, the performance of the SHBMM-based method decreases in the extreme illumination cases, such as light positions 2 and 16, while the MRF-based method is more robust to illumination variation and can maintain a good recognition performance under extreme lighting conditions. It is also important to point out that image-based approaches, such as self-quotient images [41] and correlation filters [35], [40], can achieve comparable

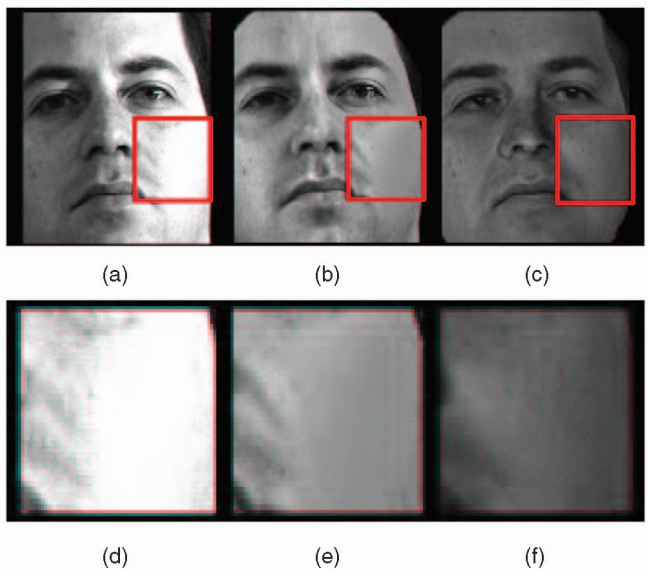


Fig. 11. Face delighting experiment on an image with saturated regions, which is highlighted in the red boxes. (a) The original image where the left side of the face is saturated. (b) The delit result from the SHBMM-based method. (c) The delit result from our MRF-based method. (d)-(f) The close-up views show that a high-quality image is synthesized by our method even in the presence of saturated areas. Note that, because there is always a scale ambiguity between the recovered albedo and illumination, the delit faces in (c) and (f) look slightly darker than the ones in (b) and (e).

(a) (b) (c) (d)

Fig. 12. Face delighting experiment on images under occlusions. (a) and (c) The original images under different occlusions. (b) and (d) The recovered albedo from our method. Our MRF-based method can generate high-quality results for the occluded areas as well.

Light Positions	SHBMM-Based Method		MRF-Based Method
	Using SHB Parameters	Using De-lit Images	
2	68.23%	70.74%	97.35%
6	91.14%	97.65%	99.55%
8	93.31%	98.93%	99.22%
12	92.66%	99.12%	99.93%
16	67.92%	69.52%	97.28%
20	89.62%	98.04%	99.86%

Fig. 13. Face recognition under different illumination conditions: We evaluate and compare the recognition performance of the SHBMM- and MRF-based methods. The same database and experiment setting are used as in Fig. 5. The results show that both methods achieve high recognition rates for images under a wide range of illumination conditions. However, in the extreme illumination conditions, such as light positions 2 and 16, the performance of the SHBMM-based methods decreases, while the MRF-based method is more robust to illumination variation and can maintain a good recognition performance under extreme lighting conditions.

or even better face recognition performance without estimating the illumination conditions, albeit with the requirement of multiple training images.

Furthermore, to compare the overall performance of our recognition methods under a wide range of illumination conditions, we tested both SHBMM and MRF-based methods on images under both single and multiple directional illumination sources. More specifically, to study the performance of our methods on images taken under multiple directional illumination sources, we synthesized images by combining face images under different illumination conditions in the CMU-PIE Database. For each subject, we randomly selected two-four images under single directional illuminant from the training data set and combined them together with random weights to simulate face images under multiple directional illumination sources. As a result, for each subject, there are 40 images under different illuminations. In the recognition step, one image of each subject was picked as the gallery set and the remaining images were used for testing. We performed the random selection five times and reported the average recognition rates in Fig. 14. Because the SHBMM-based method could not handle extreme lighting conditions, we did not include images with large illumination angles, such as light positions 2 and 16, in this experiment.

The comparison in Fig. 14 shows that both methods have high recognition rates on images even under multiple lighting sources. The proposed MRF-based method has a better recognition performance and improved robustness to illumination variation than the SHBMM-based method.

9 Conclusion

In this paper, we proposed a new framework to estimate lighting, shape, and albedo of a human face from a single image, which can even be taken under extreme lighting conditions and/or with partial occlusions. The proposed method includes two parts. In the first part, we introduced a 3D SHBMM that integrates spherical harmonics into the morphable model framework to represent faces under arbitrary lighting conditions. To handle extreme lighting conditions, we proposed a spatially varying texture morphable model in the second part to jointly recover the lighting,

Recognition Methods		Recognition Rate
SHBMM-Based Method	Using De-lit Images	98.7%
	Using SHB Parameters	91.9%
MRF-Based Method		99.3%

Fig. 14. Recognition results on images under both single and multiple directional illumination sources from the CMU-PIE Database. There are 40 images for each subject under different illuminations. One image of each subject is randomly picked as the gallery set (prior images of people to be recognized) and the remaining images are used for testing. We perform the random selection five times and report the average recognition rates. Because the SHBMM-based method could not handle extreme lighting conditions, images with large illumination angles are not included in this experiment, such as light positions 2 and 16, for fair comparison. The results show that the MRF-based method has a higher recognition rate than the SHBMM-based method.

shape, and albedo from a single face image under arbitrary unknown illumination. Different from existing methods in the literature, we decouple the estimation of the texture and illumination through a region-based scheme and incorporate the MRF constraints to ensure spatial coherence between adjacent regions. Our technique is robust to extreme lighting conditions, partial occlusions, cast shadows, and saturated image regions. We demonstrated the performance of our proposed framework through both face relighting and face recognition experiments on two publicly available face data sets: Yale Face Database B [17] and CMU-PIE Database [36]. In the future, we plan to further improve the results by incorporating face skin reflectance models and extend the current model to recover face geometry and texture under different facial expressions.

APPENDIX

As discussed in Section 6.2, the objective function F to be minimized in the MRF-based approach is

$$\begin{aligned}\mathcal{F} = &\sum_{q=1}^{Q} \sum_{(u,v)\in\Omega_q} \left\{ W_{u,v}\left(I_{u,v} - \rho_{u,v} \sum_{i=1}^{9} h_i\big(\vec{n}^M_{u,v}\big) l_i \right)^2 \right. \\ &\left. + W_{MM}\big(\rho_{u,v} - \rho^q_{u,v}\big)^2 \right\} \\ &+ W_{SM} N_{sr} \sum_{(i,j)\in\mathcal{N}} \sum_{k=1}^{m-1} \left(\frac{\beta^i_k - \beta^j_k}{\sigma^{ij}_k} \right)^2,\end{aligned}$$

where

$$\vec{n}^M_{u,v} = \left(\overline{\vec{n}_{u,v}} + \sum_{j=1}^{m-1} \lambda_j \vec{n}^j_{u,v} \right) \Bigg/ \left\| \overline{\vec{n}_{u,v}} + \sum_{j=1}^{m-1} \lambda_j \vec{n}^j_{u,v} \right\|.$$

We can derive the gradients of the objective function F as follows:

$$\frac{\partial \mathcal{F}}{\partial l_i} = 2 \sum_{q=1}^{Q} \sum_{(u,v)\in\Omega_q} \left\{ W_{u,v}\left(\rho_{u,v} \sum_{i=1}^{9} h_i\big(\vec{n}^M_{u,v}\big) l_i - I_{u,v} \right) \rho_{u,v} h_i\big(\vec{n}^M_{u,v}\big) \right\}, \tag{23}$$

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \rho_{u,v}} = 2 \Bigg\{ & W_{u,v}\left(\rho_{u,v} \sum_{i=1}^{9} h_i\big(\vec{n}^M_{u,v}\big) l_i - I_{u,v} \right) \sum_{i=1}^{9} h_i\big(\vec{n}^M_{u,v}\big) l_i \\ & + W_{MM}\big(\rho_{u,v} - \rho^q_{u,v}\big) \Bigg\},\end{aligned} \tag{24}$$

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \lambda_j} = 2 \sum_{(u,v)\in\Omega_q} & \left\{ W_{u,v}\left(\rho_{u,v} \sum_{i=1}^{9} h_i\big(\vec{n}^M_{u,v}\big) l_i - I_{u,v} \right) \right. \\ & \left. \times \left(\rho_{u,v} \sum_{i=1}^{9} \frac{\partial h_i\big(\vec{n}^M_{u,v}\big)}{\partial \lambda_j} l_i \right) \right\}.\end{aligned} \tag{25}$$

Note that given the albedo ρ, instead of computing the gradient $\frac{\partial \mathcal{F}}{\partial \beta}$, the texture PCA coefficients β are updated directly by projecting ρ to the texture PCA space.

Given the spherical harmonic bases in (5), we can derive the analytic form for each $\frac{\partial h_i(\vec{n}^M_{u,v}))}{\partial \lambda_j}$ term $(i = 1 \ldots 9)$ in (25) as follows (for clarity, we use a simple notation $\vec{n}$ for the normal instead of the original one $\vec{n}^M_{u,v}$ in (25)):

$$\begin{aligned}&\frac{\partial h_1(\vec{n}))}{\partial \lambda_j} = 0, \ \frac{\partial h_2(\vec{n}))}{\partial \lambda_j} = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} \frac{\partial \vec{n}_x}{\partial \lambda_j}, \\ &\frac{\partial h_3(\vec{n}))}{\partial \lambda_j} = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} \frac{\partial \vec{n}_y}{\partial \lambda_j}, \ \frac{\partial h_4(\vec{n}))}{\partial \lambda_j} = \frac{2\pi}{3}\sqrt{\frac{3}{4\pi}} \frac{\partial \vec{n}_z}{\partial \lambda_j}, \\ &\frac{\partial h_5(\vec{n}))}{\partial \lambda_j} = \frac{3\pi}{4}\sqrt{\frac{5}{4\pi}} \vec{n}_z \frac{\partial \vec{n}_z}{\partial \lambda_j}, \\ &\frac{\partial h_6(\vec{n}))}{\partial \lambda_j} = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} \left(\vec{n}_y \frac{\partial \vec{n}_x}{\partial \lambda_j} + \vec{n}_x \frac{\partial \vec{n}_y}{\partial \lambda_j} \right), \\ &\frac{\partial h_7(\vec{n}))}{\partial \lambda_j} = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} \left(\vec{n}_z \frac{\partial \vec{n}_x}{\partial \lambda_j} + \vec{n}_x \frac{\partial \vec{n}_z}{\partial \lambda_j} \right), \\ &\frac{\partial h_8(\vec{n}))}{\partial \lambda_j} = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} \left(\vec{n}_z \frac{\partial \vec{n}_y}{\partial \lambda_j} + \vec{n}_y \frac{\partial \vec{n}_z}{\partial \lambda_j} \right), \\ &\frac{\partial h_9(\vec{n}))}{\partial \lambda_j} = \frac{3\pi}{4}\sqrt{\frac{5}{12\pi}} \left(\vec{n}_x \frac{\partial \vec{n}_x}{\partial \lambda_j} - \vec{n}_y \frac{\partial \vec{n}_y}{\partial \lambda_j} \right),\end{aligned}$$

where

$$\frac{\partial \vec{n}_x}{\partial \lambda_j} = \frac{\vec{n}^j_x}{\|\vec{N}\|} - \frac{\vec{N}_x}{\|\vec{N}\|^3} \left(\vec{N}_x \vec{n}^j_x + \vec{N}_y \vec{n}^j_y + \vec{N}_z \vec{n}^j_z \right),$$

$$\frac{\partial \vec{n}_y}{\partial \lambda_j} = \frac{\vec{n}^j_y}{\|\vec{N}\|} - \frac{\vec{N}_y}{\|\vec{N}\|^3} \left(\vec{N}_x \vec{n}^j_x + \vec{N}_y \vec{n}^j_y + \vec{N}_z \vec{n}^j_z \right),$$

$$\frac{\partial \vec{n}_z}{\partial \lambda_j} = \frac{\vec{n}^j_z}{\|\vec{N}\|} - \frac{\vec{N}_z}{\|\vec{N}\|^3} \left(\vec{N}_x \vec{n}^j_x + \vec{N}_y \vec{n}^j_y + \vec{N}_z \vec{n}^j_z \right),$$

where $\vec{N} = \overline{\vec{n}} + \sum_{j=1}^{m-1} \lambda_j \vec{n}^j$ and the subscripts x, y, and z stand for the x, y, and z component of the vector $\vec{n}$ (and $\vec{N}$), respectively.

ACKNOWLEDGMENTS

The authors would like to thank Phil Chou for his support and helpful discussions. This work was partially supported by the US Government VACE program and by the grants: NIH R01 MH051435, US National Science Foundation (NSF) ACI-0313184, CNS-0627645, and DOJ 2004-DD-BX-1224.

REFERENCES

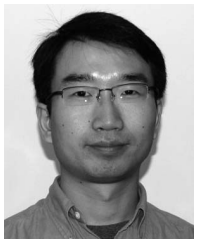
[1] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Derection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721-732, July 1997.

[2] R. Basri and D. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218-233, 2003.

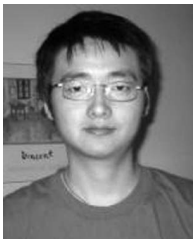
[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.

[4] P. Belhumeur and D. Kriegman, "What Is the Set of Images of an Object under All Possible Lighting Conditions," *Int'l J. Computer Vision*, vol. 28, no. 3, pp. 245-260, 1998.

[5] V. Blanz, S. Romdhani, and T. Vetter, "Face Identification across Different Poses and Illumination with a 3D Morphable Model," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 202-207, 2002.
[6] V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel, "Exchanging Faces in Images," *Proc. EuroGraphics,* 2004.
[7] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces," *Proc. SIGGRAPH,* pp. 187-194, 1999.
[8] 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.
[9] B. Cabral, M. Olano, and P. Nemec, "Reflection Space Image Based Rendering," *Proc. SIGGRAPH,* pp. 165-170, 1999.
[10] R. Chellappa, C. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE,* vol. 83, no. 5, pp. 705-740, May 1995.
[11] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active Appearance Models," *Proc. European Conf. Computer Vision,* pp. 484-498, 1998.
[12] P.E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the Reflectance Field of a Human Face," *Proc. SIGGRAPH,* pp. 145-156, 2000.
[13] M. Dimitrijevic, S. Ilic, and P. Fua, "Accurate Face Models from Uncalibrated and Ill-Lit Video Sequences," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 1034-1041 2004.
[14] R. Epstein, P. Hallinan, and A. Yullie, "5 +/− 2 Eigenimages Suffice: An Empirical Investigation of Low Dimensional Lighting Models," *Proc. IEEE Workshop Physics Based Vision,* pp. 108-116, 1995.
[15] J. Foley and A.V. Dam, *Fundamentals of Interactive Computer Graphics.* Addison-Wesley, 1984.
[16] W. Freeman, E. Pasztor, and O. Carmichael, "Learning Low-Level Vision," *Int'l J. Computer Vision,* vol. 40, no. 1, pp. 25-47, 2000.
[17] A. Georghiades, P. Belhumeur, and D. 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.
[18] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 995-999, 1994.
[19] E. Land and J. McCann, "Lightness and Retinex Theory," *J. Optical Soc. Am.,* vol. 61, no. 1, pp. 1-11, 1971.
[20] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju, "A Bilinear Illumination Model for Robust Face Recognition," *Proc. Int'l Conf. Computer Vision,* pp. 1177-1184, 2005.
[21] K.-C. Lee, J. Ho, and D. Kriegman, "Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 357-362, 2001.
[22] Z. Liu, Y. Shan, and Z. Zhang, "Expressive Expression Mapping with Ratio Images," *Proc. SIGGRAPH,* pp. 271-276, 2001.
[23] C. Loscos, G. Drettakis, and L. Robert, "Interactive Virtual Relighting of Real Scenes," *IEEE Trans. Visualization and Computer Graphics,* vol. 6, no. 4, pp. 289-305, Oct.-Dec. 2000.
[24] Q.-T. Luong, P. Fua, and Y.G. Leclerc, "Recovery of Reflectances and Varying Illuminants from Multiple Views," *Proc. European Conf. Computer Vision—Part III,* pp. 163-179, 2002.
[25] S.R. Marschner, B. Guenter, and S. Raghupathy, "Modeling and Rendering for Realistic Facial Animation," *Rendering Techniques,* pp. 231-242, Springer, 2000.
[26] S.R. Marschner, S. Westin, E. Lafortune, K. Torance, and D. Greenberg, "Image-Based brdf Measurement Including Human Skin," *Proc. Eurographics Workshop Rendering Techniques,* 1999.
[27] I. Matthews and S. Baker, "Active Appearance Models Revisited," *Int'l J. Computer Vision,* vol. 60, no. 2, pp. 135-164, Nov. 2004.
[28] R. Ramamoorthi and P. Hanrahan, "An Efficient Representation for Irradiance Environment Maps," *Proc. SIGGRAPH,* pp. 497-500, 2001.
[29] R. Ramamoorthi and P. Hanrahan, "A Signal-Processing Framework for Inverse Rendering," *Proc. SIGGRAPH,* pp. 117-128, 2001.
[30] T. Riklin-Raviv and A. Shashua, "The Quotient Image: Class Based Re-Rendering and Recongnition with Varying Illuminations," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 566-571, 1999.
[31] S. Romdhani and T. Vetter, "Efficient, Robust and Accurate Fitting of a 3d Morphable Model," *Proc. Int'l Conf. Computer Vision,* pp. 59-66, 2003.
[32] E. Sali and S. Ullman, "Recognizing Novel 3D Objects Under New Illumination and Viewing Position Using a Small Number of Examples," *Proc. Int'l Conf. Computer Vision,* pp. 153-161, 1998.
[33] D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc, "Variable Albedo Surface Reconstruction from Stereo and Shape from Shading," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 480-487, 2000.
[34] I. Sato, Y. Sato, and K. Ikeuchi, "Acquiring a Radiance Distribution to Superimpose Virtual Objects onto a Real Scene," *IEEE Trans. Visualization and Computer Graphics,* vol. 5, no. 1, pp. 1-12, Jan.-Mar. 1999.
[35] M. Savvides, B.V. Kumar, and P. Khosla, "'Corefaces'—Robust Shift Invariant PCA Based Correlation Filter for Illumination Tolerant Face Recognition," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 834-841, 2004.
[36] T. Sim, S. Baker, and M. Bsat, "The cmu Pose, Illumination, and Expression Database," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 12, pp. 1615-1618, Dec. 2003.
[37] T. Sim and T. Kanade, "Combining Models and Exemplars for Face Recognition: An Illuminating Example," *Proc. Workshop Models versus Exemplars in Computer Vision,* 2001.
[38] A. Stoschek, "Image-Based Re-Rendering of Faces for Continuous Pose and Illumination Directions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 582-587, 2000.
[39] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-96, 1991.
[40] B.V. Kumar, M. Savvides, and C. Xie, "Correlation Pattern Recognition for Face Recognition," *Proc. IEEE,* vol. 94, no. 11, pp. 1963-1976, Nov. 2006.
[41] H. Wang, S. Li, Y. Wang, and J. Zhang, "Self Quotient Image for Face Recognition," *Proc. Int'l Conf. Image Processing,* pp. 1397-1400, 2004.
[42] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras, "Face Re-Lighting from a Single Image under Harsh Lighting Conditions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2007.
[43] Z. Wen, Z. Liu, and T.S. Huang, "Face Relighting with Radiance Environment Maps," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 158-165, 2003.
[44] S. Yan, M. Li, H. Zhang, and Q. Cheng, "Ranking Prior Likelihood Distributions for Bayesian Shape Localization Framework," *Proc. Int'l Conf. Computer Vision,* pp. 51-58, 2003.
[45] L. Zhang and D. Samaras, "Pose Invariant Face Recognition under Arbitrary Unknown Lighting Using Spherical Harmonics," *Proc. Int'l Workshop Biometric Authentication,* pp. 10-23, 2004.
[46] L. Zhang and D. Samaras, "Face Recognition from a Single Training Image under Arbitrary Unknown Lighting Using Spherical Harmonics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 3, pp. 351-363, Mar. 2006.
[47] 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," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 209-216, 2005.
[48] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, 2003.
[49] W. Zhao and R. Chellappa, "Illumination Insensitive Face Recognition Using Symmetric Shape-from-Shading," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 286-293, 2000.
[50] S. Zhu, C. Guo, Y. Wang, and Z. Xu, "What Are Textons?" *Int'l J. Computer Vision,* vol. 62, nos. 1/2, pp. 121-143, 2005.



Yang Wang received the PhD degree from the Department of Computer Science at the State University of New York at Stony Brook in 2006. He is a research scientist in the Integrated Data Systems Department at Siemens Corporate Research, Princeton, New Jersey. Prior to that, he was a postdoctoral fellow in the Robotics Institute at Carnegie Mellon University from 2006 to 2008. He specializes in nonrigid motion tracking, facial expression analysis and synthesis, and illumination modeling. He is a member of the ACM, the IEEE, and Sigma Xi.



Lei Zhang received the bachelor's degree in computer science from Nanjing University in 1999, continuing there as a graduate student for two years, and the doctorate degree in computer science from Stony Brook University in 2006, specializing in computer vision and pattern recognition. From January 2006 to April 2007, he worked at Siemens Corporate Research, after which he moved to the financial industry.

Zicheng Liu received the BS degree in mathematics from Huazhong Normal University, Wuhan, China, the MS degree in operational research from the Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China, and the PhD degree in computer science from Princeton University. He is a researcher at Microsoft Research, Redmond, Washington. Before joining Microsoft, he worked as a member of the technical staff at Silicon Graphics, focusing on trimmed NURBS tessellation for CAD model visualization. His research interests include linked figure animation, face modeling and animation, face relighting, image segmentation, event detection and recognition, and multimedia signal processing. He is an associate editor of *Machine Vision and Applications*. He was the cochair of the 2003 IEEE International Workshop on Multimedia Technologies in e-Learning and Collaboration, Nice, France. He was a program cochair of the 2006 International Workshop on Multimedia Signal Processing (MMSP), Victoria, Canada. He was an electronic media cochair of the 2007 International Conference and Multimedia and Expo, Beijing, China. He is a senior member of the IEEE.

Gang Hua received the BS degree in automatic control engineering in 1999 and the MS degree in pattern recognition and intelligence system from Xian Jiaotong University (XJTU), Xian, China, in 2002, respectively, and the PhD degree in electrical and computer engineering from Northwestern University in June 2006. He is a scientist at Microsoft Live Labs Research. His current research interests include computer vision, machine learning, visual recognition, intelligent image/video/multimedia processing, visual motion and content analysis, and their applications to multimedia search. He was a research assistant of Professor Ying Wu in the Computer Vision Group of Northwestern University from 2002 to 2006. During summer 2005 and summer 2004, he was a research intern with the Speech Technology Group, Microsoft Research, Redmond, Washington, and a research intern with the Honda Research Institute, Mountain View, California, respectively. Before coming to Northwestern, he was a research assistant in the Institute of Artificial Intelligence and Robotics at XJTU. He was enrolled in the Special Class for the Gifted Young of XJTU in 1994. He received the Richter Fellowship and the Walter P. Murphy Fellowship at Northwestern University in 2005 and 2002, respectively. While he was at XJTU, he was awarded the Guanghua Fellowship, the Eastcom Fellowship, the Most Outstanding Student Exemplar Fellowship, the Sea-Star Fellowship, and the Jiangyue Fellowship in 2001, 2000, 1997, 1997, and 1995, respectively. He was also a recipient of the University Fellowship from 1994 to 2001 at XJTU. He is a member of the IEEE. As of May 2008, he holds one US patent and has 10 more patents pending.

Zhen Wen received the PhD degree in computer science from the University of Illinois at Urbana-Champaign. He is a research staff member at the IBM T.J. Watson Research Center. His research interests include visualization, computer graphics, machine learning, pattern recognition, and multimedia systems. His thesis work was on improving human-computer interaction using human face avatars. At IBM, his current research focuses on intelligent user interfaces for information analysis. His work received a Best Paper Award at the ACM Intelligent User Interfaces (IUI) Conference in 2005 and an IBM Research Division Award. He serves on technical committees for major conferences such as ACM Multimedia, ACM IUI, and IEEE Multimedia.

Zhengyou Zhang received the BS degree in electronic engineering from the University of Zhejiang, China, in 1985, the MS degree in computer science from the University of Nancy, France, in 1987, the PhD degree in computer science from the University of Paris XI, France, in 1990, and the Doctor of Science (Habilitation diriger des recherches) diploma from the University of Paris XI in 1994. He is a principal researcher with Microsoft Research, Redmond, Washington, and manages the human-computer interaction and multimodal collaboration group. He was with INRIA for 11 years and was a senior research scientist from 1991 until he joined Microsoft Research in March 1998. From 1996 to 1997, he spent a one-year sabbatical as an invited researcher at the Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. He is a fellow of the IEEE, a member of the IEEE Computer Society Fellows Committee since 2005, the chair of IEEE Technical Committee on Autonomous Mental Development, and a member of the IEEE Technical Committee on Multimedia Signal Processing. He is currently an associate editor of several international journals, including the *IEEE Transactions on Multimedia (TMM)*, the *International Journal of Computer Vision (IJCV)*, the *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, and the *Machine Vision and Applications Journal (MVA)*. He served on the editorial board of the *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* from 2000 to 2004, among others. He holds more than 50 US patents and has about 40 patents pending. He also holds a few Japanese patents for his inventions during his sabbatical at ATR. He has published more than 160 papers in refereed international journals and conferences, edited three special issues, and coauthored three books: *3D Dynamic Scene Analysis: A Stereo Based Approach* (Springer, 1992), *Epipolar Geometry in Stereo, Motion and Object Recognition* (Kluwer Academic Publishers, 1996), and *Computer Vision* (textbook in Chinese, Science Publishers, 1998, 2003). He has been on the organization or program committees for numerous international conferences, and was a program cochair of the Asian Conference on Computer Vision (ACCV '04), a technical cochair of the International Workshop on Multimedia Signal Processing (MMSP '06), and a program cochair of the International Workshop on Motion and Video Computing (WMVC '07).

Dimitris Samaras received the 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. He is an associate professor in the Department of Computer Science at the State Unversity of New York at Stony Brook, where he has been working since September 2000. 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 the IEEE.