Face Re-Lighting from a Single Image under Harsh Lighting Conditions

Yang Wang[1], Zicheng Liu[2], Gang Hua[3], Zhen Wen[4], Zhengyou Zhang[2], Dimitris Samaras[5]

[1]The Robotics Institute, Carnegie Mellon University, wangy@cs.cmu.edu
[2]Microsoft Research, {zliu,zhang}@microsoft.com
[3]Microsoft Live Labs Research, ganghua@microsoft.com
[4]IBM T.J.Watson Research Center, zhenwen@us.ibm.com
[5] Computer Science Department, Stony Brook University, samaras@cs.sunysb.edu

Abstract

In this paper, we present a new method to change the illumination condition of a face image, with unknown face geometry and albedo information. This problem is particularly difficult when there is only one single image of the subject available and it was taken under a harsh lighting condition. 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 spherical harmonic representation. However, the approximation error can be large under harsh lighting conditions [2] 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 harsh lighting conditions, but insensitive to partial occlusions as well. The performance of our framework is demonstrated through various experimental results, including the improvement to the face recognition rate under harsh lighting conditions.

1. Introduction

Recovering the geometry and texture of a human face from images remains to be a very important but challenging problem, with wide applications in both computer vision and computer graphics. One typical application is to generate photo-realistic images of human faces under arbitrary lighting conditions [28, 8, 31, 11, 24, 23]. This problem is particularly difficult when there is only one single image of the subject available and it was taken under a harsh lighting condition. Using spherical harmonic representation [2, 26], 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. However, under harsh lighting conditions,

the approximation error can be large [2], which remains an unsolved problem for both graphics and vision applications, such as face relighting and face recognition. Furthermore, this problem becomes even more challenging in the presence of cast shadows, saturated areas, and partial occlusions.

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. But there are two main problems with such a region based approach. First, if the majority of the pixels in a 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 be inconsistent 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 it is well known in other fields such as super-resolution and texture synthesis [15, 43], Markov Random Field (MRF) is an effective theoretical framework 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.

All these distinguish our approach from previous techniques such as the one proposed by Zhang *et al*. [40]. They used a 3D spherical harmonic basis morphable model (SHBMM) by adding the spherical harmonic illumination representation into the morphable model method. It produces photo-realistic rendering results under regular lighting conditions, but obtains poor results in the saturated face image area. Furthermore, because the texture is not separated from the spherical harmonic bases in SHBMM, their

method can not handle the harsh lighting conditions due to the large approximation errors in the spherical harmonic representation. To address these problems, we decouple the texture, the geometry (including the surface normal), and the illumination by modeling them separately. Compared to the 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 harsh 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 harsh 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. The experimental results further demonstrate the superb performance of our approach.

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 photo-realistic images of objects including human faces [11, 37, 12, 40] and face recognition under different lighting conditions [1, 30, 42, 16, 20, 34]. Marschner *et al.* [24, 25] measured the geometry and reflectance field of faces from a large number of image samples in a controlled environment. Georghiades *et al.* [16] and Debevec *et al.* [11] used a linear combination of basis images to represent face reflectance. Ramamoorthi and Hanrahan [27] presented a signal processing framework for inverse rendering which provides analytical tools to handle general lighting conditions.

Furthermore, Sato *et al.* [32] and Loscos *et al.* [22] used the ratio of illumination to modify the input image for relighting. Interactive relighting was achieved in [22, 37] 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 [28] used the color ratio (called the quotient image) to generate an image of the second face under the second lighting condition. Stoschek [35] combined the quotient image with image morphing to generate re-lit faces under continuous changes of poses. Recently, Liu *et al.* [21] used the 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 photo-realistic 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 [41, 9]. Georghiades *et al.* [16] presented a new method using the illumination cone. Sim and Kanade [34] proposed a model and exemplar based approach for recognition. Both [16] and [34] 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, 29] and face image synthesis [7, 6]. The illumination effects are modeled by the Phong model [14]. In order to handle more general lighting conditions, Zhang *et al.* [40] integrated the spherical harmonic illumination representation into the Morphable Model approach, by modulating the texture component with the spherical harmonic bases.

Generally, in order to handle the illumination variability, appearance-based methods such as Eigenfaces [36] and AAM [10] 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, 26, 13, 17]. Using the spherical harmonic presentation of Lambertian reflection, Basri *et al.* [2] and Ramamoorthi [26] 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 9 single light source images of a face is reported in [20]. However, to use these recognition schemes, the basis images spanning the illumination space for each face are required. Zhao and Chellappa [42] 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 [39] proposed to recover the 9 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.* [19] used a bilinear illumination model to reconstruct a shape-specific illumination subspace. However, it requires a large dataset 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, 26]. After that, a new subregion based framework is proposed to recover both the shape, texture, and illumination from an input face image, by incorporating the statistical distribution and spatial coherence of face texture. The proposed method decouples the texture from

the geometry and illumination models and integrates them into an energy minimization problem 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 T_{model} of a new face can be generated by a convex combination of the shapes and texture of the m exemplar 3D faces, i.e.,

$$S_{model} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i; \; T_{model} = \overline{T} + \sum_{i=1}^{m-1} \beta_i t_i \quad (1)$$

where 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 [29], a realistic face shape can be generated by

$$S^{2D} = fPR(\overline{S^{3D}} + \sum_{i=1}^{m-1} \alpha_i s_i^{3D} + t^{3D}) + t^{2D}, \quad (2)$$

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. The t^{3D} and t^{2D} are translation vectors in 3D and 2D respectively. 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 pre-selected feature points in the 3D Morphable Model and their correspondences S_f^{img} detected in the target image:

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

where $\overline{S(F)^{3D}}$ and $s_i(F)^{3D}$ are the shape of the corresponding feature points in the Morphable Model in Eqn.(1).

3.2. Spherical Harmonics Representation

In general, spherical harmonics are the sphere analog of the Fourier basis 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 9 spherical harmonic bases [2, 26]:

$$I(\vec{n}) = \rho(\vec{n})E(\vec{n}) \approx \rho(\vec{n}) \sum_{i=1}^{9} h_i(\vec{n}) \cdot l_i \quad (4)$$

where I denotes the image intensity, $\vec{n}$ the surface normal, ρ the surface albedo, E the irradiance, l_i the weighting co-

efficient, and h_i the spherical harmonic basis as follows:

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

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.,

$$I \approx [\rho_1 H_1, \rho_2 H_2, \ldots, \rho_n H_n]^T \cdot l \quad (6)$$

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.

3.3. Energy Minimization Framework

Since lighting in smaller image regions is more homogeneous, we subdivide a face into smaller regions to better fit the image under a harsh lighting condition. The idea of subdivision was also used by Blanz and Vetter in [7], where a face is subdivided along feature boundaries (such as eyes, nose, mouth, etc.) to increase the expressiveness of the morphable models. They estimate morphable model parameters independently over each region and perform smoothing along region boundaries to avoid visual discontinuity. However, this approach can not be applied to images under harsh lighting conditions because of the inconsistency of the estimated textures in different regions (e.g. Fig. 1(c)). 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.

We divide a face into regular regions with a typical size of 50×50 pixels. For each region, we represent its face texture by using a PCA texture model similar to Eqn. (1):

$$\rho^q = \overline{T}^q + \sum_{k=1}^{m-1} \beta_k^q t_k^q, \quad q = 1, \ldots, Q \quad (7)$$

where Q is the total number of regions and t_k^q 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 $(\sigma_k^{q_i q_j})^2$, and 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

$$\sum_{k=1}^{m-1} (\frac{\beta_k^{q_i} - \beta_k^{q_j}}{\sigma_k^{q_i q_j}})^2. \tag{8}$$

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 [27] and the perception literature (such as Land's retinex theory [18]), 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

$$\vec{n}_{u,v}^M = (\overline{\vec{n}_{u,v}} + \sum_{i=1}^{m-1} \lambda_i \vec{n}_{u,v}^i) / \|\overline{\vec{n}_{u,v}} + \sum_{i=1}^{m-1} \lambda_i \vec{n}_{u,v}^i\| \quad (9)$$

where λ is the weighting coefficient to be estimated.

Following the discussion in Section 3.2, the illumination model in Eqn.(4) can also be added as another constraint to fit the image I. Note that for pixels which are saturated or in cast shadows, Eqn.(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 above illumination model. $W_{u,v}$ 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:

$$\arg\min_{\rho,\lambda,\beta,l} \sum_{q=1}^{Q} \sum_{(u,v)\in\Omega_q} \{W_{u,v}(I_{u,v} - \rho_{u,v} \sum_{i=1}^{9} h_i(\vec{n}_{u,v}^M)l_i)^2$$
$$+ W_{MM}(\rho_{u,v} - \rho_{u,v}^q)^2\}$$
$$+ W_{SM} N_{sr} \sum_{(i,j)\in\mathcal{N}} \sum_{k=1}^{m-1} (\frac{\beta_k^i - \beta_k^j}{\sigma_k^{ij}})^2 \tag{10}$$

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 Eqn.(9), ρ^q is constrained by the texture subspace defined in Eqn.(7), 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 $(\sigma_k^{ij})^2$ is estimated from the exemplar texture data in the morphable models[7].

The objective function in Eqn. (10) is an energy function of a Markov Random Field. The first two terms in Eqn. (10) 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 Eqn.(9) to the shape constraint in Eqn.(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, $W_{u,v}$ 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 2500. Due to the nonlinearity of the objective function (10), the overall optimization problem is solved in an iterative fashion, i.e., fixing the albedo ρ and the surface normal $\vec{n}$, we solve for the global lighting l, and fixing the lighting l, we solve for the albedo ρ and the surface normal $\vec{n}$. Because gradients of Eqn.(9,10) can be derived analytically, the standard conjugate method is used for the optimization.

However, the linear equation system (6) is underconstrained because the surface albedo ρ varies from point to point. Therefore, it is impossible to obtain the initial lighting l_{init} without any knowledge of the surface albedo ρ. Fortunately, based on the observation in [37], the albedo of a human face, though not constant, does not have low-frequency components other than the constant component. If we expand $\rho(\vec{n})$ by using spherical harmonics as $\rho(\vec{n}) = \rho_{00} + \Psi(\vec{n})$, where ρ_{00} is the constant component and $\Psi(\vec{n})$ contains other higher order components, Eqn.(4) can be further simplified as $\rho(\vec{n})E(\vec{n}) \approx \rho_{00} \sum_{i=1}^{9} h_i(\vec{n}) \cdot l_i$. Consequently, the original under-constrained problem (6) can be approximated by the following linear equation system

$$I \approx \rho_{00} [H_1, H_2, \ldots, H_n]^T \cdot l \tag{11}$$

Therefore, given an image of a face with known surface normal $\vec{n}$, we can solve for the initial values of the 9 spherical harmonic coefficients $l = [l_1, l_2, \ldots, l_9]^T$ using a least squares procedure, up to a constant albedo ρ_{00}.

For clarity purposes, the outline of the optimization algorithm is presented in Table 1, and an example is shown in Fig.1, where Fig.1(a) is the original image taken under

a harsh lighting condition, Fig.1(b) shows the recovered surface normal from our method, and the recovered albedo from our method is shown in Fig.1(c) without the spatial coherence term and Fig.1(d) with the spatial coherence term. As we can see, the region inconsistency artifacts in Fig.1(c) are significantly reduced in Fig.1(d).

1. **Initial Shape Estimation**: Detect the face feature points S_f^{img} on the input image using an automatic face feature detection method [38]. Then, based on the set of detected feature points and the corresponding pre-selected feature points in the 3D Morphable Model, the pose parameters f, ϕ, γ and θ and the shape parameter α can be recovered using Eqn.(3), as described in Section 3.1.

2. **Image Segmentation**: Segment the input face images 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. **Initial Illumination and Albedo Estimation**: Compute the constant albedo scale factor ρ_{00} by averaging the intensity values of the input face image. Estimate the initial lighting coefficient l_{init} using Eqn.(11), based on the constant albedo scale factor ρ_{00} and the initial shape recovered in the first step. After that, the initial albedo ρ_{init} is computed by Eqn. (4).

4. **Iterative Minimization**: Solve the objective function (10) in an iterative fashion – typically, only 2 iterations were used in our experiments to generate photo-realistic 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.

Table 1. The outline of our estimation algorithm.
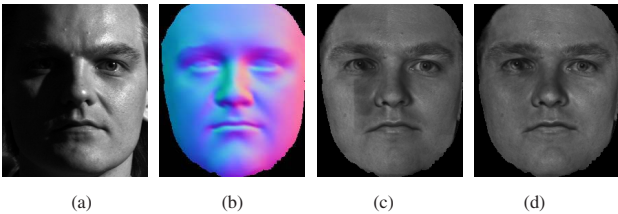

(a)   (b)   (c)   (d)

Figure 1. Example result: (a) is the original image taken under a harsh lighting condition, (b) shows the recovered surface normal 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).

4. Experimental Results

Using the approach proposed in Section 3, 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 re-lighting and face recognition based on the recovered parameters. The advantage of this approach is that it only requires one image as the input. Furthermore, compared to the methods proposed in [37, 39, 40], our proposed framework can also handle images with saturated areas and partial occlusions and is robust to harsh lighting conditions.

4.1. Image Re-Lighting and De-Lighting

Based on the recovered albedo ρ, surface normal $\vec{n}$, and illumination parameter l, we can render a face I' using the recovered parameters by setting different values to the illumination parameter l' [2, 37, 40]:

$$I'(\vec{n}) = \rho(\vec{n}) \sum_{i=1}^{9} h^i(\vec{n}) \cdot l_i' \qquad (12)$$

However, because certain texture details might be lost in the estimated face albedo ρ, we also use the ratio image technique to preserve photo-realistic quality. Although the ratio image technique used in [37, 40], which is based on the spherical harmonic illumination representation, has generated promising results under regular lighting conditions, it can not be adopted in our framework because of the large approximation error in the spherical harmonic approximation for harsh lighting conditions. Instead, we smooth the original image using a Gaussian filter and then compute the pixel-wise ratio between the original image and its smoothed version. After that, we apply this pixel-wise ratio to the re-lit image computed by Eqn.(12) 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 $\sigma = 2$.

Moreover, a de-lighting result of the input image can also be generated by simply replacing the original pixel intensity values with the recovered albedo ρ. In order to evaluate the performance of our framework, we conducted the experiments on two publicly available face data sets: Yale Face Database B[16] and CMU-PIE Database[33]. 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 5 subsets according to the angles of the light source direction from the camera optical axis, i.e., (1) less than $12°$, (2) between $12°$ and $25°$, (3) between $25°$ and $50°$, (4) between $50°$ and $77°$, and (5) larger than $77°$. Fig.2(a) shows one sample image per group of Yale Face Database B. The corresponding re-lit results from our method are shown in Fig.2(c). Compared to the results from Wen *et al.*'s method [37], which are shown in Fig.2(b), the
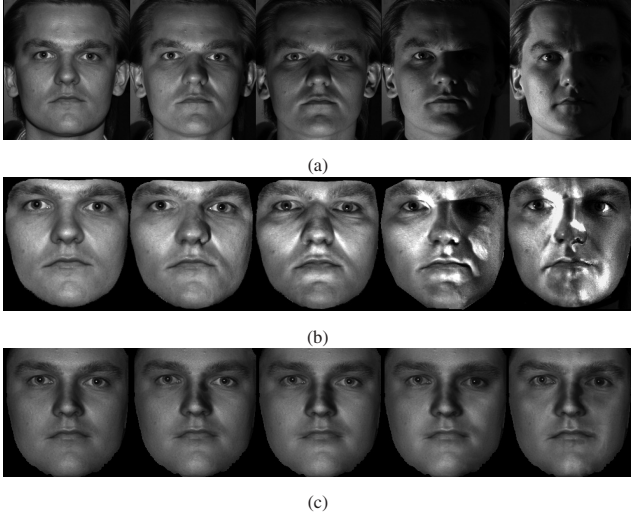
(a)

(b)

(c)

Figure 2. Face re-lighting experiment on Yale Face Database B [16]. (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*. [37]. (c) The re-lit results from our method. As we can see, our method preserves photo-realistic quality, especially under harsh lighting conditions such as the images in rightmost 2 columns, i.e., in group (4-5).
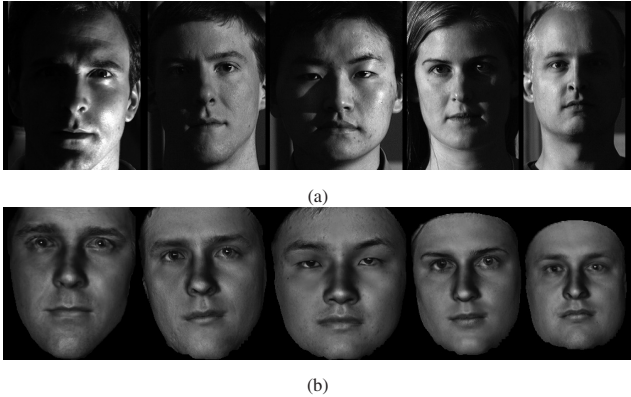

(a)

(b)

Figure 3. Face re-lighting experiment on subjects in both Yale Database B [16] and CMU-PIE Database[33]. (a) Example input images taken under different harsh lighting conditions. (b) The synthesized frontal lighting results generated by our method with high quality.

results generated by our method have much higher quality especially under harsh lighting conditions such as the images in group (4-5). Fig.3 shows more face re-lighting results on both Yale Face Database B [16] and CMU-PIE Database[33]. Despite the different harsh lighting conditions in the input images (Fig.3(a)), our method can still generate high quality re-lit results as shown in Fig.3(b).

Fig.4 shows an example of the face de-lighting experiment on the CMU-PIE Database, where some image pixel values are saturated (Fig.4(a)). Fig.4(b) shows the de-lit image generated by Zhang *et al*.'s method [40], which has poor quality in the saturated area. However, our subregion based

method decouples the estimation of the illumination and albedo and can handle this situation successfully (Fig.4(c)). In the close-up views Fig.4(d-f), we can see that an image with remarkable quality is synthesized by our method even in the presence of saturated areas.

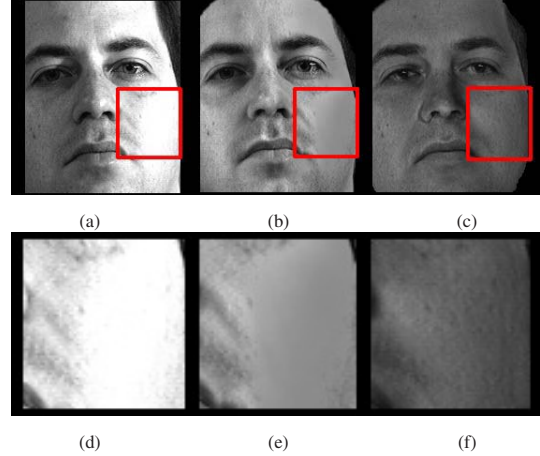(a)   (b)   (c)

(d)   (e)   (f)

Figure 4. Face de-lighting experiment on an image with saturated regions, which is highlighted in the red boxes: (a) The original image where the right side of the face is saturated. (b) The de-lit result from the method proposed by Zhang *et al*. [40] (The image is taken from [40] directly.) (c) The de-lit result from our MRF-based method. The close-up views (d-f) show that a remarkable quality image is synthesized by our method even in the presence of saturated areas.

Furthermore, since our framework models spatial dependence, it can handle image occlusions as well. This is in spirit similar to super resolution and texture synthesis[15, 43]. But we are able to recover missing information and remove lighting effects simultaneously. Fig.5 shows two examples of the face de-lighting experiment on images under occlusions, where Fig.5(a,c) are the original images under different occlusions and Fig.5(b,d) are the recovered albedo from our method. As we can see our method can generate high quality results for the occluded areas as well.

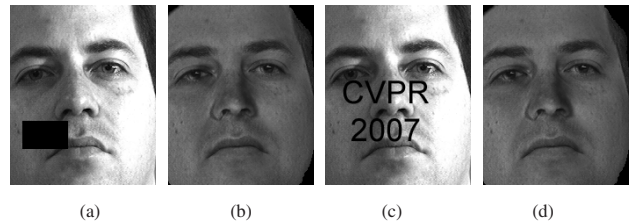(a)   (b)   (c)   (d)

Figure 5. Face de-lighting experiment on images under occlusions: (a,c) are the original images under different occlusions and (b,d) are the recovered albedo from our method. Our method can generate high quality results for the occluded areas as well.

4.2. Face Recognition

In this section, we show that our framework on face re-lighting from a single image can be used for face recog-

nition. 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 Eqn. (12). Once the illumination effects in images are normalized, any existing face recognition algorithms, such as Eigenfaces (PCA) [36] and Fisherfaces (LDA) [3], can be used on the re-lit face images for face recognition. In our experiments, we tested our approach using two publicly available face database: Yale Face Database B [16] and CMU-PIE Database[33].

In Yale Face Database B, there are 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions). In our current experiment, we only consider 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 training image. The remaining 630 images are used as testing images. In order to evaluate the face recognition performance of our proposed method under different lighting conditions, we compare our recognition results with other existing methods in the literature and show experimental results in Fig. 6.

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[39]	0.0	0.3	3.1
BIM (30 Bases)[19]	0.0	0.0	0.7
Wen et al. [37]	0.0	1.7	30.7
Our Method	0.0	0.0	0.1

Figure 6. Recognition results on the Yale Face Database using various previous methods in the literature and our method. Except for Wen et al.'s method [37] and our method, the data were summarized from [19].

We can see that 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 harsh lighting conditions (e.g. lighting angles $> 50°$).

In the CMU-PIE Database, there are 68 subjects with 13 poses and 43 different illumination conditions each pose. As explained above, in our current experiment we focus on illumination variations and perform face recognition for the frontal pose images only. In order to compare our method to the method proposed by Zhang et al. [40] which is closely related to our work, we also tested our method on all 68 subjects on CMU-PIE Database in the same experiment setting and report the recognition result in Fig. 7. We can see that

Methods	Recognition Rate
Zhang et al. [40]	99.3%
Our Method	99.8%

Figure 7. Recognition results on frontal pose images from the CMU-PIE Database. The data of Zhang et al.'s method was summarized from [40].

the face recognition performance is also improved by our new method.

5. Conclusion

In this paper, we proposed a novel MRF-based energy minimization framework to jointly recover the lighting, shape, and albedo from a single face image under arbitrary unknown illumination. Our technique is robust to harsh 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 [16] and CMU-PIE Database[33]. In the future, we plan to further improve the results by incorporating face skin reflectance models and to expand the current model to recover face geometry and texture under different poses and facial expressions.

6. Acknowledgements

References

[1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination derection. *PAMI*, 19:721–732, 1997. 2

[2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. volume 25, pages 218–233, 2003. 1, 2, 3, 5

[3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, pages 711–720, 1997. 7

[4] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions. *IJCV*, 28(3):245–260, 1998. 2

[5] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illumination with a 3d morphable model. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 202–207, 2002. 2

[6] V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel. Exchanging faces in images. In *EuroGraphics*, 2004. 2

[7] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH'99*, pages 187–194. 2, 3, 4

[8] B. Cabral, M. Olano, and P. Nemec. Reflection space image based rendering. In *SIGGRAPH'99*, pages 165–170. 1

[9] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceeding of IEEE*, 83(5):705–740, 1995. 2

[10] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV'98*, pages 484–498. 2

[11] P. E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH'00*, pages 145–156. 1, 2

[12] M. Dimitrijevic, S. Ilic, and P. Fua. Accurate face models from uncalibrated and ill-lit video sequences. In *CVPR'04*, II: 1034-1041. 2

[13] R. Epstein, P. Hallinan, and A. Yullie. 5 +/- 2 eigenimages suffice: An empirical investigation of lowdimensional lighting models. In *IEEE Workshop on PhysicsBased Vision*, pages 108–116, 1995. 2

[14] J. Foley and A. V. Dam. *Fundamentals of interactive computer graphics*. Addison-Wesley, 1984. 2

[15] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *IJCV*, 40(1):25–47, 2000. 1, 6

[16] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643–660, 2001. 2, 5, 6, 7

[17] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *CVPR'94*, pages 995–999. 2

[18] E. Land and J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971. 4

[19] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. A bilinear illumination model for robust face recognition. In *ICCV05*, pages II: 1177–1184. 2, 7

[20] K.-C. Lee, J. Ho, and D. Kriegman. Nine points of light: Accquiring subspaces for face recognition under variable lighting. In *CVPR'01*, pages 357–362. 2

[21] Z. Liu, Y. Shan, and Z. Zhang. Expressive expression mapping with ratio images. In *SIGGRAPH'01*, pages 271–276. 2

[22] C. Loscos, G. Drettakis, and L. Robert. Interactive virtual relighting of real scenes. *IEEE Trans. on Visualization and Computer Graphics*, 6(3), 2000. 2

[23] Q. Luong, P. Fua, and Y. Leclerc. Recovery of reflectances and varying illuminants from multiple views. In *ECCV02*, page III: 163 ff. 1

[24] S. R. Marschner, B. Guenter, and S. Raghupathy. Modeling and rendering for realistic facial animation. In *Rendering Techniques*, pages 231–242. Springer Wien New York, 2000. 1, 2

[25] S. R. Marschner, S. Westin, E. Lafortune, K. Torance, and D. Greenberg. Image-based brdf measurement including human skin. In *Rendering Techniques*, 1999. 2

[26] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH'01*, pages 497–500. 1, 2, 3

[27] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH'01*, pages 117–128. 2, 4

[28] T. Riklin-Raviv and A. Shashua. The quotient image: Class based re-rendering and recongnition with varying illuminations. In *CVPR'99*, pages 566–571. 1, 2

[29] S. Romdhani and T. Vetter. Efficient, robust and accurate fitting of a 3d morphable model. In *ICCV'03*, pages 59–66. 2, 3

[30] E. Sali and S. Ullman. Recognizing novel 3-d objects under new illumination and viewing position using a small number of examples. In *ICCV'98*, pages 153–161. 2

[31] D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *CVPR00*, pages I: 480–487. 1

[32] I. Sato, Y. Sato, and K. Ikeuchi. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *TVCG*, 5(1):1–12, 1999. 2

[33] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *PAMI*, 25(12):1615–1618, 2003. 5, 6, 7

[34] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. In *Proc. of Workshop on Models versus Exemplars in Computer Vision*, 2001. 2

[35] A. Stoschek. Image-based re-rendering of faces for continuous pose and illumination directions. In *CVPR'00*, pages 582–587. 2

[36] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991. 2, 7

[37] Z. Wen, Z. Liu, and T. S. Huang. Face relighting with radiance environment maps. In *CVPR'03*, pages 158–165. 2, 4, 5, 6, 7

[38] S. Yan, M. Li, H. Zhang, and Q. Cheng. Ranking prior likelihood distributions for bayesian shape localization framework. In *ICCV'03*, pages 51–58. 5

[39] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *PAMI*, 28(3):351–363, 2006. 2, 5, 7

[40] 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. In *CVPR'05*, pages II:209–216. 1, 2, 5, 6, 7

[41] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003. 2

[42] W. Zhao and R.Chellappa. Illumination-insensitive face recognition using symmetric shape-from-shading. In *CVPR'00*, pages 286–293. 2

[43] S. Zhu, C. Guo, Y. Wang, and Z. Xu. What are textons? *IJCV*, 62(1-2):121–143, 2005. 1, 6