Eurographics Symposium on Rendering (2005)
Kavita Bala, Philip Dutré (Editors)

Estimation of 3D Faces and Illumination from Single Photographs Using A Bilinear Illumination Model

Jinho Lee[†‡] Raghu Machiraju[‡] Hanspeter Pfister[†] Baback Moghaddam[†]

[†] Mitsubishi Electric Research Laboratories
Cambridge, MA, USA

[‡] The Ohio State University
Columbus, OH, USA

Abstract

3D Face modeling is still one of the biggest challenges in computer graphics. In this paper we present a novel framework that acquires the 3D shape, texture, pose and illumination of a face from a single photograph. Additionally, we show how we can recreate a face under varying illumination conditions. Or, essentially relight it. Using a custom-built face scanning system, we have collected 3D face scans and light reflection images of a large and diverse group of human subjects . We derive a morphable face model for 3D face shapes and accompanying textures by transforming the data into a linear vector sub-space. The acquired images of faces under variable illumination are then used to derive a bilinear illumination model that spans 3D face shape and illumination variations. Using both models we, in turn, propose a novel fitting framework that estimates the parameters of the morphable model given a single photograph. Our framework can deal with complex face reflectance and lighting environments in an efficient and robust manner. In the results section of our paper, we compare our methods to existing ones and demonstrate its efficacy in reconstructing 3D face models when provided with a single photograph. We also provide several examples of facial relighting (on 2D images) by performing adequate decomposition of the estimated illumination using our framework.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

Modeling faces from single photographs has many applications in computer graphics (computer games, virtual reality, expression synthesis, face replacement, reconstruction of skin reflectance) and in computer vision (pose and illumination invariant face recognition, face tracking in videos). It also has many potential applications in human computer interaction and digital photography. To date, a morphable face model [BV99] is the most effective way to reconstruct 3D face shape, texture, pose, and illumination from photographs. A morphable model compactly encapsulates the space of human faces based on the statistics of measured face data from a group of human subjects.

The main challenge in the application of a morphable model for face reconstruction is the formulation of an efficient, robust, and general fitting procedure. In an approach often called analysis-by-synthesis [BV99], the model and rendering parameters are optimized such that the rendering of the model yields the closest match to the given input image under the assumption of a suitable image-error metric. This task is complicated by the complex reflectance properties of human faces and the unknown types and numbers of light sources in the input photographs. Previous work makes several simplifying assumptions, such as Lambertian reflectance for faces [AGR96], simple analytic illumination models (usually Phong) and simple illumination (usually one ambient and one directional light) [BBPV03, BSVS04]. Unfortunately, these simplifications make it hard to achieve photorealistic and geometrically accurate face models.

Inspired by recent research in linear illumination subspaces [BJ03, RH01], we introduce a novel framework to fit a morphable model to single photographs. Instead of making arbitrary assumptions about skin reflectance and the illumination environment we use a bilinear illumination model

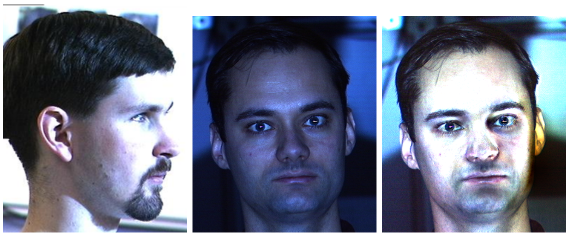


Figure 1: *An example of transferring an illumination of a face in one image to another face in a different image. (left) Source or reference image. (middle) Target image. (right) Resulting image. Image sources are from [SBB03]*

for 3D faces that has been computed from measured data of faces under varying illumination. During the reconstruction of a query face, we first estimate the 3D shape, texture, and pose parameters (initially specified by a user) of the morphable model. Instead of comparing the rendered face to the input image, we measure the distance of the face in the input image to a projection in the illumination subspace that is specific to the estimated 3D shape. Using this simplified cost function, we reconstruct the 3D shape, texture, pose, as well as the illumination parameters of the bilinear model. We can then apply the reconstructed illumination to another reconstructed 3D face for illumination transfer in digital photography (see Figure 1). In addition, the reconstructed illumination can be further decomposed to estimate the colors and intensities of the individual light sources that constitute the lighting environment of the input image. This feature of our illumination model is exploited to solve a more general problem of *face relighting*.

The main contributions of our paper are: (1) a novel optimization framework based on the illumination subspace method for the fitting of morphable face models; (2) a method to construct a measurement-based illumination model to deal with arbitrary lighting conditions and complex reflectance properties of human faces; (3) a novel method to relight a face in a photograph in an intuitive and flexible manner using the proposed illumination model.

In Section 2 we describe relevant previous work. In Section 3 we describe our system for capturing 3D faces and reflectance images. In Section 4 we propose our novel optimization framework to estimate 3D face and illumination from single 2D images. Section 5 describes how we construct a bilinear illumination model from real measurements of many subjects. Section 6 shows some results of single image reconstruction, illumination transfer, and face relighting.

2. Related Work

In this section, we describe related work in morphable models, illumination subspaces and face relighting.

To fit a morphable model to single images, Blanz and Vetter [BV99] introduce a cost function based on the difference between the input image and the rendered image. They use a stochastic gradient decent algorithm to minimize that function. In subsequent work with different applications [BV03, BBPV03, BSVS04], they extend this cost function to include the distances between the 2D feature points in the image and the projection of the corresponding 3D feature points. This extension constraints further the search space of 3D shape and pose during optimization. Related work that employed morphable models was reported using sparse feature points [BMVS04] and a set of silhouette images taken from different viewpoints [LMPM03]. Our strategy is different from these approaches in that we do not use any rendering parameters explicitly during optimization. *It is purely a data-driven approach*.

Previous work in linear illumination subspaces includes Georghiades et al. [GBK01] who use a photometric stereo approach to reconstruct 3D face geometry and albedo to generate synthetic images of a subject's face. In [HHB03] a similar technique is used to generate synthetic 2D images from a morphable model to train a component-based classifier for face recognition. Basri and Jacobs [BJ03] showed that an arbitrary illumination of a convex Lambertian object can be approximated accurately by a low dimensional linear subspace spanned by *nine harmonic images*. These nine harmonic images can be generated analytically given surface normals and albedo of the object. Zhang and Samaras [ZS04] combined nine spherical harmonics and morphable model approach to estimate the illumination bases from a single photograph and applied their method to pose invariant face recognition.

Similar to our work, Vasilescu and Terzopoulos [VT03] perform a multilinear analysis to model different factors in forming facial images explicitly. They perform a higher-order SVD on the tensor data of 2D images to compute a space that spans identity, expression, pose, and illumination. A key distinction of our work is that we explore a bilinear illumination subspace of human faces using high-resolution 3D geometry, not 2D images, and combine it with the morphable model of Blanz and Vetter [BV99]. Vasilescu and Terzopoulos [VT04] showed that multilinear analysis is effective for representing a texture space which incorporates various viewpoints and illumination conditions.

Though a number of techniques are reported for the problem of face relighting, we only describe a recent paper that employs a similar approach. Wen et al. [WLH03] used radiance environment maps to relight faces in photographs. After computing an approximated radiance environment map using spherical harmonics from a single photograph, the estimated lighting condition can be applied from different orientations to relight the input face. To relight a face from arbitrary lighting conditions, they modify the estimated nine harmonic coefficients interactively. However, it is often difficult to relate the harmonic coefficients directly to the numbers, colors and intensities of the individual light sources to obtain the desired illumination effect. We present a method for face relighting under arbitrary lighting conditions, allowing users to explicitly control individual light sources.

3. Data Acquisition and Registration

We first describe our custom-built system for capturing 3D faces and reflectance images. Later, we describe the methods to register the acquired illumination samples into a common vector space.

Face scanning dome To create a repository of models, we acquire high-resolution 3D geometry of faces using a commercial system (from 3QTech) that employs structured-light to scan the face. The acquired 3D mesh consists of more than 40,000 vertices. We also use a custom-built, dome-structured device to acquire reflectance images of the face, which is equipped with 16 digital cameras and 146 directional LED light sources. Each light source consists of 103 white LEDs and a diffuser. The 16 cameras are controlled by eight client PCs and are synchronized with the light sources. We obtain a photograph of the face from each camera with each light source turned on in a sequential fashion. This results in 2,336 images of the face illuminated by 146 light sources from 16 different viewpoints.

We use the freely available OpenCV calibration package to determine spherical and tangential lens distortion parameters. External camera calibration is performed by synchronously acquiring a few hundred images of an LED swept through the dome center. Nonlinear optimization is used to optimize for the remaining camera intrinsics and extrinsics.

Registration of illumination samples All 3D face geometries acquired through our system are rigidly aligned in a common coordinate system. Then we select 40 feature points in the facial area of each 3D face and compute point-to-point correspondence. First, we choose a reference face and improve the geometry so that it has the desired number of points in the facial region. For each target face, we warp the reference face so that it matches the target face in terms of 40 feature points using scattered data interpolation. Later, we perform a cylindrical resampling from the warped reference face to the target face to obtain a corresponding point in the target face for each point in the warped reference face. More details can be found in [LMPM03].

To register the 2D illumination samples of the acquired face images to the 3D geometry of the subject face we find the similarity transformation between the coordinate systems of the dome cameras and the 3D scanning system. To determine this transformation we use a 3D calibration target with nine markers and create images of the target using all 16 cameras. Using the intrinsic and extrinsic calibration data and the acquired 2D correspondence of those markers, we can obtain the 3D positions of those markers in the dome coordinate system using non-linear optimization. The objective function employed here is:

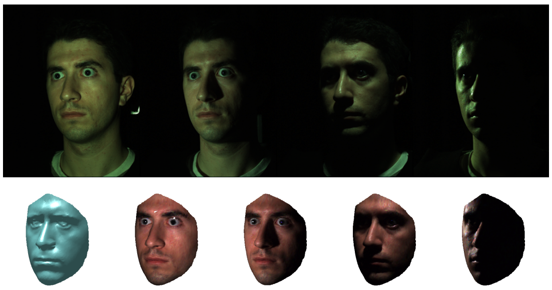


Figure 2: *(Top) Raw reflectance images. (Bottom) A 3D shape and the registered illumination samples for the corresponding images.*

$$\mathbf{y}_i = \arg\min \sum_{k=1}^{K} \|\mathbf{x}_{i,k} - P_k(\mathbf{y}_i)\|^2 \qquad (1)$$

where K is the number of cameras, P_k is a projection matrix of camera k, $\mathbf{x}_{i,k}$ is the 2D location of feature point i observed by camera k, and $\mathbf{y}_i$ is the i^{th} 3D feature point.

After obtaining the corresponding feature point $\mathbf{z}_i$ from the 3D geometry of the target, we compute the similarity transformation Q from the coordinate space of the geometry acquisition subsystem to the dome coordinate system. A Procrustes analysis is performed between the point sets $\mathbf{y}$ and $\mathbf{z}$. Using the acquired similarity transform and extrinsic calibration data of each camera, we know the 2D-3D correspondence between the points on the 3D face and the pixels on the 2D images from all 16 cameras. The final mapping is computed by:

$$\mathbf{x}_{i,k} = P_k(Q(\mathbf{z}_i)). \qquad (2)$$

We apply this transformation to all surface points of the face geometry and obtain the corresponding illumination samples from the images captured by all 16 cameras and 146 lighting conditions. Figure 2 shows some raw reflectance images and registered illumination samples on a common 3D shape space.

Occlusion filling and albedo estimation For the non-observable points from a certain camera viewpoint, we resolve or locate the *holes* using the illumination samples observed from other camera viewpoints. Given the positions of the holes in camera viewpoint V_i, we choose a camera viewpoint V_j that is near V_i but has no holes at those positions. Then, we perform principal component analysis (PCA) for all illumination images obtained at view V_j. To approximate the illumination of hole points in V_i, we project only the *observed* points onto the subspace spanned by the first M eigenvectors of V_j and reconstruct the closest illumination in the illumination subspace of V_j. This reconstruction includes valid illumination values in the holes w.r.t viewpoint V_i.

Finally, to estimate the diffuse texture (albedo) from all reflectance images, we first compute the average of 146 illumination samples per each visible vertex for each viewpoint. Then, the samples from the 16 viewpoints at the specific vertex are blended by a weighted average using the cosine of the angles between the vertex normal and the view vector as weights.

4. Estimation of 3D Faces Using Illumination Subspace

In this section we describe the morphable model and how we construct it from a mixture of two different 3D face databases, and present an optimization framework using a dynamically generated illumination subspace combined with the morphable model.

4.1. Morphable Model

To derive a morphable model for 3D shape and texture we combine data from the USF Human ID database [usf] (134 subjects) and our own database (71 subjects). We first compute the point-to-point correspondence across all scans of the two databases so that they are all in the common vector space using the method described in Section 3. After constructing a vector $\mathbf{s} = (x_1 \cdots x_N, y_1 \cdots y_N, z_1 \cdots z_N)$ for each shape and $\mathbf{t} = (r_1 \cdots r_N, g_1 \cdots g_N, b_1 \cdots b_N)$ for each texture, we perform PCA on all shape vectors $\mathbf{S}$ and texture vectors $\mathbf{T}$ separately. Using the first M eigenvectors and model parameters α and β, an arbitrary shape and texture can be reconstructed as following [BV99]:

$$\mathbf{s} = \bar{\mathbf{S}} + \sum_{i=1}^{M} \alpha_i \mathbf{e}_i^s, \quad \mathbf{t} = \bar{\mathbf{T}} + \sum_{i=1}^{M} \beta_i \mathbf{e}_i^t, \tag{3}$$

where $\bar{\mathbf{S}}$ and $\bar{\mathbf{T}}$ are the average shape and texture across subjects, $\mathbf{e}_i^s$ and $\mathbf{e}_i^t$ are the i^{th} eigenvector for shape and texture respectively.

4.2. Fitting Procedure

Given a photograph of an unknown person, we generate arbitrary shape (α) and texture (β) coefficients based on the morphable model. We also estimate an arbitrary pose (rotation and translation) of the face (γ) (using 9 initial user-specified feature points).

Given α and γ (and assuming fixed camera parameters), we project the geometry of the synthesized face to the image plane of the photograph. Using a simple visibility test we then acquire the corresponding pixel values of each projected visible point on to the surface of the face. If this projection is properly aligned to the input face, the acquired pixel values comprise the illuminated texture ($\hat{\mathbf{t}}$) of the face with the specific shape (α) and pose (γ) parameters. *Now, it is useful to ask where does one get an illuminated texture.* This is the major departure point from existing methods.

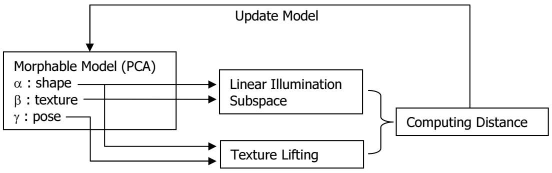


Figure 3: *An optimization framework using a dynamically generated illumination subspace.*

We assume that there exists an illumination subspace, represented by a matrix $\mathbf{B}$ whose columns constitute basis vectors that span a subspace of all possible illumination of the shape $\mathbf{s}$ and texture $\mathbf{t}$ of a face. We explain how we construct this illumination space in Section 5. Here we describe the overall fitting process.

From Eq.3 $\mathbf{s}$ and $\mathbf{t}$ can be reconstructed by α and β. Therefore, using the illumination subspace $\mathbf{B}$ for a given α and β, the distance of the *lifted texture* ($\hat{\mathbf{t}}$) to $\mathbf{B}$ can be used as a cost function to find the optimal α, β, and γ for the given photograph. By lifted texture, we mean a vector of pixel values of the input photograph extracted by the projection of the 3D points of a face with the given shape and pose all of which have passed the visibility test for self-occlusion. By updating the model parameters iteratively based on this distance function, we find the estimate of shape, texture, pose and illumination of the face in the image. Figure 3 describes this procedure in the form of a flow diagram.

Based on this framework, we formulate a new cost function for fitting the morphable model to the input image:

$$f(\alpha,\beta,\gamma) = \sum_{c=r,g,b} \|\hat{\mathbf{t}}_c - \mathbf{B}_c \mathbf{B}_c^T \hat{\mathbf{t}}_c\|, \tag{4}$$

where $\hat{\mathbf{t}}_c$ is an $N \times 1$ vector obtained by extracting pixel values (for each channel c separately) from the input image using the *geometric* projection of the 3D model template computed by the current α and γ. $\mathbf{B}_c$ is an $N \times M$ matrix which contains M orthonormal bases for the texture-weighted illumination subspace. $\mathbf{B}_c \mathbf{B}_c^T \hat{\mathbf{t}}_c$ is considered a projection of the lifted texture $\hat{\mathbf{t}}_c$ to the illumination subspace spanned by the column vectors of $\mathbf{B}_c$. A similar distance metric is reported in [BJ03] for the purpose of illumination invariant face recognition.

The orthonormal bases $\mathbf{B}_c$ of the texture-weighted illumination subspace are computed with the input of current model parameters α, β. The following procedure explains how we obtain $\mathbf{B}_c$ using nine harmonic images:

1. Given α, β, compute the geometry and diffuse texture of the face using the computed morphable model (Section 4.1): $\mathbf{s}$ and $\mathbf{t}$ (See Eq 3).

2. Compute vertex normals **vn** from **s** and the mesh connectivity of the morphable model.
3. Update the first nine harmonic reflectance vectors **R** as described in [BJ03] from **vn**.
4. Build the nine harmonic images $\hat{\mathbf{B}}_c$ by element-wise multiplication of $\mathbf{t}_c$ with each column vector of **R**.
5. Perform QR decomposition to obtain the orthonormal bases $\mathbf{B}_c$ from $\hat{\mathbf{B}}_c$.

We use the downhill simplex method [PFTV88] to optimize the cost function (Eq 4). It is a non-linear minimization algorithm that requires only cost function evaluations. Although the simplex method is not very efficient in the number of function evaluations until convergence [PFTV88], it works robustly with our problem. It provides a tangible way to deal with the relatively large range of initialization settings by adjusting the initial size of simplex. It should be noted that a gradient-based optimization method such as Levenberg-Marquardt algorithm can be used. Since the analytic derivatives are not known, we will have to rely on a numerical procedure to compute the derivatives.

Note that the illumination subspace spanned by nine harmonic images is one example that can be used together with our framework. Although this analytical linear subspace is fast to compute on the fly, it has an inherent limitation on handling non-Lambertian objects such as human faces. In the following section we present a measurement based illumination model that works together with this framework.

5. Bilinear Illumination Model

The problem we solve can be formally described as follows: "Given a statistical model for shape and texture, what is the most appropriate illumination subspace for the given shape and texture parameters according to the real illumination measurements of the samples in the training dataset?" For this purpose, we use a bilinear illumination model based on the higher-order SVD or N-mode SVD [LMV00, VT03]. Note that we do not use any Lambertian assumptions when constructing our illumination model. Since our model database (See Section 3) is built from real photographs of human faces, self-shadowing and specularity is implicitly included in our model.

We start by decoupling the underlying shape and texture in our acquired face data. One should observe that a pure diffuse texture part (albedo) can be separated from facial illumination maps. Much of reflectance and illumination including shadows is dependent on the shape (geometry) of a face assuming similar reflectance properties across different faces and different parts of a face. By factoring out the diffuse texture, we are now able to capture the subtleties that arise from specular reflectance and shadowing effects that come solely from the form of the facial surface. Assuming that facial texture can be decorrelated with the shape and reflectance, we factor out the diffuse texture (albedo) from the illumination samples in the following manner:

$$w_n = \hat{t}_n / t_n, \quad n = 1..N, \tag{5}$$

where $\hat{t}_n$ is an illumination sample, t_n is the diffuse texture at a 3D point $\mathbf{p}_n$ with N being the number of 3D mesh points. We call w_n a *texture-free illumination component*, which is different from reflectance since it also includes cast shadows. For the subsequent data analysis we use this shape-dependent, texture-free illumination component. Similar to the nine harmonic images [BJ03], the diffuse texture is multiplied with the reflectance bases to build a *texture-weighted* illumination subspace.

We now build a bilinear illumination model using the 3D shape and lighting conditions as the axes of variation. For each pair of shape i and lighting condition j, we have $N(= 10,006)$ 3D positions and texture-free illumination components ($[xyzw]$ tuples) for 33 subjects. We assemble them as a long vector $\mathbf{a}_{i,j} = (x_1 \cdots x_N, y_1 \cdots y_N, z_1 \cdots z_N, w_1 \cdots w_N)$ with length $4N$.

We choose one near-frontal viewpoint with the occlusion filling procedure described in Section 3 for further analysis. The size of our data tensor $\mathcal{D}$ is $33 \times 146 \times 4N$. $\mathcal{D}$ can be expressed as follows:

$$\mathcal{D} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \tag{6}$$

where $\times_n$ is a *mode-n product* defined between a tensor $\mathcal{A} \in \Re^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ with order $N \geq 2$ and a matrix $\mathbf{U} \in \Re^{J_n \times I_n}$. It is an operation that replaces every column vector $\mathbf{a}_{i_1 \cdots i_{n-1} i_{n+1} \cdots i_N} \in \Re^{I_n}$ in $\mathcal{A}$ with the column vector obtained by $\mathbf{Ua}$. The result is a tensor with dimension $I_1 \times \cdots \times J_n \times \cdots \times I_N$.

A *core tensor* $\mathcal{C} \in \Re^{33 \times 146 \times 4N}$ governs the interaction between mode matrices $\mathbf{U}_k, k = 1..3$. Note that, unlike the sigular value matrix in a traditional matrix SVD, C does not have diagonal structure but usually is a full matrix. The *mode matrices* $\mathbf{U}_k$ can be computed by performing SVD on a matrix $\mathbf{D}_{(k)} \in \Re^{I_k \times (I_1 \cdots I_{k-1} I_{k+1} \cdots I_3)}$, which is composed of all column vectors $\mathbf{d}_{i_1 \cdots i_{k-1} i_{k+1} \cdots i_3} \in \Re^{I_k}$ in $\mathcal{D}$, where $I_1 = 33, I_2 = 146, I_3 = 4N$ of our tensor data $\mathcal{D}$. $\mathbf{U}_k$ constitutes orthonormal bases of the column space of $\mathbf{D}_{(k)}$.

Using the associative property of the *mode-n* product [LMV00], the last mode matrix can be incorporated in $\mathcal{Z} = \mathcal{C} \times_3 \mathbf{U}_3$, resulting in a simplified equation:

$$\mathcal{D} = \mathcal{Z} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2, \tag{7}$$

where $\mathbf{U}_1 \in \Re^{33 \times 33}$ and $\mathbf{U}_2 \in \Re^{146 \times 146}$ capture the variation along the shape and lighting axes, respectively. $\mathcal{Z} \in \Re^{33 \times 146 \times 4N}$ governs the interaction between $\mathbf{U}_1$ and $\mathbf{U}_2$. It can be computed using the orthonomality of $\mathbf{U}_k$ with:

$$\mathcal{Z} = \mathcal{D} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T. \tag{8}$$

The result is a bilinear model that captures the variation of 3D shape and texture-free illumination. This model provides us with 146 illumination bases given the coefficient vector

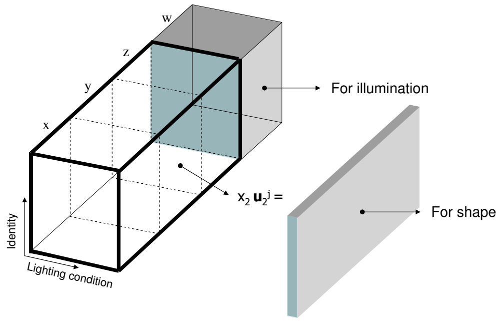


Figure 4: *Visualization of dividing a core tensor into two parts and generating a more compact model exploiting the redundancy of geometry data along the illumination axis*

of a person's face geometry. The number of bases can be reduced to a lower dimensional space to handle larger numbers of training subjects. Similar to traditional matrix SVD, this compression can be done easily by retaining only the first M_k columns of $\mathbf{U}_k$. However unlike the matrix SVD, this truncation generally does not yield an optimal solution in the least square sense though it provides a good approximation [LMV00].

We reduce the dimension of shape and lighting condition axis from 33 to 20 and 146 to 30, respectively using the algorithm described in [VT03]. This yields $\tilde{\mathbf{U}}_1 \in \Re^{33\times 20}$, $\tilde{\mathbf{U}}_2 \in \Re^{146\times 30}$, and a core tensor $\tilde{\mathcal{Z}}$ with dimensions of $20 \times 30 \times 4N$. The approximation of $\mathcal{D}$ is obtained using:

$$\tilde{\mathcal{D}} = \tilde{\mathcal{Z}} \times_1 \tilde{\mathbf{U}}_1 \times_2 \tilde{\mathbf{U}}_2. \tag{9}$$

Due to the redundancy of the geometry data ($[xyz]$ tuples) along with the lighting conditions of the original data tensor $\mathcal{D}$, we keep only part of the core tensor $\tilde{\mathcal{Z}}$ without loss of information. We can divide $\tilde{\mathcal{Z}}$ into two parts: the geometry matrix $\tilde{\mathbf{Z}}_s \in \Re^{20\times 3N}$ and the illumination tensor $\tilde{\mathcal{Z}}_l \in \Re^{20\times 30\times N}$. $\tilde{\mathbf{Z}}_s$ can be computed using:

$$\tilde{\mathbf{Z}}_s = \tilde{\mathcal{Z}} \times_2 \tilde{\mathbf{u}}_2^j, \tag{10}$$

where $\tilde{\mathbf{u}}_2^j$ can be any row vector of $\tilde{\mathbf{U}}_2$ and $\tilde{\mathcal{Z}}_l$ is obtained by keeping only the last N slices along the third dimension of $\tilde{\mathcal{Z}}$. Figure 4 visualizes this procedure.

In this formulation of data compression, a reconstruction of geometry and illumination bases of subject i can be computed by:

$$\mathbf{s}_i = \tilde{\mathbf{u}}_1^i \tilde{\mathbf{Z}}_s; \tag{11}$$

$$\mathbf{R}_i = \tilde{\mathcal{Z}}_l \times_1 \tilde{\mathbf{u}}_1^i, \tag{12}$$

where $\tilde{\mathbf{u}}_1^i$ represents the i^{th} row vector of $\tilde{\mathbf{U}}_1$. If we replace $\tilde{\mathbf{u}}_1^i$ with a linear combination of the row vectors of $\tilde{\mathbf{U}}_1$, then the above equations will generate a geometry and illumination bases for the linearly combined face. By spanning the reconstructed illumination bases we can represent any linear combination of all 146 light sources to a reasonable level of accuracy.

To use this bilinear model together with the fitting framework we described in Section 4.2, we relate the model parameters from the external morphable model (see Section 4.1) to a shape space in the bilinear model and compute the person-specific illumination subspace through the following procedure:

1. Given α, β, compute the geometry and diffuse texture of the face using the morphable model: $\mathbf{s}$ and $\mathbf{t}$ (See Eq 3).
2. Solve an overdetermined linear system $\mathbf{s} = \tilde{\mathbf{Z}}_s^T \hat{\alpha}$ with respect to $\hat{\alpha}$ (See Eq 11).
3. Obtain the illumination bases $\mathbf{R}$ by replacing $\tilde{\mathbf{u}}_1^i$ with $\hat{\alpha}$ in Eq 12.
4. Build a texture-weighted illumination bases $\hat{\mathbf{B}}_c$ by element-wise multiplication of $\mathbf{t}_c$ with each column vector of $\mathbf{R}$.
5. Perform QR decomposition to obtain the orthonormal bases $\mathbf{B}_c$ from $\hat{\mathbf{B}}_c$.

Note that in *step 3*, the linear system can be efficiently solved by performing QR decomposition of $\tilde{\mathbf{Z}}_s$ off-line in advance. $\mathbf{B}_c$ is used the same manner as the nine harmonic image bases we presented in Section 4 along with the cost function Eq 4. Figure 5 show the first nine texture-weighted illumination bases obtained using our method for an average shape and texture ($\alpha = \beta = \mathbf{0}$) and compares them to the analytic nine harmonic images. It is difficult to capture the high frequency components of the face illumination such as specularities and cast shadows using linear analysis with only nine dimensions. This is the reason we use more than nine bases upto thirty bases.

6. Results

In this section, we present the results of fitting the morphable model to single photographs using different methods. We also describe methods and results for face relighting using the reconstructed illumination from our illumination model

6.1. Fitting to Single Photographs

In our implementation of the proposed fitting framework, the cost function Eq 4 is computed using DGELS in LAPACK.Using the downhill simplex method as our optimization method we iterate several times starting from the previous fitted parameters and increasing the dimensionality of the model parameters in each iteration. Thus, we obtain coarse yet expedient fitting in a lower-dimensional space, while achieving a more detailed closer fitting in a higher dimensional space at a higher computational cost. A typical fitting process often requires $1-3$ minutes on a Pentium 4 2GHz PC.

Figure 6 shows the reconstructions of two illuminated 3D

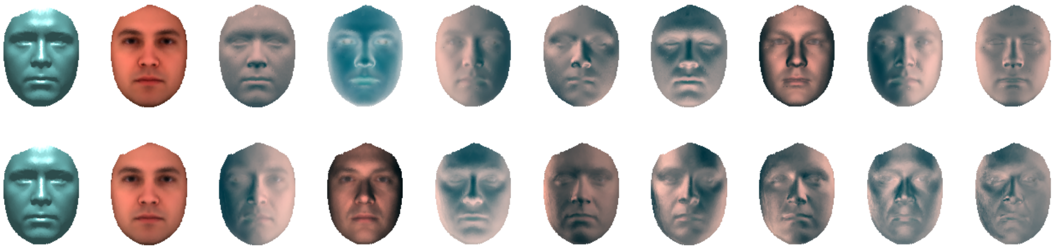


Figure 5: *The illumination bases of 9-D using nine harmonic images (upper row) and our bilinear model (bottom row) for a certain 3D face model with average shape and texture. Each base is scaled independently to cover the full color range for purposes of visualization.*

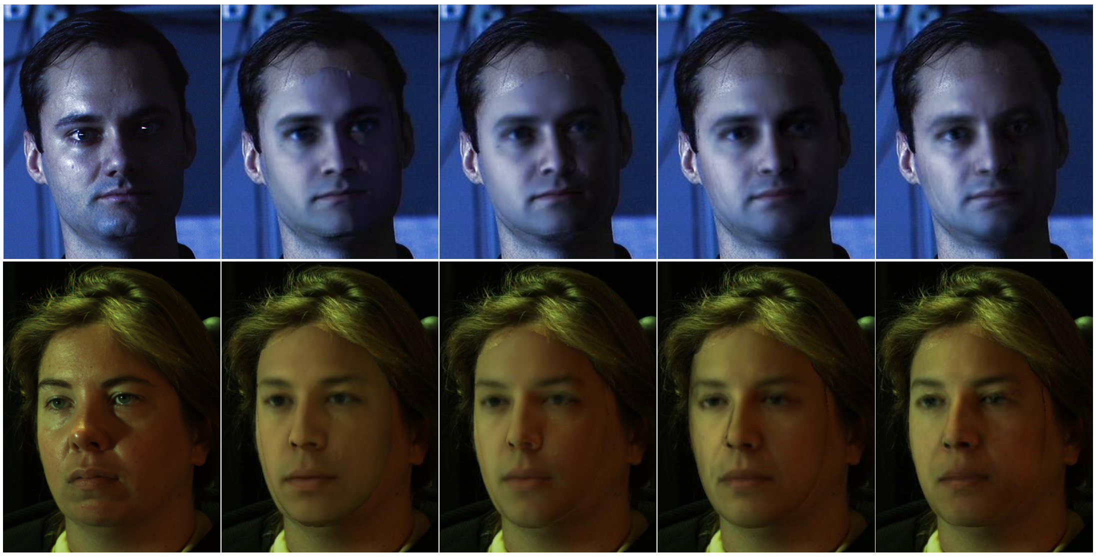


Figure 6: *Reconstruction of illuminated 3D faces with various optimization methods. Column 1: input images; Column 2: Using explicit lighting parameters; Column 3-5: Our framework with nine harmonic images (column 3), bilinear model with 9 bases (column 4), and 30 bases (column 5).*

faces using different algorithms and different illumination bases. The left column in Figure 6 shows the input images. The second column shows results using our implementation of the morphable model approach by Blanz and Vetter [BV99] with explicit illumination parameters (one ambient and one diffuse intensity for each channel) and a realization of Phong illumination model. The remaining columns show results obtained from our proposed fitting framework. The results shown in the third column use nine harmonic images, whereas the fourth and fifth columns show results using the bilinear illumination model with 9 bases and 30 basis vectors, respectively.

The input image in the first row is one of the images in the PIE database [SBB03], the subject being illuminated by a single point light source. The input image in the second row was captured by our scanning system with a mixture of two point light sources (one LED and one Halogen light) and one fluorescent area light source (the subject in the image is outside our training data samples). It should be noted that our fitting framework using the illumination subspace works robustly well under both harsh illumination and high saturation inducing illumination (as embodied in the images), as shown in Figure 7.

6.2. Face Relighting

Given a photograph of a face lit by arbitrary complex illumination, we may want to apply the same illumination on another photograph with a different face and a different lighting environment. This problem can be considered to be a special case of the more general *relighting* problem. We describe our efforts towards solving both problems.

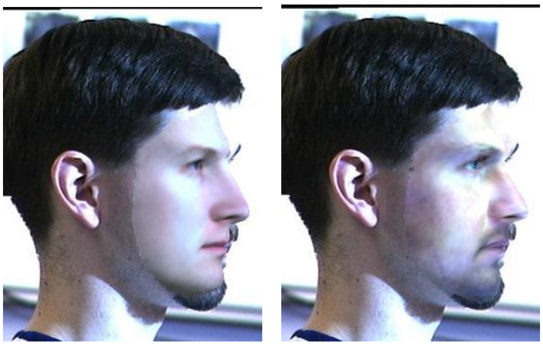


Figure 7: *For a harsh illumination environment that induces image saturation the method by Blanz and Vetter (left) is easily captured in local minima. Our proposed method (right) is not hampered by this situation since it does not optimize for explicit lighting parameters.*

6.2.1. Illumination Transfer

We assume that we have two photographs: the source image and the target image to be re-lighted using the source image. An approximate solution for the problem of illumination transfer can be achieved by leveraging the ability of our model to reconstruct complex illumination environments. After fitting the morphable model to both source and target images, we reconstruct the diffuse textures $\mathbf{t}_s$ and $\mathbf{t}_t$ using the coefficient vectors β_s and β_t, respectively (Eq.3). The texture-free illumination of target face $\mathbf{w}_t$ is then decoupled from the reconstructed texture of target illumination $\tilde{\mathbf{t}}_t$ as follows:

$$\mathbf{w}_t = \tilde{\mathbf{t}}_t./\mathbf{t}_t. \tag{13}$$

To replicate the illumination from the source image onto the target face, we apply the source illumination parameters to the texture-weighted illumination bases of the target face using a similar procedure:

$$\mathbf{w}_s = (\mathbf{B}_t\mathbf{B}_s^T\tilde{\mathbf{t}}_s)./\mathbf{t}_s, \tag{14}$$

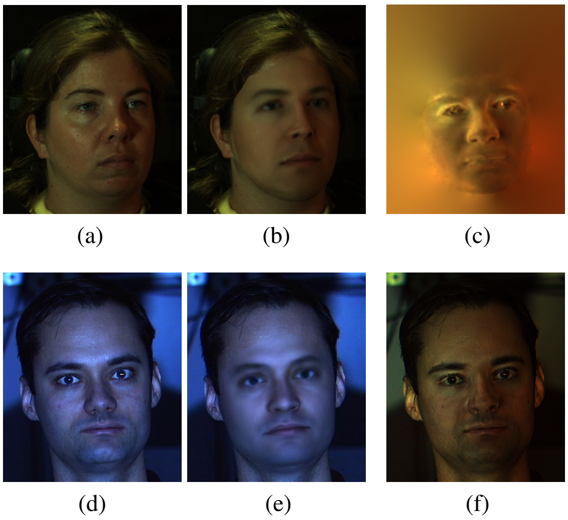
where $\mathbf{B}_s, \mathbf{B}_t$ is the source and target illumination bases, respectively, and $\tilde{\mathbf{t}}_s$ is the reconstructed texture of source illumination. Operator ./ indicates element-wise matrix division.

Figure 8 shows an example of this procedure. $\tilde{\mathbf{t}}_s$ and $\tilde{\mathbf{t}}_t$ are the pixel colors of all vertices on the fitted models shown in Figure 8(b) and Figure 8(e), respectively. Assuming $\mathbf{w}_s$ and $\mathbf{w}_t$ are close to the real face illumination, the transfer from source to target can be approximated as follows:

$$\mathbf{t}_{xfer} = \hat{\mathbf{t}}_t. * \mathbf{w}_s./\mathbf{w}_t, \tag{15}$$

where $\hat{\mathbf{t}}_t$ are the corresponding pixel values of the original target image. Since we wish to apply this formulation to all the pixels in the original source image, we first perform interpolation of the projection of $\mathbf{w}_q = \mathbf{w}_s./\mathbf{w}_t$ to fill the entire image plane, resulting $W_q \in \Re^{H\times W}$, where $H \times W$ is the image resolution. This generates the *reflection ratio image* shown in Figure 8(c). The final transfer (Figure 8(f)) is performed by multiplying each pixel of the original image (Figure 8(d)) with the corresponding pixel of the reflection ratio image as shown below:

$$\text{Image}_{xfer} = \text{Image}_{target}. * W_q. \tag{16}$$



(a) (b) (c)

(d) (e) (f)

Figure 8: *Illumination transfer: (a,d) Source and target images. (b,e) Estimation of illumination. (c) Reflection ratio image. (f) Resulting image obtained by multiplying the target image to the reflection ratio image.*

6.2.2. Illumination Decomposition

We can exploit further the illumination estimated by our bilinear illumination model. Using the spherical harmonic method [BJ03], it is difficult to derive individual light sources from the estimated illumination coefficients and thereby limiting the application of relighting. In a similar work exploiting spherical harmonics [WLH03], to relight a face from arbitrary lighting conditions, the authors modify the estimated nine harmonic coefficients interactively. This leads to much difficulty in obtaining the desired illumination as the combination of individual light sources. By contrast, since our bilinear model is derived from a few hundred (146 to be exact) explicit lighting conditions, we can decompose the individual light sources using the estimated illumination bases. We solve the following linear system in terms of $\mathbf{x}$:

$$\mathbf{B}\tilde{\mathbf{U}}_2\mathbf{x} = \tilde{\mathbf{t}}, \tag{17}$$

with constraints:

$$0 \le x_i \le 1, \quad \sum_i x_i = 1,$$

where $\mathbf{B} \in \Re^{N\times 30}$ is the reconstructed *texture weighted* illumination bases, $\tilde{\mathbf{U}}_2 \in \Re^{30\times 146}$ is the mode matrix along the illumination axis (Eq.9), $\tilde{\mathbf{t}} \in \Re^{N\times 1}$ is the reconstructed illuminated texture, and $\mathbf{x} \in \Re^{146\times 1}$ is a weight vector of 146 dome light sources. Eq.17 is a constrained linear least square problem and can be solved by an optimization method based on quadratic programming. For each color channel c, we obtain $\mathbf{x}_c$ by using $\mathbf{B}_c$ and $\tilde{\mathbf{t}}_c$ in Eq.17. Each element of the optimized weight vector $\mathbf{x}_c$ represents the (relative) intensity of the corresponding physical light source of our face scanning dome. To reconstruct the illuminated texture under different combination of dome light sources, we simply generate a new weight vector $\mathbf{x}_{new}$, replace $\mathbf{x}$ in Eq.17 and reconstruct new texture $\tilde{\mathbf{t}}_{new}$. The next step is straightforward using the same technique used in illumination transfer. Since we are using the same face before and after the transfer in this case, we do not need to compute the texture-free illumination component with different diffuse textures. Thus, the corresponding equation of Eq.15 is:

$$\mathbf{t}_{relight} = \hat{\mathbf{t}}. * \tilde{\mathbf{t}}_{new}./\tilde{\mathbf{t}}, \tag{18}$$

where $\hat{\mathbf{t}}$ is the lifted texture from the original input image.

Figure 9 and Table 1 show this procedure. In Figure 9, column labeled (a) is an input image to be relighted and column labeled (b) is the fitting result using bilinear illumination model, which yields $\tilde{\mathbf{t}}$, the reconstructed texture of the input illumination (Eq.17). Table 1 shows the result of computing $\mathbf{x}$ for each color channel. We applied a suitable threshold (0.1) to show the effect of significant light sources. Note the strong contributions from lights numbered 139 and 144 (both of them are located near top in our dome). By setting to zero the intensities in red and green channel respectively, we reconstruct image (c) and (d). Note the changes in illumination on the subject who retains his essential features. Also increasing the intensities for the two light sources $\mathbf{x}(82) = [0.3\ 0.3\ 0]$ and $\mathbf{x}(7) = [0.1\ 0.1\ 0.4]$ (a right and a bottom light source in our dome), we reconstruct image Figure 9(e) and Figure 9(f). Using this approach it is easy to relight a face under arbitrary combinations of densely sampled directional light sources. Figure 10 shows more results of relighting using another input image. In this example, we added individual light sources to the original lighting environment as dictated by our dome light configuration. Input images of Figure 9 and Figure 10 were adopted from FRGC V1.0 database [frg].

7. Conclusions and Future Directions

We presented a novel optimization framework to fit a morphable face model onto photographs. We also presented a novel bilinear illumination model using higher-order SVD that describes 3D shape and illumination variations. Combined, these two approaches lead to a simple and general fitting method with the ability to deal with arbitrary illumination environments and complex face reflectance. We applied our new fitting method to the problem of illumination transfer and face relighting. Our approach for face relighting provides an intuitive and flexible way to change the illumination of a face in a 2D image.

Channel	Light No.	Intensity
Red	15	0.28
	106	0.21
	139	0.31
Green	114	0.18
	144	0.37
Blue	95	0.11
	125	0.10
	144	0.48

Table 1: *Illumination decomposition of the input image in Figure 9.*

Currently, we use the reflectance images acquired from the near-frontal camera to build the bilinear model. Since face reflectance depends on the viewpoint (it is anisotropic and specular), we could construct a separate illumination model for each viewpoint. Then, we could exploit this view-dependent illumination model in our fitting framework. During optimization when the pose parameter (γ) changes, we could dynamically pick the view-dependent illumination model that is closest to the given pose parameter. This would allows us to achieve more accurate reconstructions in even more challenging lighting environments. A vexing problem of storage will become even more challenging and will have to be addressed in earnest.

References

[AGR96] Atick J. J., Griffin P. A., Redlich N.: Statistical approach to shape from shading: Reconstruction of 3D face surfaces from single 2D images. *Neural Computation 8*, 6 (1996), 1321–1340. 1

[BBPV03] Blanz V., Basso C., Poggio T., Vetter T.: Reanimating faces in images and video. In *Proc. of EUROGRAPHICS* (2003). 1, 2

[BJ03] Basri R., Jacobs D.: Lambertian reflectance and linear subspace. *IEEE Transaction on Pattern Analysis and Machine Intelligence 25*, 2 (2003). 1, 2, 4, 5, 8

[BMVS04] Blanz V., Mehl A., Vetter T., Seidel H. P.: A statistical method for robust 3D surface reconstruction from sparse data. In *Int. Symp. on 3D Data Processing, Visualization and Transmission* (2004). 2

[BSVS04] Blanz V., Scherbaum K., Vetter T., Seidel H. P.: Exchanging faces in images. In *Proc. of EUROGRAPHICS* (2004). 1, 2

[BV99] Blanz V., Vetter T.: A morphable model for the synthesis of 3d faces. In *Proceedings of SIGGRAPH 1999* (1999). 1, 2, 4, 7

[BV03] Blanz V., Vetter T.: Face recognition based on fitting a 3D morphable model. *IEEE Transactions on*

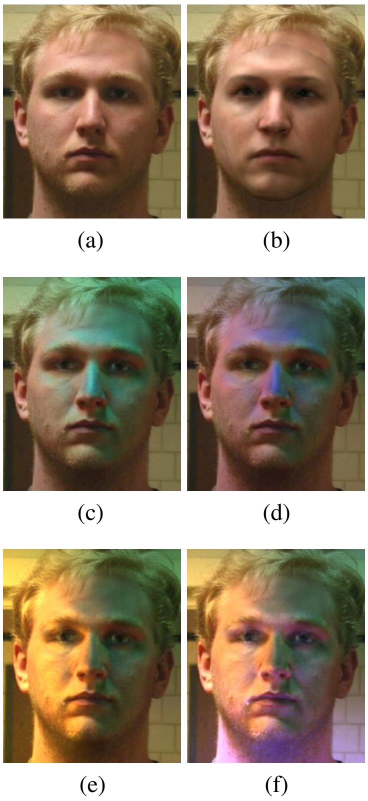


Figure 9: *Face relighting: (a) Input image. (b) Fitted image. (c) Removing the red component of intensity from an upper-positioned light. (d) Removing the green component from the same light. (e) Adding a yellowish light from left. (f) Adding a bluish light from the bottom of the dome.*

Pattern Analysis and Machine Intelligence 25, 9 (2003), 1063–1074. 2

[frg] The NIST Face Recognition Grand Challenge (http://www.frvt.org/FRGC/). 9

[GBK01] GEORGHIADES A. S., BELHUMEUR P. N., KRIEGMAN D. J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 6 (2001), 643–660. 2

[HHB03] HUANG J., HEISELE B., BLANZ V.: Component-based face recognition with 3D morphable models. In *Proc. of the 4th Int. Conf. on Audio- and Video-Based Biometric Person Authenticitation Surrey* (2003). 2

[LMPM03] LEE J., MOGHADDAM B., PFISTER H., MACHIRAJU R.: Silhouette-based 3D face shape recovery. In *Proceedings of Graphics Interface* (2003). 2, 3

[LMV00] LATHAUWER L. D., MOOR B. D., VANDEWALLE J.: A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications 21*, 4 (2000). 5, 6

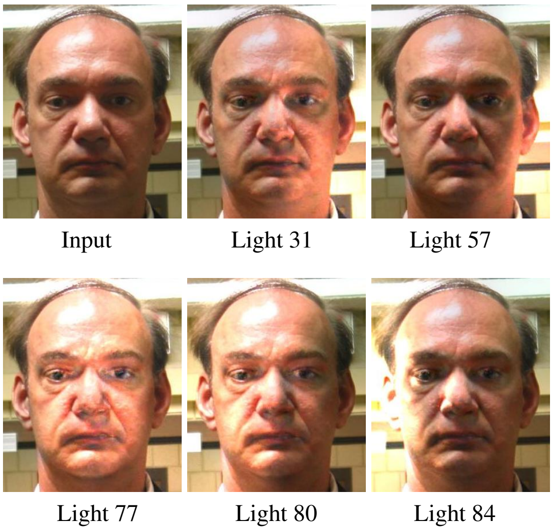


Figure 10: *Face relighting: Adding individual light sources to the original input image as dictated by our dome light configuration.*

[PFTV88] PRESS W. H., FLANNERY B. P., TEUKOLOSKY S. A., VETTERLING W. T.:. In *Numerical Recipes in C: The Art of Scientific Computing* (1988), Cambridge University Press, New York. 5

[RH01] RAMAMOORTHI R., HANRAHAN P.: An efficient representation for irradiance environment. In *Proceedings of SIGGRAPH* (2001), pp. 497–500. 1

[SBB03] SIM T., BAKER S., BSAT M.: The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 12 (2003), 1615–1618. 2, 7

[usf] USF HumanID 3-D Database, Courtesy of Sudeep Sarkar, University of South Florida, Tampa, FL. 4

[VT03] VASILESCU M. A. O., TERZOPOULOS D.: Multilinear subspace analysis of image ensembles. In *Proceedings of Computer Vision and Pattern Recognition* (2003). 2, 5, 6

[VT04] VASILESCU M. A. O., TERZOPOULOS D.: Tensortextures: Multilinear image-based rendering. In *Proceedings of SIGGRAPH* (2004), pp. 336–342. 2

[WLH03] WEN Z., LIU Z., HUANG T.: Face relighting with radiance environment maps. In *Proc. of Computer Vision and Pattern Recognition* (2003). 2, 8

[ZS04] ZHANG L., SAMARAS D.: Pose invariant face recognition under arbitrary unknown lighting using spherical harmonics. In *Proc. of Biometric Authentication Workshop* (2004). 2