Image-Driven Re-targeting and Relighting of Facial Expressions

Lei Zhang, Yang Wang, Sen Wang, Dimitris Samaras*
Computer Science Department
SUNY at Stony Brook, NY, USA

Song Zhang†, Peisen Huang‡
Mechanical Engineering Department
SUNY at Stony Brook, NY, USA

ABSTRACT

Synthesis and re-targeting of facial expressions is central to facial animation and often involves significant manual work in order to achieve realistic expressions, due to the difficulty of capturing high quality expression data. Recent progress in dynamic 3-D scanning allows very accurate acquisition of dense point clouds of facial geometry and texture moving at video speeds. Often the new facial expressions need to be rendered in different environments where the illumination is different from the original capture conditions. In this paper we examine the problem of re-targeting captured facial motion under different illumination conditions when the information we have about the face we want to animate is minimal, a single input image. Given an input image of a face, a set of illumination example images (of other faces captured under different illumination) and a facial expression motion sequence, we aim to generate novel expression sequences of the input face under the lighting conditions in the illumination example images. The input image and illumination example images can be taken under arbitrary unknown lighting. In this paper, we propose two methods in which a 3D spherical harmonic morphable model (SHBMM) can generate images under new lighting conditions with remarkable quality even if only one single image under unknown lighting is available, not only for static poses but for dynamic sequences as well where the face is undergoing subtle high-detail motion.

Keywords: Spherical Harmonics, Morphable Models, Facial expressions

1 INTRODUCTION

Synthesis and re-targeting of facial expressions is central to facial animation and often involves significant manual work in order to achieve realistic expressions, due to the difficulty of capturing high quality expression data. Recent progress in dynamic 3-D scanning allows very accurate acquisition of dense point clouds of facial geometry and texture moving at video speeds. Often the new facial expressions need to be rendered in different environments where the illumination is different from the original capture conditions. Great progress has been made in generating photo-realistic images of objects including human faces [8][9][13][39][11]. However, when only one single image under unknown lighting is available, facial expression re-targeting becomes particularly challenging. In this paper we examine the problem of re-targeting captured facial motion under different illumination conditions, when the information about the face we want to animate is minimal, a single input image under arbitrary unknown illumination. Given an input image of a face, a set of illumination example images (of other faces captured under different illumination) and a facial expression motion sequence, we aim to generate novel expression sequences of the input face under the lighting conditions in the illumination example images. The input image and illumination example images can be taken under arbitrary unknown lighting. In this paper, we demonstrate how a 3D spherical harmonic morphable model (SHBMM) generate images under new lighting conditions with remarkable quality even if only one single image under unknown lighting is available, not only for static poses but for dynamic sequences as well where the face is undergoing subtle high-detail motion (see Fig. 1).[1]

*e-mail: {lzhang, yangwang, swang, samaras}@cs.sunysb.edu
†e-mail:szhang@oml.eng.sunysb.edu
‡e-mail:peisen.huang@sunysb.edu

Figure 1: Facial expression re-targeting under novel illumination conditions from a single input image: the top left image is the input image. The images in the first row are synthesized images of different expressions and illumination conditions. The second row shows the synthesized images of the same expression under different illumination conditions.

Facial animation is an active area of research in computer graphics (see [28] for an overview of older work). In 2D facial animation, many advanced examples of talking faces have been produced with image-based methods [6, 12], mainly focused on the mouth region. In 2D methods imaging conditions can only be those of the original video. To allow 3-D animations, several techniques have been developed to create photo-realistic face models from 2D images [30, 5]. Physics-based models are used to simulate the surface deformations caused by muscle forces [24]. Mathematical approximation models include free form deformations [21], B-Spline surfaces [25] and variational approaches [10]. Recently, both static 3-D scans of expressions [5, 3] and time-sequences of 3-D motion [20] have been used to collect 3-D facial expressions. Expression cloning [26] can produce facial animations by reusing existing motion data. Morph-based approaches [30], geometry-based approaches [43, 19] and high level control mechanisms [7] generate photo-realistic facial expressions.

Previous research suggested that illumination variability in face images is low-dimensional e.g. [16][2][35]. Recently, using spherical harmonics, it has been shown [1][32] that the set of images of a

[1]We would like to thank Sudeep Sarkar and Simon Baker for providing databases and Thomas Vetter and Sami Romdhani for helpful discussions. This research was supported by grants from U.S. Department of Justice(2004-DD-BX-1224) and National Science Foundation(ACI-0313184)

convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a 9 dimensional linear subspace spanned by basis images for each face. These images can be rendered from a 3D scan of the face or can be approximated by acquiring a number of images of the same subject under different illuminations [31][23]. This requirement for a number of training images and/or 3D scans of the subjects in the database necessitates specialized equipment and procedures for the capture of the training set. In a single image based approach [40] for face recognition under arbitrary illumination conditions, a statistical model of spherical harmonics is based on a collection of 2D basis images. To recover a new set of basis images, the input image should be accurately aligned with the bootstrap images, hence images obtained under different viewpoints cannot be manipulated efficiently. In this work, we base both statistical models (for shape information and for spherical harmonic bases) and perform statistical analysis directly on 3D to estimate the most appropriate spherical harmonic basis even though we maintain the single training image requirement. The ability to manipulate the spherical harmonic basis in 3D space allows the use of input poses that do not exist in the training data.

Inverse rendering methods [8][9] suggest that one can generate photo-realistic renderings of objects under new lighting conditions by capturing the lighting environment and recovering surface reflectance properties, by using a number of images to model the environment map and face reflectance. In the Quotient image method[37], the set of images generated by varying lighting conditions on a collection of Lambertian objects (same shape different texture) can be characterized using images of a prototype object and a illumination invariant "signature" image per object of the class. In this work, instead of assuming that all faces have the same shape, we use a set of shape parameters to represent shape information for each face and we use the spherical harmonic basis which has been proven to be illumination invariant as our "signature" basis. Related to our work in [39], spherical harmonics approximate the radiance environment map for any given image. The lighting conditions of one person's face can be modified so that it matches the lighting conditions of a different person's face image by assuming two faces have similar skin albedos and using a generic face model. In this work, given an image of a face, we explicitly recover the shape information and estimate the spherical harmonic basis of the face (containing albedo information). Thus the input image and illumination example images can have different skin albedos and poses. In 3D face Morphable Models [5], each face can be represented by linear combinations of a set of 3D face exemplars. The method in [3] is bound to images taken under single directional illumination and requires the knowledge of light direction which is often impractical. In the proposed method, the illumination variations are captured by the spherical harmonic basis, thus, there is no illumination limitation on the input images. Another important difference lies in the process of face synthesis. In the face synthesis applications [5][4] of the Morphable Model, new faces were synthesized by setting different shape and texture parameters, i.e. a new face was represented by a linear combination of a set of 3D face exemplars resulting to the possible loss of detail for the specific face.

By using a 3D Spherical Harmonic Basis Morphable Model (SHBMM), any face under arbitrary illumination conditions can be represented simply by a set of SHBMM parameters: shape parameters, spherical harmonic basis parameters, illumination coefficients and pose parameters. The problem we will discuss in this paper can be stated as following: given an input image of a face, a set of illumination example images (of other faces captured under different illumination) and a facial expression motion sequence, how can we generate novel expression sequence of the input face under the lighting conditions in the illumination example images? The input image and illumination example images can be taken under arbitrary unknown lighting. Our method can be described by the following steps:
1) extracting the motion field from the input expression sequence by using a multi-resolution deformable face model;
2) estimating the SHBMM parameters for both the input image and illumination example images by fitting the Spherical Harmonic Basis Morphable Model to those images and thus removing the illumination from the input image (face "de-lighting") and generate images under new illumination conditions (face "re-lighting");
3) generating a face model as the initial frame with the de-lit or re-lit texture and the recovered shape and aligning the model to the deformable model used in step 1);
4) synthesizing novel facial expression sequences by applying the extracted motion field to the initial frame.

This paper is organized as follows. In the next section, we will briefly describe the Spherical Harmonic Basis Morphable Model and explain how to recover the SHBMM parameters from a single input image. We will also introduce how to perform image de-lighting and re-lighting by using the recovered SHBMM parameters. In Section 3, we will explain the process of facial expression re-targeting and propose two methods for expression re-targeting under novel illumination conditions. In Section 4, we describe our experiments and their results. The final Section presents conclusions and future work directions.

2 Spherical Harmonic Basis Morphable Model and Image De/Re-Lighting

In this section, we will briefly describe the method proposed in [42] of constructing the 3D Spherical Harmonic Basis Morphable Model (SHBMM) and explain how to recover the SHBMM parameters from one single input image under arbitrary lighting. We will also introduce image de-lighting and re-lighting using the constructed SHBMM.

2.1 Spherical Harmonic Basis Morphable Model

A Spherical Harmonic Basis Morphable Model is a 3D model of faces with separate shape and spherical harmonic basis models that are learnt from a set of exemplar spherical harmonic basis. Morphing between faces requires complete sets of correspondences between all of the faces. When building such a model, the shape and spherical harmonic basis spaces are transformed into vector spaces. In our method, we used a collection of 3D faces supplied by USF [5] as the bootstrap data set. For each 3D face, we computed a set of 9 spherical harmonic basis as follows [1]:

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

where the superscripts e and o denote the even and the odd components of the harmonics respectively, λ denote the vector of the object's albedos, n_x, n_y, n_z denote three vectors of the same length that contain the x, y and z components of the surface normals. Further, n_{xy} denote a vector such that $n_{xy,i} = n_{x,i}n_{y,i}$. Thus, we present the a face with 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 $B = (B^1_1, ..., B^9_1, ..., B^1_n, ..., B^9_n)^B \in \Re^{9n}$. Thus, the Spherical

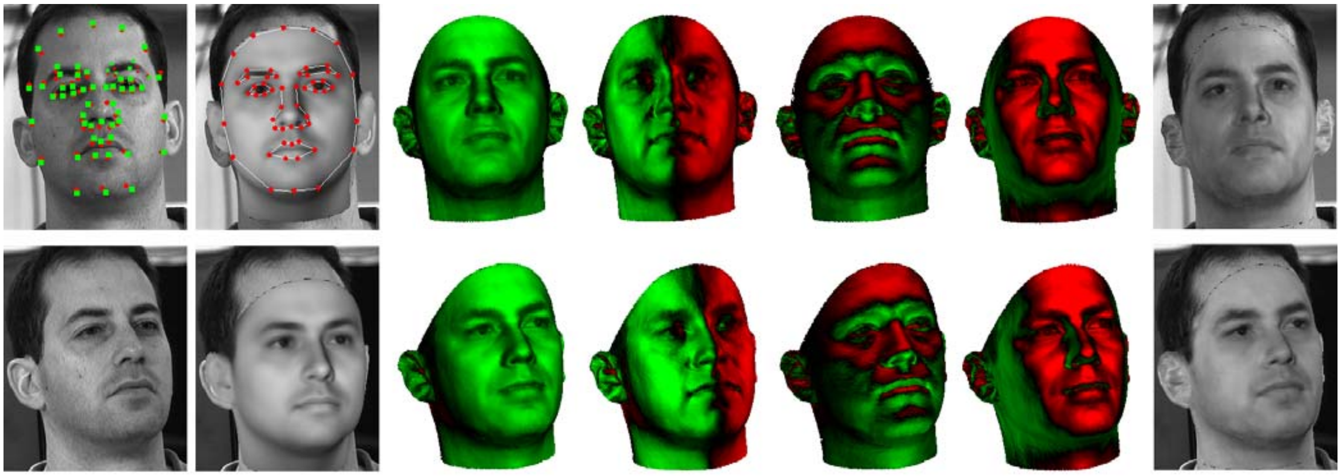

Figure 2: Fitting SHBMM to images: In each row, the first image is the input image followed by initial fitting and recovered spherical harmonic basis (here we show 4 basis instead of 9 due to space limits). The last image is the rendered image using the recovered parameters. In the first row, the red points are pre-picked major features, the green points are the corresponding features and the points lying in the white line are secondary features.

Harmonic basis morphable model can be constructed using a data set of m exemplar faces; exemplar i is represented by the shape-vector S_i and SHB-vector B_i. New shapes s and spherical harmonic basis b can be generated by convex combinations of the shapes and textures of the m exemplar faces as Eq. 2 shows:

$$s = \bar{s} + S\alpha; b = \bar{b} + B\beta \tag{2}$$

With the set of basis images, any image under arbitrary illumination conditions can be approximately represented by the linear combination of the basis as:

$$I \approx b\ell \tag{3}$$

Eq. 3 states that any image I under arbitrary illumination conditions can be represented by the weighted combination of the basis images and we call ℓ in Eq. 3 illumination coefficients.

Thus, combining Eq. 3 and 2, we see that, any face under arbitrary illumination conditions and pose can be represented by four low dimensional vectors (SHBMM parameters): $\{\alpha, \beta, \ell, M\}$ with α representing the geometry parameters, β the spherical harmonic basis parameters ℓ representing the illumination coefficients and M representing the pose parameters including projection and transformation information.

2.2 Fitting SHBMM to Images under Unknown Lighting

2.2.1 Shape Parameters Recovery

Similar to [3]. A realistic face shape can be generated by: $s_{2d} = fPR(\bar{s} + S\alpha + t_{3d}) + t_{2d}$ where f is a scale parameter, P an orthographic projection matrix and R a rotation matrix with ϕ, γ and θ the three rotation angles for the three axes. t_{3d} and t_{2d} are translation vectors in 3D and 2D respectively. Given an input image of a face, 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-picked feature points in SHBMM and their correspondence s_f^{img} in the training image:

$$argmin_{f,\phi,\gamma,\theta,t_{2d}} \| s_f^{img} - (fPR(\bar{s}_f + S_f\alpha + t_{3d}) + t_{2d}) \|^2 \tag{4}$$

where $\bar{s}_f$ and S_f is the corresponding shape information of the feature points in the SHBMM in Equation 2.

2.2.2 Estimating Spherical Harmonic Basis Parameters and Illumination Coefficients

According to Eq. 3 and 2, a realistic face image can be generated by: $I = (\bar{b} + B\beta)\ell$ where $\bar{b} + B\beta$ is the spherical harmonic basis component of the SHBMM and ℓ is the vector of illumination coefficients. Given an input image I_{input} of a face, the spherical harmonic basis parameters β and the illumination coefficients ℓ can be estimated by minimizing the difference between the input image and the rendered image from SHBMM:

$$min_{\beta,\ell} \| (\bar{b} + B\beta)\ell - I_{input} \|^2 \tag{5}$$

Figure 2 shows the fitting process and results. Please refer to [3][40][42][22] for the details of forward and inverse face rendering and the SHBMM parameters recovery.

2.3 Image De-Lighting and Re-Lighting

In the previous section, we demonstrated the recovery the set of SHBMM parameters $\{\alpha_s, \beta_s, \ell_s\}$ from an input face I_s. Inversely, we can render a face $I_s^{'}$ using the recovered parameters to approximate:

$$I_s^{'} = (\bar{b} + B\beta_s)\ell_s \tag{6}$$

Thus, the face texture (de-lit face) can be directly computed from the estimated spherical harmonic basis according to Eq. 1. Hence, face re-lighting can be performed by setting different values to the illumination parameters ℓ similar to [1]. However, in that method, a face was represented by a linear combination of a set of 3D face exemplars which results to possible loss of detail for the specific face. Alternatively, ignoring cast shadows and specularities, we notice that:

$$\frac{I_s}{I_d} = \frac{H(n_t)\lambda_t \ell}{\lambda_t} \approx \frac{H(n_e)\lambda_e \ell}{\lambda_e} = \frac{I_s^{'}}{\lambda_e} \tag{7}$$

where $H(n)\ell$ is the spherical harmonic basis, n_t and n_e are the actual and estimated surface normals and λ_t and λ_e are the real and estimated face textures.

Eq. 7 states that the intensity ratio of the input image to the de-lit image should be equal to that of the rendered face and the corresponding face texture(albedo). The face texture(albedo) of the

Figure 3: Face de-lighting and re-lighting results: the images in the first row are the illumination example images and those in the first column are input images. Images in the second column are the de-lit images. Images with remarkable quality are synthesized even if only one input image is available.

rendered face can be simply computed: $\lambda = \sqrt{4\pi}b_{00}$ according to Eq. 1.

Rewriting Eq. 7, an input image can be de-lit:

$$I_d = \frac{I_s \times \sqrt{4\pi}b_{00}}{(\bar{b}+B\beta_s)\ell_s} \tag{8}$$

Given two images I_s, I_t with the recovered parameters $\alpha_s, \beta_s, \ell_s$ and $\alpha_t, \beta_t, \ell_t$, we have:

$$\frac{I_s}{I_d} = \frac{(\bar{b}+B\beta_s)\ell_s}{\sqrt{4\pi}b^s_{00}} \tag{9}$$

and

$$\frac{I_r}{I_d} = \frac{(\bar{b}+B\beta_s)\ell_t}{\sqrt{4\pi}b^s_{00}} \tag{10}$$

thus, the re-lit image can be computed:

$$I_r = \frac{(\bar{b}+B\beta_s)\ell_t \times I_s}{(\bar{b}+B\beta_s)\ell_s} \tag{11}$$

A simple solution to the cast shadow problem is also provided in [42] by using the image difference (image error E) between input image I_{input} and the re-rendered image from SHBMM. Figure 3 shows examples of face de-lighting and re-lighting.

3 Facial Expression Re-targeting

We will discuss the following problem: given an input image of a face I_s, an expression sequence of another face S_{exp} and a set of illumination example images of other faces I_t, how can we generate new facial expression sequences of the input face simulating the expression S_{exp} under the novel illumination conditions according to the illumination example images I_{target}? In our work, there is no limitation on the illumination conditions of the input image and illumination example images. In this section, we will introduce two methods for facial expression re-targeting under novel illumination conditions.

3.1 Motion Field from Input Expression Sequences

The first step of our method is to extract the motion field from the input expression sequence S_exp. Recent technological advances in digital projection display, digital imaging, and personal computers, are making 3-D shape acquisition in real time increasingly available [17, 41]. In this work, the expression sequences are collected using a structured light technique [38]. The samples returned by our system are not registered in object space and hence there is no guarantee of intra-frame correspondences, which would make tracking of facial features problematic. For this reason, we use a multi-resolution deformable face model. At the coarse level, we use a mesh with 1K nodes that is suitable for facial animation. The coarse mesh was first developed for robust face tracking in low quality 2-D images [15] and extended to 3-D data. This method is fast, and the deformation parameters for each facial motion are few and intuitive. However it cannot capture accurately the large number of local deformations and expression details in our data, so we use it for a coarse-level initial tracking.

The highly local deformations and details in expressions are captured in a second level fitting process. For each frame of the range scan, the resulting mesh from the coarse-level tracking is used to initialize a subdivided refined mesh with 8K nodes. This finer mesh is registered to the frame based on the 3-D extension of a variational algorithm for non-rigid shape registration [18]. This algorithm integrates an implicit shape representation [27] and the cubic B-spline based Free Form Deformations (FFD) model [34, 33], and generates a motion/deformation field that is smooth, continuous and gives dense one-to-one correspondences. In the following sections, the extracted motion field will be employed in expression re-targeting.

3.2 Facial Expression Re-targeting

In this section, we will introduce two methods for facial expression re-targeting under novel illumination conditions by using the extracted motion field.

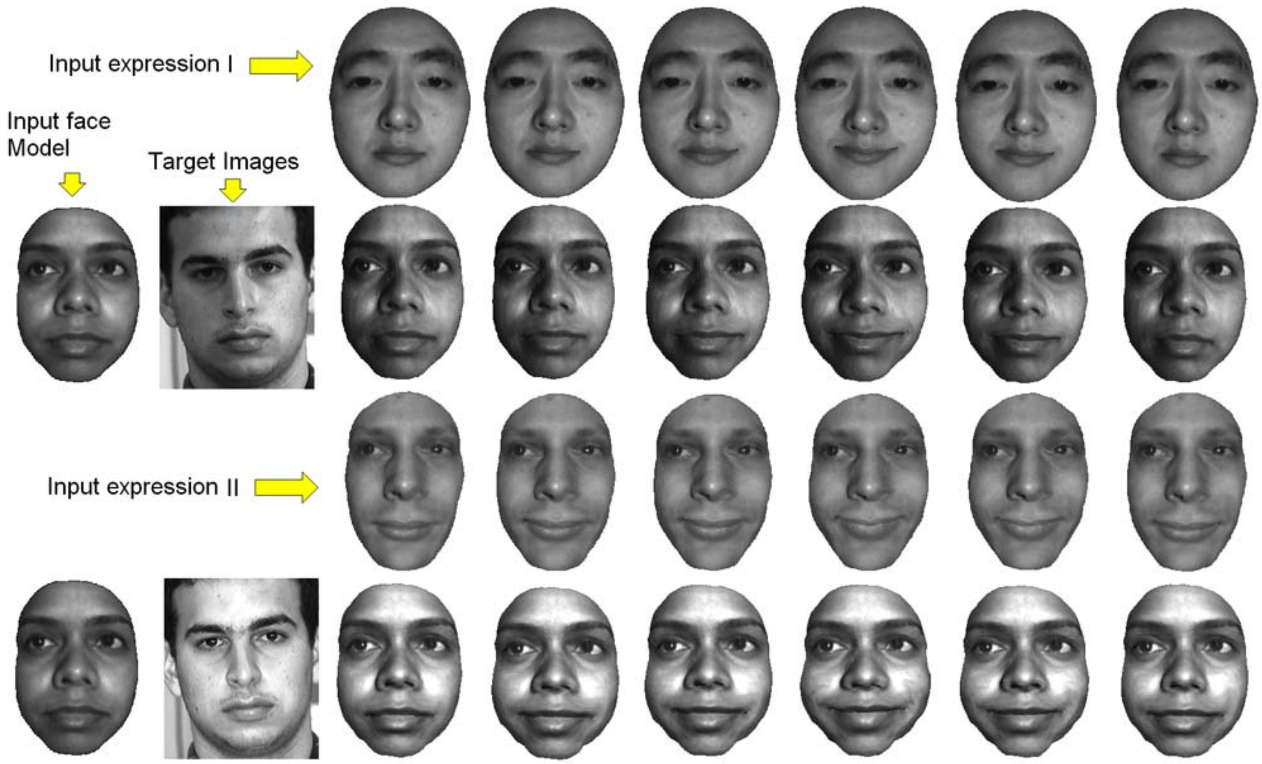

Figure 4: Expression re-targeting from a 3D face model: the first column shows the input face model and the second column shows the illumination example images I_{target}. The images in the first row (expression I) and third row (expression II) are the input expression images and the images in the second and fourth rows are the synthesized expression images of the input face model under the lighting conditions estimated from the illumination example images.

3.2.1 Method I: Morphing the Re-Lit Texture

As described in Sec. 2.2, given an input image of a face I_s and a illumination example image of a second face I_t, the SHBMM parameters $\{\alpha_s, \beta_s, \ell_s, M_s\}$ and $\{\alpha_t, \beta_t, \ell_t, M_t\}$ of the two images can be estimated respectively by solving the minimization problems in Eq. 4 and 5. Thus, according to Sec. 2.3, the re-lit texture I_r of the input image I_s can be computed by using Eq. 11. A fitting process is then performed to align the recovered face model to the multi-resolution deformable face model described in 3.1. Hence, a novel facial expression sequence of the input image I_s under illumination condition ℓ_t can be generated by morphing the re-lit texture with the extracted motion field.

3.2.2 Method II: Morphing and Relighting the De-Lit Images

Similar to Method I, we first estimate the SHBMM parameters of the input image I_s and illumination example images I_t. Then we compute the de-lit texture I_d of the input face according to Eq. 8. Since the structure of the input face is estimated and the motion field has been extraced in 3D, a new 3D shape and hence a set of spherical harmonic bases b_i of each image frame i can be computed according to Eq. 1. Using the computed spherical harmonic basis, a novel expression sequence of I_s under illumination condition ℓ_t can be generated by relighting the de-lit texture, where the appearance I_r^i of the frame i can be computed as $I_r^i = b_i \ell_t$.

The difference of the above two methods is that, in Method I, we apply the projected motion field directly to the re-lit image while in Method II, we first apply the motion field to the de-lit image and perform relighting to generate novel appearance. Method I is faster but may be inaccurate since the texture is fixed along all the frames thus cannot represents the illumination effects due to the 3D structure deformation. The results of both methods are reported and compared in Section 4.

4 Experiments and Results

We used the CMU-PIE data set [36] which provides images of illumination variations for facial expression re-targeting. The CMU-PIE database contains 68 individuals, none of which is also in the USF set used to compute the Spherical Harmonic Basis Morphable Model.

The first set of experiments is to perform expression re-targeting under novel illumination conditions given 3D face models. The results of the first expression re-targeting experiment are shown in Figure 4, where the first column shows the input face model, the second column shows the illumination example images I_{target}. The images in the first row (expression I) and third row (expression II) are the input expression images and the images in the second and fourth rows are the synthesized expression images of the input face model under the lighting conditions estimated from the illumination example images. Figure 5 and 6 show the results of expression re-targeting by using the re-lit and de-lit images respectively. These are harder experiments in the sense that the 3D face shape is not as accurate as in the experiments of Figure 4 since it's recovered through the morphable model and not directly captured. In both images, the first column shows the input image I_{input} and the second column shows the illumination example images I_{target}. The images in the first two rows are the synthesized expression images simulating expression II and the images in the last rows are the synthesized images simulating expression I. As described in Sec. 3.2.2, in the method of using re-lit images, the illumination effects

Figure 5: Facial expression re-targeting using **Re-lit** images: the first column shows the input image and the second column shows the illumination example images from which we want to transfer illumination parameters. The images in the first two rows are synthesized expressions simulating input expression II while those in the last row are synthesized according to input expression I.

in the re-lit texture are fixed for all the frames thus might not be able to represent the illumination effects due to the face structure variations. Fig. 7 compares the synthesized images using re-lit and de-lit techniques where the images in the top row are synthesized using re-lit images and those in the second row are synthesized using de-lit images. To clearly visualize the illumination effect we assumed the surface has constant albedo, hence all changes in the texture image are due to illumination. We can see from Fig. 7 that, the technique of relighting de-lit images generates images that are modulated correctly by illumination and are hence more photorealistic. In *www.cs.sunysb.edu/ lzhang/cgi*$_v$*ideo.avi*, the result I sequence shows the expression re-targeting of an input face model (Fig. 3), result II and III show the expression re-targeting of an input image (Fig. 4) under different illumination conditions, and result IV shows the re-targeted facial expression of a second input expression.

5 Conclusions and Future Work

We have shown that, given an input image of a face, a set of illumination example images (of other faces captured under different illumination) and a facial expression motion sequence, we can generate new expression sequences of the input face under the lighting conditions in the illumination example images. We demonstrated that with the two proposed methods, a 3D spherical harmonic morphable model (SHBMM) can generate images under new lighting conditions with remarkable quality even if only one single image under unknown lighting is available, not only for static poses but for dynamic sequences as well where the face is undergoing subtle high-detail motion.

In our expression re-targeting methods, the extracted motion field was applied directed to the input face. Thus, if the face of the input expression and the input face are very different in shape, the synthesized expression will not be realistic. In future work, we plan to incorporate the high level control [26][7] and editing [41]

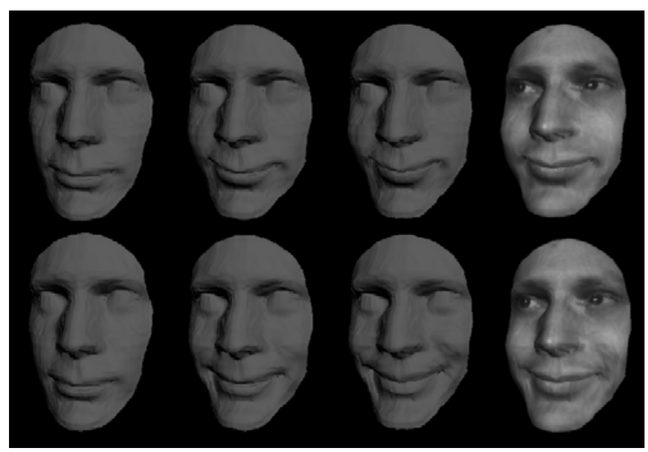

Figure 7: Comparison of Method I and II: the images in the top row are synthesized using Method I and those in the second row are synthesized using Method II. The first three columns are rendered using constant albedo, showing only the illumination effects, the fourth column has the same geometry and illumination effects as the third but true (variable) albedo). Method II generates images that are modulated correctly by illumination, as can be seen in the right column where correct illumination reveals the deformation of the cheeks .

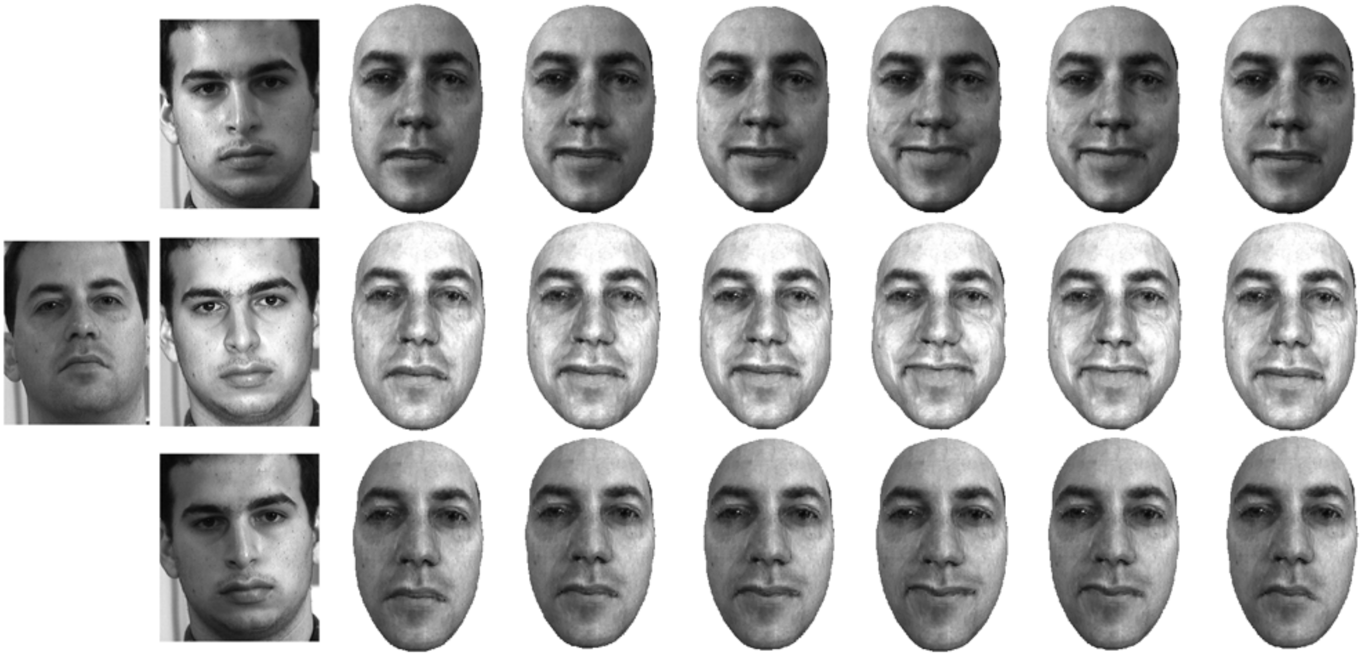

Figure 6: Facial expression re-targeting using **De-lit** images: the first column shows the input image and the second column shows the illumination example images. The images in the first two rows are synthesized expression simulating input expression II while those in the last row are synthesized according to input expression I.

mechanisms into our method to generate photo-realistic facial expressions.

References

[1] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, February 2003.

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

[3] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. In *Eurographics'03*, pages 641–650.

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

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

[6] M. Brand. Voice puppetry. In *SIGGRAPH'99*, pages 21–28.

[7] M. Byun and N. I. Badler. Facemote: qualitative parametric modifiers for facial animations. In *Symposium on Computer Animation*, pages 65–71, 2002.

[8] P.E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, pages 189–198, 1998.

[9] P.E. Debevec, T. Hawkins, C. Tchou, H Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH*, pages 145–156, 1998.

[10] D. DeCarlo, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In *SIGGRAPH'98*, pages 67–74.

[11] M. Dimitrijevic, S. Ilic, and P Fua. Accurate face models from uncalibrated and ill-lit video sequences. In *CVPR*, 2004.

[12] T. Ezzat, G. Geiger, and T. Poggio. Trainable videorealistic speech animation. In *SIGGRAPH'02*, pages 388–398.

[13] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting. In *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, pages 47–54, 1999.

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

[15] S. K. Goldenstein, C. Vogler, and D. Metaxas. Statistical cue integration in dag deformable models. *PAMI*, 25(7):801–813, 2003.

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

[17] P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang. Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Opt. Eng.*, 38(6):1065–1071.

[18] X. Huang, N. Paragios, and D. Metaxas. Establishing local correspondences towards compact representations of anatomical structures. In *MICCAI'03*, pages 926–934, 2003.

[19] P. Joshi, W. C. Tien, M. Desbrun, and F. Pighin. Learning controls for blend shape based realistic facial animation. In *Symposium on Computer Animation*, pages 187–192, 2003.

[20] G. A. Kalberer and L. Van Gool. Realistic face animation for speech. *Intl. Journal of Visualization and Computer Animation*, 2002.

[21] P. Kalra, A. Mangili, N. Magnenat-Thalmann, and D. Thalmann. Simulation of facial muscle actions based on rational free form deformations. In *Eurographics'92*, pages 59–69.

[22] J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Silhouette-based 3d face shape recovery. In *Graphics Interface*, 2003.

[23] K.C. Lee, J. Ho, and D.J. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. In *CVPR01*, pages I:519–526, 2001.

[24] Y. Lee, D. Terzopoulos, and K. Walters. Realistic modeling for facial animation. In *SIGGRAPH'95*, pages 55–62.

[25] H. Huitric M. Nahas and M. Saintourens. Animation of a b-spline figure. In *The Visual Computer*, pages 272–276, 1988.

[26] J.-Y. Noh and U. Neumann. Expression cloning. In *SIGGRAPH'01*, pages 277–288.

[27] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[28] F. I. Parke and K. Waters. *Computer facial animation*. 1996.

[29] A.P. Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *PAMI*, 22(1):107–119, January 2000.

[30] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin. Synthesizing realistic facial expressions from photographs. In *SIGGRAPH'98*, pages 75–84.

[31] R. Ramamoorthi. Analytic pca construction for theoretical analysis of lighting variability in images of a lambertian object. *PAMI*, 24(10), Oct. 2002.

[32] R. Ramamoorthi and P Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH*, pages 497–500, 2001.

[33] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 8:712–721, 1999.

[34] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *SIGGRAPH'86*, pages 151–160.

[35] A. Shashua. On photometric issues in 3d visual recognition from a single 2d image. *IJCV*, pages 21:99–122, 1999.

[36] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *PAMI*, pages 1615–1618, 2003.

[37] Riklin-Raviv T. and Shashua A. The quatient image: Class based rerendering and recognition with varying illuminations. In *CVPR*, pages 566–571, 1999.

[38] Y. Wang, Z. Huang, C.S. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgamma, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3-d facial expressions. In *EuroGraphics*, volume 23, 2004.

[39] Z. Wen, Z. Liu, and T. S. Huang. Face relighting with radiance environment maps. In *CVPR*, 2003.

[40] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *CVPR*, volume I, pages 19–25, 2003.

[41] L. Zhang, N. Snavely, B. Curless, and S.M. Seitz. Spacetime faces: High-resolution capture for modeling and animation. In *ACM SIGGRAPH*, 2004.

[42] L. Zhang, S. Wang, and D. Samaras. Title withheld to preserve anonymous of cvpr05 submission.

[43] Q. Zhang, Z. Liu, B. Guo, and H. Shum. Geometry-driven photorealistic facial expression synthesis. In *Symposium on Computer Animation*, pages 177–186, 2003.

[44] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: A literature survey. In *UMD*, 2000.