A Bilinear Illumination Model for Robust Face Recognition

A Bilinear Illumination Model for Robust Face Recognition

Jinho Lee Baback Moghaddam Hanspeter Pfister Raghu Machiraju

Mitsubishi Electric Research Laboratories (MERL)
201 Broadway, Cambridge MA 02139, USA

Abstract

We present a technique to generate an illumination subspace for arbitrary 3D faces based on the statistics of measured illuminations under variable lighting conditions from many subjects. A bilinear model based on the higher-order singular value decomposition is used to create a compact illumination subspace given arbitrary shape parameters from a parametric 3D face model. Using a fitting procedure based on minimizing the distance of the input image to the dynamically changing illumination subspace, we reconstruct a shape-specific illumination subspace from a single photograph. We use the reconstructed illumination subspace in various face recognition experiments with variable lighting conditions and obtain accuracies which are very competitive with previous methods that require specific training sessions or multiple images of the subject.

1. Introduction

The performance of any face recognition system is adversely affected by facial appearance changes caused by lighting and pose variation. Many attempts have been made to overcome these problems yet it still remains an active area of research in the vision community. One prevalent trend is to exploit the 3D information of human faces to overcome the limitation of traditional 2D images [7, 11, 14, 12, 8, 3, 4]. The 3D shape information can be obtained directly from a range scanner [3, 5] or estimated from a single image [14, 4, 8] or from multiple images [7]. Although the cost of acquiring 3D geometry is decreasing, the fact remains that the majority of existing face databases still consist of single or multiple 2D images. Therefore, it is more practical (if not more accurate) to obtain 3D shape from a single photograph rather than requiring multiple images or 3D data.

There are currently three different approaches in using 3D shape information for robust face recognition: first, using 3D directly as a pose/illumination independent signature [14, 4]; second, using 3D data to generate synthetic imagery under various viewpoints and lighting conditions in order to generate a pose/illumination invariant representation in 2D image space [7, 11, 8]; and third, using 3D shape to derive an analytic illumination subspace of a Lambertian object with spherical harmonics [3, 17].

The first approach is typified by Morphable Models [4] for example, to obtain the 3D shape and 2D texture of the face from a single photograph. The fitted model's shape and texture, extracted from a probe and a gallery image, are matched directly based on their respective PCA coefficients. This approach has showed some promise in dealing with variable pose and lighting, for example in the CMU PIE database [15]. However, it requires careful manual initialization of facial landmarks and uses an iterative nonlinear optimization technique for fitting which can take several minutes to converge (if at all) and then only to a local minimum. Thus, it is not ultimately clear whether this face capture/modeling approach can be used for *real-time* face recognition.

The second and third approaches are qualitatively different and are related to the popular recognition paradigm of "distance-from-a-subspace" which dates back to the early work on 2D appearance-based modeling. Although these two approaches may also use 3D morphable models, it is mostly in the form of a tool for subsequent invariant modeling and subspace generation, as opposed to the final choice of representation for recognition. The methodology proposed in this paper belongs to this latter camp, yet has key differences and presents both computational and modeling advantages not shared by other techniques. For example, it easily surpasses the limitations of a Lambertian (constant BRDF) reflectance model, is based on the statistics of detailed and highly accurate measurements of actual surface reflectance data from a multitude of human subjects under variable illumination, and it can be used to generate a "tailor-made" or *shape-specific* illumination subspace for a given face from a *single* photograph.

2. Background

Several approaches have been reported for generating a linear subspace to capture the illumination variations of a face. Georghiades *et al.* [7] used photometric stereo to reconstruct 3D face geometry and albedo from seven frontal images under different illuminations. The estimated 3D face was then used to render synthetic images from various poses and lighting conditions to train (learn) a person-specific *illumination cone*. In our approach, these three steps (3D estimation → rendering → training) are accomplished by fitting an illumination subspace directly to a *single* image, thereby bypassing the intermediate rendering and training steps.

A similar "short-cut" is implicit in Basri & Jacobs [3] who proposed that the arbitrary illumination of a convex Lambertian 3D object should be approximated by a low-dimensional linear subspace spanned by *nine harmonic images*. The nine harmonic images can easily be computed analytically given surface normals and the albedo. This analytic model makes for rapid generation of an illumination subspace for recognition use. A more practical variation on this theme was proposed by Lee *et al.* [11] who empirically found nine directions of a point source with which to approximate the span of the *nine harmonic images*. These nine images were found to be adequate for face recognition and had the further advantage of not requiring 3D shape (surface normals) and albedo. Of course, it is not always practical to acquire nine images of every subject in a real operational setting, so synthetic images were used instead and their experiments gave results comparable to [7].

Recently, Zhang & Samaras [17] proposed a technique to estimate the *nine harmonic images* from a single image by using a 3D bootstrap dataset (*eg.* the USF HumanID 3D face database [1]). Their approach is closer in spirit to ours but with two key differences. First, since the human face is neither (exactly) Lambertian nor (entirely) convex, spherical harmonics have an inherent limitation, especially when dealing with specularities and cast shadows (not to mention inter-reflections and subsurface scattering). To address this shortcoming, we specifically went after a generative illumination model that was based on detailed measurement of facial reflectance data as captured from a "large" population of subjects under a multitude of lighting conditions and under controlled laboratory conditions (see Figure 1).[1] Secondly, unlike Zhang & Samaras we do not require a bootstrap 3D dataset to (indirectly) estimate the basis images. Instead, we estimate the actual 3D shape from the input 2D image. Moreover, since this shape is already registered (in point-to-point correspondence) with our illumination bases, we can easily render our model to fit images under arbitrary viewpoints (*i.e.*, these canonical basis images are *object-centered* for greater flexibility).

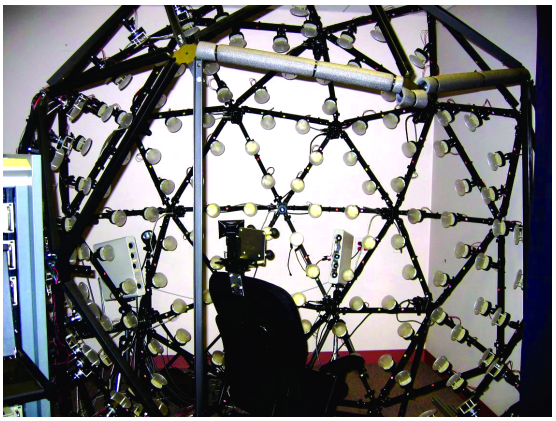


Figure 1. Our face scanning dome for precise (registered) 3D shape and reflectance measurement. The system uses 16 calibrated cameras and 146 light sources.

Since our statistical model was estimated with a great number of high-dimensional training data, generated under multiple factors (shape/identity and illumination), we had to make sure it was compact and manageable in both form and function. While it is possible to implement (linear) interactions with PCA alone, we found that multilinear tensor decompositions were better suited to this particular task. Recently, Vasilescu & Terzopoulos [16] carried out a multilinear analysis of facial images, using higher-order SVD on 2D images under multiple factors such as identity, expression, pose, and illumination. The higher-order decomposition chosen for our specific goal (coupling face shape and reflectance) was a 3-mode tensor SVD, resulting in a *bilinear* illumination model. [2] We should point out however, that we stack 3D shape and reflectance data into our data tensor and not just pixel intensities as in [16].

3. Modeling Methodology

We first describe how we build a generative illumination model from the measurement of subjects under variable lighting conditions. Using our custom-made *face scanning dome*, shown in Figure 1, we acquired faces of 33 subjects under 146 different directional lighting conditions along with the underlying 3D geometry of the faces. The intrinsic and extrinsic parameters of all cameras are accurately calibrated and 3D points of a face are projected onto the corresponding 2D points in each reflectance image through a 3D-2D registration technique (see [10] for a more detailed description).

[1] Naturally, once estimated the model is then easily applied to a variety of arbitrary images in unconstrained "real-world" settings.

[2] Previous examples of the power of bilinear models in computer vision applications include disambiguating *style* and *content* [6].

3.1. Bilinear Illumination Model

We first obtain 3D point-to-point correspondences across different 3D faces using the method described in [10]. Illumination samples (intensities) from each reflectance image are projected from the 3D points on the face, yielding registered 2D samples which are thereby aligned with the 3D shape, all in one common vector space. We also compute a diffuse texture from all illuminated images for each subject. Assuming that facial texture is not coupled with the shape and reflectance, we factor out diffuse texture from the illumination samples:

$$w_k = \hat{t}_k / t_k, \quad k = 1..N,$$

where $\hat{t}_k$ is an illumination sample, t_k is diffuse texture at a 3D point $\mathbf{p}_k$ with N as the number of 3D mesh points. We call w a *texture-free illumination component*, which differs from reflectance since it includes cast shadows.

Consequently, for each subject (i) we have $N(= 10,006)$ 3D points (xyz) and texture-free illumination components (w) for each lighting condition (j) from a specific viewpoint. We align them as a vector $\mathbf{a}_{i,j} = (x_1 \cdots x_N, y_1 \cdots y_N, z_1 \cdots z_N, w_1 \cdots w_N), i = 1..33, j = 1..146$. We stack these vectors along two axes, labeled shape and illumination, and perform a higher-order (3-mode) SVD [9] to capture the joint statistics of both factors. The resulting data array is a tensor $\mathcal{D} \in \Re^{33 \times 146 \times 4N}$ expressed as the product:

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

where *mode matrices* $\mathbf{U}_1 \in \Re^{33 \times 33}$, $\mathbf{U}_2 \in \Re^{146 \times 146}$ and $\mathbf{U}_3 \in \Re^{4N \times 4N}$ capture the variation along the shape, illumination, and data axes, respectively. A *core tensor* $\mathcal{C} \in \Re^{33 \times 146 \times 4N}$ governs the interaction between $\mathbf{U}_1$,$\mathbf{U}_2$, and $\mathbf{U}_3$. See [9] for more details on the *mode-k product* operator, $\times_k$. Using the associative property of the *mode-k product*, the mode matrix $\mathbf{U}_3$ can be incorporated into $\mathcal{Z} = \mathcal{C} \times_3 \mathbf{U}_3$, resulting in a simplified equation:

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

For a more compact representation, we truncate the highest-order singular vectors and retain a reduced lower-dimensional subspace (20 for shape and 30 for illumination) using the algorithm described in [16], thus yielding:

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

where $\tilde{\mathbf{U}}_1 \in \Re^{33 \times 20}$, $\tilde{\mathbf{U}}_2 \in \Re^{146 \times 30}$, and $\tilde{\mathcal{Z}} \in \Re^{20 \times 30 \times 4N}$.

To exploit the redundancy of shape data ($[xyz]$ tuples) along the illumination axis, we divide the core tensor $\tilde{\mathcal{Z}}$ into two parts, $\tilde{\mathcal{Z}}_{xyz} \in \Re^{20 \times 30 \times 3N}$ and $\tilde{\mathcal{Z}}_w \in \Re^{20 \times 30 \times N}$ as in Figure 2. Thus, a data vector with i^{th} shape and j^{th} lighting condition is reconstructed by:

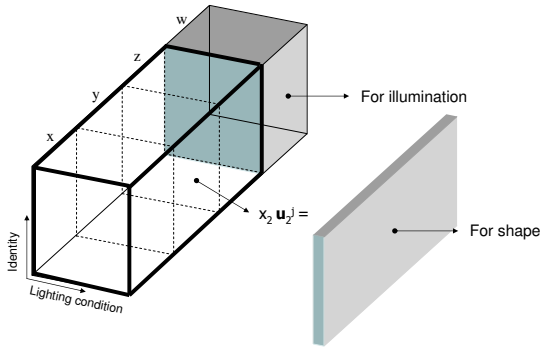


Figure 2. The two (bilinear) components of our tensor, which yields a more compact model by exploiting the redundancy of geometry along the lighting condition axis).

$$\tilde{\mathbf{d}}_{i,j} = (\tilde{\mathcal{Z}}_{xyz} \times_1 \tilde{\mathbf{u}}_1^i \times_2 \tilde{\mathbf{u}}_2^j, \tilde{\mathcal{Z}}_w \times_1 \tilde{\mathbf{u}}_1^i \times_2 \tilde{\mathbf{u}}_2^j). \quad (1)$$

Since the underlying geometry is independent of lighting condition (j), we pre-compute $\tilde{\mathcal{Z}}_{xyz} \times_2 \tilde{\mathbf{u}}_2^j$ with any j, remove a singleton dimension, and yield shape basis row vectors $\mathbf{Z}_s \in \Re^{20 \times 3N}$. Also, shape-specific illumination bases $\mathbf{R}_i \in \Re^{30 \times N}$ are obtained by computing $\tilde{\mathcal{Z}}_w \times_1 \tilde{\mathbf{u}}_1^i$ and removing a singleton dimension. Eq. 1 now becomes:

$$\tilde{\mathbf{d}}_{i,j} = (\tilde{\mathbf{u}}_1^i \mathbf{Z}_s, \tilde{\mathbf{u}}_2^j \mathbf{R}_i), \quad (2)$$

where $\tilde{\mathbf{u}}_1^i$ and $\tilde{\mathbf{u}}_2^j$ can be considered shape and illumination coefficients of $\tilde{\mathbf{d}}_{i,j}$ respectively.

With the two components ($\mathbf{Z}_s, \tilde{\mathcal{Z}}_w$), given any linear combination of shape parameters (α), we can reconstruct the corresponding shape and illumination bases with:

$$\mathbf{s} = \alpha \mathbf{Z}_s; \quad (3)$$

$$\mathbf{R} = \tilde{\mathcal{Z}}_w \times_1 \alpha; \quad (4)$$

$$\alpha = \sum_{i=1}^{33} \alpha_i \tilde{\mathbf{u}}_1^i, \quad (5)$$

where $\mathbf{s}$ is a shape vector (xyz) and the rows of $\mathbf{R}$ are the illumination basis vectors for the specific shape parameter α. Although α is given by Eq. 5, there are cases when an arbitrary shape $\hat{\mathbf{s}}$ is already available from an external parametric shape model. However, we can simply fit $\hat{\mathbf{s}}$ to find the closest shape parameter $\hat{\alpha}$ in our internal shape model by solving the following linear system:

$$\hat{\mathbf{s}} = \hat{\alpha} \mathbf{Z}_s. \quad (6)$$

We have used this technique to estimate an illumination subspace from any external generic shape model (*eg.*, a standard Morphable Model) which we describe in Section 4.

3.2. Beyond Nine Spherical Harmonics

Building the bilinear illumination model with data obtained from one near-frontal camera viewpoint, we performed an experiment to see how well the subspace created by this illumination model could reconstruct the original data. We also compare our accuracy to the that obtained by using nine spherical harmonics as basis images [3].

Since we have ground truth for the 3D shape and illumination samples from 146 lighting conditions and 16 viewpoints for 33 subjects, we measured the variation in reconstruction error from different numbers of bases in each method. For each subject i, we have 3D shape $\mathbf{s}_i = (x_i \cdots x_N, y_1 \cdots y_N, z_1 \cdots z_N)$, diffuse texture $\mathbf{t}_i = (t_1 \cdots t_N)^T$ and illumination samples $\hat{\mathbf{t}}_{i,j,k} = (\hat{t}_1 \cdots \hat{t}_N)^T$ for all lighting conditions $j = 1..146$ and camera viewpoints $k = 1..16$. Since not all the illumination samples are available for each viewpoint (due to occlusion) we use the notation $\tilde{\mathbf{t}}$ for the vector that contains only valid samples.

Given $\mathbf{s}$ and $\tilde{\mathbf{t}}$ (also omitting the indices), we first compute the illumination bases $\mathbf{R}$ using our method (Eq. 3 and 4) and also with nine harmonic images (see [3]). Then the diffuse texture $\mathbf{t}$ is multiplied by each column of $\mathbf{R}^T$ in element-wise manner. This yields the texture-weighted illumination bases $\mathbf{B}$ and the reconstruction error for $\tilde{\mathbf{t}}$ is:

$$error = \|\tilde{\mathbf{t}} - \tilde{\mathbf{B}}\hat{\mathbf{B}}^T\tilde{\mathbf{t}}\|, \tag{7}$$

where $\hat{\mathbf{B}}$ is computed with a standard QR decomposition of $\tilde{\mathbf{B}}$ which contains only the valid rows of $\mathbf{B}$ corresponding to $\tilde{\mathbf{t}}$.

We computed the reconstruction errors for all combination of subjects, lighting conditions, camera viewpoints, and the number of bases used for reconstruction and for each method to generate an illumination subspace. Figure 3 shows the resulting reconstruction errors. The top plot compares the two methods with different (total) number of bases. The bottom plot compares the two methods for the different camera viewpoints (see Table 1 for camera angles).

4. Estimation with a Single Image

We now describe how to obtain the person-specific illumination subspace given a single image of an individual. The illumination bases are derived from our bilinear illumination model (BIM) after fitting a morphable model to the input image, by minimizing the distance of the input image to the dynamically generated illumination subspace.

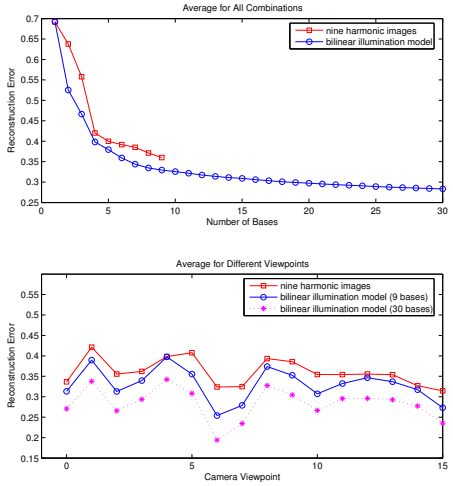


Figure 3. (Top) Average reconstruction error *vs.* number of bases in our bilinear model (blue) and 9 harmonic images (red). (Bottom) Reconstruction error *vs.* camera viewpoint.

4.1. Shape-Specific Illumination Subspace

We build a morphable model [4] from the combined imagery of the USF Human ID database [1] (134 subjects) and our own internal database (71 subjects). 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 principal component analysis (PCA) on all shape vectors $\mathbf{S}$ and texture vectors $\mathbf{T}$ separately. Using the first M eigenvectors and model parameters α and β, arbitrary shape and texture are reconstructed using,

$$\mathbf{s} = \bar{\mathbf{S}} + \sum_{i=1}^{M} \alpha_i \mathbf{e}_i^s, \quad \mathbf{t} = \bar{\mathbf{T}} + \sum_{i=1}^{M} \beta_i \mathbf{e}_i^t, \tag{8}$$

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

The optimization parameters are α, β and γ which is a 6-dimensional pose parameter (3 for translation, 3 for rotation). During each iteration, we generate shape ($\mathbf{s}$) and diffuse texture ($\mathbf{t}$) from parameters α and β and then extract texture $\hat{\mathbf{t}}$ by projecting $\mathbf{s}$ to the input image at the given pose γ. The optimal parameters are found by minimizing an error function similar to Eq. 7. Instead of $\hat{\mathbf{t}}$, we use $\tilde{\mathbf{t}}$ which contains only the visible points in the extracted texture, thus

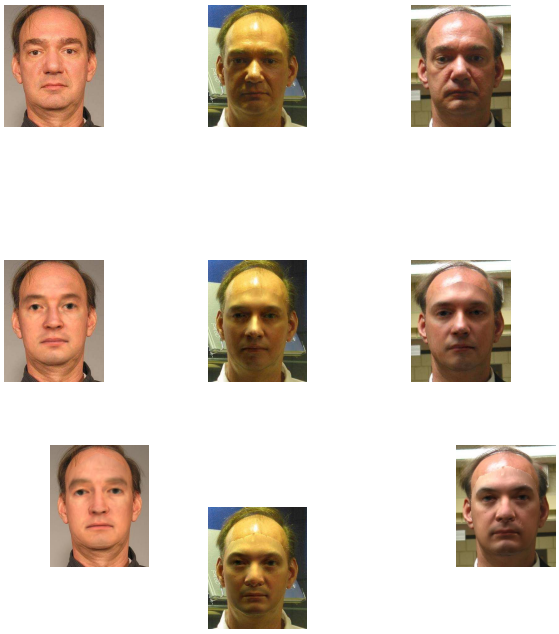


Figure 4. (Top) Input images (from FRGC dataset [2]) and fitted models (overlaid): (Middle) using bilinear model with 30 bases. (Bottom) using 9 harmonic images.

yielding the following optimization problem:

$$\arg\min_{\alpha,\beta,\gamma} \|\tilde{\mathbf{t}} - \hat{\mathbf{B}}\hat{\mathbf{B}}^T\tilde{\mathbf{t}}\|, \tag{9}$$

We use the Downhill Simplex Method [13], a robust non-linear optimization algorithm requiring cost function evaluations only (no gradients needed). Figure 4 shows results for three images taken under different illuminations of a subject from the FRGC database [2], where our bilinear illumination model (middle row) has reconstructed scene illumination better than nine harmonic images (*eg.*, see the forehead and nose in 2nd and 3rd examples). Note that since our model uses an adaptive illumination subspace during optimization, the final reconstructed shape and texture in the figure need not be the same for both methods.

While the shape, texture and pose parameters estimated by this optimization framework are important in reconstruction, we are mainly concerned with optimal characterization of illumination bases $\hat{\mathbf{B}}_{opt}$. These bases span the illumination subspace of the person with the shape $\mathbf{s}(\alpha_{opt})$ and the diffuse texture $\mathbf{t}(\beta_{opt})$.

Figure 5 shows the fit (top row) and the first 3 reconstructed illumination bases (middle row) for an image from the Yale Face Database B. In this case, the illumination bases were weighted by the synthesized texture $\mathbf{t}(\beta_{opt})$.

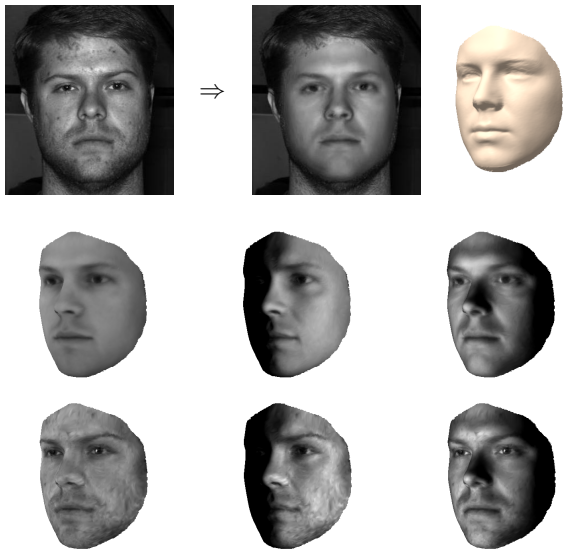


Figure 5. Input image (top left), its overlaid model texture (top middle) and 3D shape (top right) using bilinear illumination model. First 3 illumination bases with synthesized texture (middle row) and with real texture (bottom row).

However, the synthesized texture cannot capture all the details of the face in the original input image. For purposes of face recognition, it is important to obtain the real-texture weighted illumination bases. We will use the following notation in subsequent discussion:

- $\mathbf{t}_s$: synthesized diffuse texture (known)
- $\hat{\mathbf{t}}_s$: synthesized illuminated texture (known)
- $\mathbf{t}_r$: real diffuse texture (unknown)
- $\hat{\mathbf{t}}_r$: real illuminated texture (known)
- define $\mathbf{A} \otimes \mathbf{b}$, $\mathbf{A} \oslash \mathbf{b}$ as element-wise multiplication (division) of vector $\mathbf{b}$ with all column vectors of $\mathbf{A}$

In each iteration, illumination bases are first computed by:

$$\mathbf{B} = \mathbf{R} \otimes \mathbf{t}_s, \tag{10}$$

and new bases are obtained by replacing $\mathbf{t}_s$ with $\mathbf{t}_r$ such as:

$$\mathbf{B}^* = \mathbf{B} \oslash \mathbf{t}_s \otimes \mathbf{t}_r. \tag{11}$$

Assuming that our estimated illumination approximates the original illumination, we get

$$\mathbf{t}_r \approx \hat{\mathbf{t}}_r \otimes \mathbf{t}_s \oslash \hat{\mathbf{t}}_s. \tag{12}$$

Finally, substituting Eq. 12 into Eq. 11 yields:

$$\mathbf{B}^* \approx \mathbf{B} \otimes \hat{\mathbf{t}}_r \oslash \hat{\mathbf{t}}_s. \tag{13}$$

The bottom row of Figure 5 shows the first three illumination bases weighted by the real diffuse texture.

4.2. Illumination Bases for Face Recognition

Although illumination bases in a common vector space are useful for pose-invariant face recognition, they have one disadvantage. Since all the extracted textures are registered in a shape-free vector space, we lose all shape information for matching. It is generally accepted that texture is an important identity cue, but 3D shape is increasingly important under extreme lighting conditions. In the majority of face recognition systems, probe and gallery images are often aligned using only the eye locations, with other facial areas transformed accordingly. Shape information is exploited by all algorithms either implicitly or explicitly. Therefore, it is often more practical to have illumination bases in the 2D image space as opposed to in a shape-free 3D space.

Figure 9 shows the illumination bases rendered in the same size as the gallery. They were computed using a similar method to that described in Section 4. First the bases registered with a 3D shape are divided by the corresponding reconstructed illumination samples $(\mathbf{B} \oslash \mathbf{t}_s)$ and projected to the image plane in which the fitting is performed. The projected pixel data is densely computed using push-pull interpolation in the cropped image plane and finally multiplied by the original cropped image. This procedure is performed for each reconstructed basis.

5. Experiments

5.1. Experiment 1

We first performed an experiment using *nine harmonic images* to see how the corresponding linear subspace can accommodate images under different illumination models. This experiment was carried out with the USF 3D Face Database [1]. The necessary 3D point-to-point correspondence was established across different faces and all surface points and texture color values are registered in a common vector space ($48707 \cdot 3$). For gallery face representations, we generate the *nine harmonic images* for each of the 138 3D faces in the database. For the probe faces, we use color values of surface points that are computed using two different illumination models for four different lighting directions: (i) Lambertian (ii) Phong model. We project these $4 \cdot 138 = 552$ illuminated faces to the illumination subspace spanned by nine harmonic images of each gallery face and measure the distance and find the rank of the matching scores. Figure 6 shows the cumulative recognition rates obtained with different illumination models and compared with simple template-matching. Note that the *nine harmonic images* performance is perfect with faces rendered with a Lambertian model but not so when applied to imagery from different (more realistic) illumination.

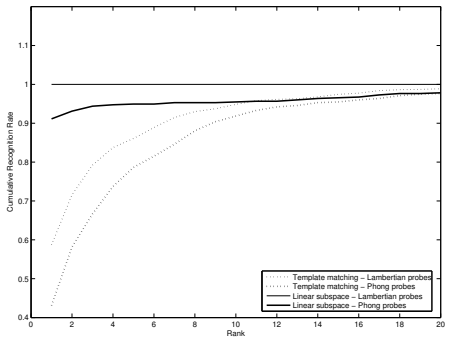
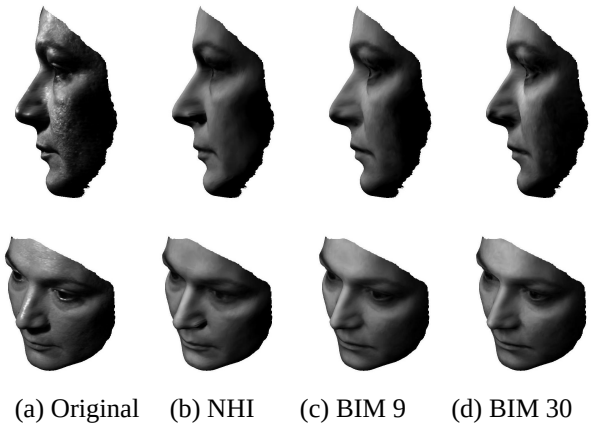


Figure 6. Recognition performance of nine harmonic images with Lambertian and Phong lighting models.

5.2. Experiment 2

In this experiment we compare the recognition performance of our bilinear illumination model with *nine harmonic images*. Our model comes from a single near-frontal viewpoint and being a view-specific model it is tested for extrapolation to data obtained from fifteen other viewpoints. Given a 3D shape along with the illuminated texture (Figure 7(a)), we first create the illumination bases using the 3D shape with different methods and compute the closest illuminated texture to the subspaces spanned by different bases. Figure 7(b,c,d) show the reconstructed illuminated faces using *nine harmonic images* as well as our model with nine and thirty bases. Note that highlights and shadows are partially preserved using our bilinear illumination model.

As with Experiment 1, we perform recognition tests using various lighting conditions from all 16 different viewpoints. Twenty harsh illuminations were chosen for each viewpoint (Figure 8) and a total of $20 \times 33 = 660$ probe illuminations were used for the test in each viewpoint. For each probe subject, the distance to the subspaces spanned by the three different set of illumination bases (b,c,d) of each gallery subject is computed. The resulting recognition rates reported in Table 1 show that our model is more competitive even with just 9 bases. Note that the dimensionality of our illumination model is reduced along both the shape and illumination axes in a single viewpoint yet it still retains superior recognition performance with other viewpoints and under very harsh illumination conditions. The imagery



(a) Original (b) NHI (c) BIM 9 (d) BIM 30

Figure 7. Subspace reconstruction: (a) original illumination (b) nine harmonic images (c) bilinear model with 9 bases (d) bilinear model with 30 bases.

Figure 8. 20 illumination conditions from viewpoint #10 used for the recognition test in Experiment 2.

Viewpoint		Rec. rate (%): Illum. Bases		
	(Az.,El.)	**NHI**	**BIM 9**	**BIM 30**
0	(0,-61)	66	75	86
1	(72,-29)	77	94	94
2	(36,-35)	80	97	98
3	(-36,-35)	78	90	96
4	(-72,-29)	80	78	92
5	(54,0)	83	98	99
6*	(18,0)	90	99	100
7	(-18,0)	90	98	100
8	(-54,0)	87	90	98
9	(72,35)	88	96	98
10	(36,29)	84	97	98
11	(-36,29)	93	98	98
12	(-72,35)	89	87	92
13	(36,61)	80	88	92
14	(-36,61)	80	84	89
15	(0,35)	80	96	99

Table 1. Recognition under view extrapolation for bilinear model (from viewpoint 6) and nine spherical harmonics.

Comparison of Recognition Methods			
Methods	Error Rate (%) vs. Illum.		
	Subset 1,2	Subset 3	Subset 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	0.0	0.3	3.1
BIM (9 Bases)	0.0	0.0	7.1
BIM (30 bases)	0.0	0.0	0.7

Table 2. Recognition results using various methods in the literature (data summarized from [17]).

from viewpoint 6 was used as reference to build our model.

5.3. Experiment 3

Finally, we did recognition experiments on the Yale Face Database B [7]. For the recognition test, we used 450 images of 10 subjects under 45 different lighting directions at a fixed (frontal) pose. The different lighting conditions are divided into four subsets according to the angle between the light direction and camera axis [7]. We use a single image for each subject in the gallery and use the remaining 440 images for probe images. From each gallery image, we reconstruct the computed subspace by the method presented in Section 4.2. Figure 9 shows two subjects in the database and the first nine reconstructed bases. The matching scores between the probe and gallery images are computed using Eq. 7. Table 2 shows the results of the recognition rate compared to other published results. We note that all these methods — except 9PL (*nine points of light* [11]) and Zhang & Samaras [17] — require off-line processing using multiple training images of each gallery subject. Although 9PL does not require *training* images *per se*, it does require nine images taken under preset lighting conditions in order to span the illumination subspace of each individual. Our model requires only a single image.

6. Conclusion

We proposed a novel method for computing an illumination subspace by extracting 3D shape from a single image.

Figure 9. Single input image for 2 subjects (1st column, top/bottom). The subsequent 9 columns are the first 9 illumination bases computed from the original input images. All basis images are scaled independently to display full range of spatial variation.

Figure 10. Recognition under harsh illumination: 3 probe images (left column) and BIM reconstructions (middle 9D, right 30D) with minimum distance to the input probes.

To deal with the complex reflectance properties of human faces, we exploited a compact illumination model derived from the joint statistics of 3D surface points and precisely registered illumination samples under varied lighting conditions. The experimental results show that this model has better reconstruction and recognition performance than related analytic models. Moreover, it has good extrapolation across pose. With the Yale Face Database B, our method was (at the very least) comparable to the prior art despite the much simpler computation for obtaining an illumination-invariant face representation from a single image. Finally, our method has the potential for robust pose-invariant recognition using reconstructed illumination bases that are registered with the recovered 3D shape.

Acknowledgments

We thank Tim Weyrich (ETH Zürich) for his invaluable work on the scanning dome, and also Addy Ngan (MIT).

References

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

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

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspace. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003.

[4] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[5] K. I. Chang, K. Bowyer, and P. Flynn. Face recognition using 2D and 3D facial data. In *Multimodal User Authentication Workshop*, 2003.

[6] W. Freeman and J. Tennenbaum. Learning bilinear models for two-factor problems in vision. In *Proc. of Computer Vision & Pattern Recognition*, pages I:19–25, 1997.

[7] 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[8] J. Huang, B. Heisele, and V. Blanz. Component-based face recognition with 3D morphable models. In *Proc. of the 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authenticitation Surrey*, 2003.

[9] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4), 2000.

[10] J. Lee, R. Machiraju, H. Pfister, and B. Moghaddam. Estimation of 3D faces and illumination from a single photograph using a bilinear illumination model. In *Proc. of Eurographics Symposium on Rendering*, 2005.

[11] K. Lee, J. Ho, and D. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. In *Proc. of Computer Vision & Pattern Recognition*, volume 1, pages 519–526, 2001.

[12] B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju. Model-based 3D face capture with shape-from-silhouettes. In *Proc. of Advanced Modeling of Faces & Gestures*, 2004.

[13] W. H. Press, B. P. Flannery, S. A. Teukolosky, and W. T. Vetterling. In *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, New York, 1988.

[14] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In *European Conference on Computer Vision*, pages 3–19, 2002.

[15] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Proc. Int'l Conf. Automatic Face & Gesture Recognition*, pages 53–58, 2002.

[16] M. A. O. Vasilescu and D. Terzopoulos. Multilinear subspace analysis of image ensembles. In *Proc. of Computer Vision & Pattern Recognition*, 2003.

[17] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *Proc. Computer Vision & Pattern Recognition*, pages I:19–25, 2003.