To Appear in the IEEE Trans. on Pattern Analysis and Machine Intelligence

From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose

Athinodoros S. Georghiades Peter N. Belhumeur

Departments of Electrical Engineering
and Computer Science
Yale University
New Haven, CT 06520-8285
[georghiades, belhumeur]@yale.edu

David J. Kriegman

Beckman Institute
University of Illinois, Urbana-Champaign
Urbana, IL 61801
kriegman@uiuc.edu

Abstract

We present a generative appearance-based method for recognizing human faces under variation in lighting and viewpoint. Our method exploits the fact that the set of images of an object in fixed pose, but under all possible illumination conditions, is a convex cone in the space of images. Using a small number of training images of each face taken with different lighting directions, the shape and albedo of the face can be reconstructed. In turn, this reconstruction serves as a generative model that can be used to render—or synthesize—images of the face under novel poses and illumination conditions. The pose space is then sampled, and for each pose the corresponding illumination cone is approximated by a low-dimensional linear subspace whose basis vectors are estimated using the generative model. Our recognition algorithm assigns to a test image the identity of the closest approximated illumination cone (based on Euclidean distance within the image space). We test our face recognition method on 4050 images from the Yale Face Database B; these images contain 405 viewing conditions (9 poses $\times$ 45 illumination conditions) for 10 individuals. The method performs almost without error, except on the most extreme lighting directions, and significantly outperforms popular recognition methods that do not use a generative model.

Index Terms: Face Recognition, Image-Based Rendering, Appearance-Based Vision, Face Modeling, Illumination and Pose Modeling, Lighting, Illumination Cones, Generative Models.

1 Introduction

It has been observed that "the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity" [46]. As is evident in Figures 1, 2, and 4, the same person, with the same facial expression, can appear strikingly different when light source direction and viewpoint vary.

Over the last few years, numerous algorithms have been proposed for face recognition, see surveys [59, 7, 17, 50]. For decades, geometric feature-based methods [21, 33, 35, 34, 27, 26, 59, 6, 69, 42] have used properties and relations (e.g., distances and angles) between facial features such as eyes, mouth, nose, and chin to perform recognition under variable illumination and pose. Despite their economical representation and their insensitivity to variations in illumination and viewpoint, feature-based methods are quite sensitive to the feature extraction and measurement process. It has been argued that existing techniques for the extraction and measurement of facial features are not reliable enough [12]. It has also been claimed that methods for face recognition based on finding local image features and inferring identity by the geometric relations of these features are ineffective [6].

Methods have recently been introduced which use low-dimensional representations of images of objects to perform recognition. See for example [36, 66, 23, 51, 55, 47, 45, 25]. These methods, often termed appearance-based methods, differ from feature-based techniques in that their low-dimensional representation is, in a least-squares sense, faithful to the original image. Techniques such as SLAM [47] and Eigenfaces [66] have demonstrated the power of appearance-based methods both in ease of implementation and in accuracy.

Despite their success, much of the previous appearance-based methods suffer from an important drawback: recognition of an object under a particular lighting and pose can be performed reliably *provided the object has been previously seen under similar circumstances*. In other words, these methods in their original form cannot extrapolate to novel viewing conditions.

In this paper, we present a generative model for face recognition. Our approach is, in spirit, an appearance-based method. However, it differs substantially from previous methods in that a small number of training images are used to synthesize novel images under changes in lighting and viewpoint. Our face recognition method exploits the following main observations:

1. The set of images of an object in fixed pose but under all possible illumination conditions is

a convex cone (termed the illumination cone) in the space of images [1].

2. When the surface reflectance can be approximated as Lambertian, this illumination cone can be constructed from a handful of images acquired under variable lighting [1].
3. An illumination cone can be well approximated by a low-dimensional linear subspace [1].
4. Under variable lighting and pose, the set of images is characterized by a family of illumination cones parameterized by the pose. The illumination cones for non-frontal poses can be constructed by applying an image warp on the extreme rays defining the frontal cone.

To construct the illumination cone, the shape and albedo of each face is reconstructed using our own variant of photometric stereo. We use as few as seven images of a face seen in a fixed pose, but illuminated by point light sources at varying, unknown positions, to estimate its surface geometry and albedo map up to a generalized bas-relief (GBR) transformation [2, 20, 18, 19]. (A GBR transformation scales the surface and introduces an additive plane.) We then exploit the symmetries and similarities in faces to resolve the three parameters specifying the GBR transformation.

Using the estimated surface geometry and albedo map, synthetic images of the face could then be rendered by varying lighting directions and viewpoint. However, because the space of lighting conditions is infinite dimensional, sampling this space is no small task. To get around this, we take advantage of the fact that, under fixed viewpoint, the set of all n-pixel images of a face (or any object), under arbitrary illumination, forms a convex cone in the image space $\mathbb{R}^n$ [1]. Since this illumination cone is convex, it is characterized by a set of extremal rays (i.e., images of the face illuminated by appropriately chosen single point light sources), and all other images in the cone are formed by convex combinations of the extreme rays. The cone can be simplified in two ways: Using a subset of the extreme rays and approximating it as a low-dimensional linear subspace [20, 19]. This method for handling lighting variability in images of human faces differs from [23, 24, 25] in that our model is generative—it requires only a few images to predict large image changes. It is, in spirit, most closely related to the synthesis approaches suggested in [61, 56] and stands in stark contrast to the illumination insensitivity techniques argued for in [3, 8].

To handle image variation due to viewpoint, we warp the images defining the frontal illumination cone in a manner dictated by 3-D rigid transformations of the reconstructed surface geometry. This method for handling pose variability differs from [4, 38, 30, 43] in that we warp synthetic frontal-pose images of each face using its estimated surface geometry. Thus, for each face we

generate a collection of illumination cones—one for each sampled viewpoint. Each pose-specific illumination cone is generated by warping the images corresponding to its extremal rays.

For our recognition algorithm, we could assign to a test image the identity of the closest cone (based on Euclidean distance). However, this would require computing the distance of the test image to all illumination cones for all people over all viewpoints and would be computationally expensive. To avoid this expense, we use SVD (singular value decomposition) to compress each face representation independently. This compression is done in two stages. First, we approximate each of the illumination cones with its own low-dimensional subspace, as suggested in [1]. We then project the compressed illumination cones of a single face down to a low-dimensional subspace specific to the face. This low dimensional subspace can be determined by applying SVD on all the images in the cones of the face.

We should point out that our dimensionality reduction techniques are similar to those used in [45], but they differ on three important counts. First, in our method the representation of *each* face is a collection of low-dimensional subspaces, one per sampled viewpoint. That is, each representation is face-specific. This is in contrast to [45] where all faces are modeled by a single collection of low-dimensional subspaces, with each subspace modeling the appearance of *all* faces in one particular view. Second, in our method each subspace explicitly *models* the image variability of one face in a particular pose under different illumination conditions. Third, the images used to generate our subspaces are synthetic, rendered from the small set of training images.

This generative, or extrapolating, ability of our method distinguishes it from previous approaches to face recognition under variable illumination and pose. In a recent approach, a 3-D model of a face (shape and texture) is utilized to transform the input image into the same pose as the stored prototypical faces, and then direct template matching is used to recognize faces [4, 5, 68, 67]. Similarly, a simple, generic 3-D model of a face is used to estimate the pose and light source direction in the input image [75]. This image is then converted into a synthesized image with virtual frontal view and frontal illumination. This virtual image is finally fed into a system such as LDA [76] (similar to the Fisherfaces method [3]) for recognition. In another approach, an Active Appearance Model of a generic face is deformed to fit to the input image, and the control parameters are used as a feature vector for classification [10, 39, 14, 11]. In our method, on the other hand, image variability is explicitly modeled, and we therefore skip the intermediate steps of fitting parameters or estimating the pose and light source direction.

To test our method, we perform face recognition experiments on a 4050 image subset of the publicly available Yale Face Database B[1]. This subset contains 405 viewing conditions (9 poses $\times$ 45 illumination conditions) for 10 individuals. Figure 3 shows the 10 individuals from the Yale Face Database B used to test our method, while Figure 4 shows a single individual seen under the 405 viewing conditions used in the testing. Our method performs almost without error on this database, except on the most extreme lighting directions.

While there is an ever growing number of face databases, we have developed the Yale Face Database B to allow for systematic testing of face recognition methods under large variations in illumination and pose. This is in contrast to other databases, such as the FERET, which contain images of a large number of subjects but captured under limited variations in pose and illumination. Even though many face recognition algorithms have been tested and evaluated using the FERET database [52, 53, 54], FERET does not allow for a systematic study of the effects of illumination and pose. Although the Harvard Robotics Lab database [23, 24, 25] contains images of faces with large variations in illumination, the pose is fixed throughout.

We concede that in our experiments, and in our database, we have made no attempt to deal with variations in facial expression, aging, or occlusion (e.g., beards and glasses). Furthermore, we assume that the face to be recognized has been located (but not necessarily accurately aligned) within the image, as there are numerous methods for finding faces in images [13, 55, 60, 38, 9, 41, 44, 40, 32, 22, 72, 45, 58, 57, 71]. Instead, our method performs a local search over a set of image transformations.

In Section 2, we will briefly describe the illumination cone representation and show how we can construct it using a small number of training images. We will then show how to synthesize images under differing lighting and pose. In Section 3, we will explain the construction of face representations and then describe their application in new face recognition algorithms. Finally, Section 4 presents experimental results.

2 Modeling Illumination and Pose Variability

2.1 The Illumination Cone

In earlier work, it was shown that the set of all n-pixel images of an object under arbitrary combinations of point or extended light sources forms a convex cone $\mathcal{C}$ in the image space $\mathbb{R}^n$. This

[1] `ftp://plucky.cs.yale.edu/CVC/pub/images/yalefacesB/`

cone, termed the illumination cone, can be constructed from as few as three images [1] if the object is convex in shape and has Lambertian reflectance.

Let the surface of a convex object be specified by the graph of a function $z(x, y)$. Let the surface have a Lambertian reflectance [37] with albedo $\alpha(x, y)$ and be viewed orthographically. Let $\mathbf{b}(x, y)$ be a row vector determined by the product of the albedo with the inward pointing unit normal of a point (x, y) on the surface. We can write $\mathbf{b}(x, y)$ as

$$\mathbf{b}(x, y) = \alpha(x, y)\frac{(z_x(x, y), z_y(x, y), -1)}{\sqrt{z_x^2(x, y) + z_y^2(x, y) + 1}}, \tag{1}$$

where $z_x(x, y)$ and $z_y(x, y)$ are the $x-$ and $y-$partial derivatives. Let the object be illuminated by a point light source at infinity; let the light source be specified by $\mathbf{s} \in \mathbb{R}^3$ signifying the product of the light source intensity with a unit vector in the direction of the light source.

Let the vector $\mathbf{x} \in \mathbb{R}^n$ denote an n-pixel image of the object whose surface is specified by $z(x, y)$. A coordinate of $\mathbf{x}$ specifies the image irradiance of a pixel which views a surface patch centered at some point (x, y). Let $B \in \mathbb{R}^{n\times 3}$ be a matrix where each row is given by $\mathbf{b}(x, y)$ as defined above. Let the rows of the $\mathbf{x}$ and the rows of B correspond to the same points (x, y) on the object's surface. Under the Lambertian model of reflectance, the image $\mathbf{x}$ is given by

$$\mathbf{x} = \max(B\mathbf{s}, \mathbf{0}), \tag{2}$$

where $\max(B\mathbf{s}, \mathbf{0})$ sets to zero all negative components of the vector $B\mathbf{s}$. The pixels set to zero correspond to the surface points lying in attached shadows. Convexity of the object's shape is assumed at this point to avoid cast shadows. While attached shadows are defined by local geometric conditions, cast shadows must satisfy a global condition. Note that when no part of the surface is shadowed, $\mathbf{x}$ lies in the 3-D "illumination subspace" $\mathcal{L}$ given by the span of the columns of B [23, 48, 62]; the subset $\mathcal{L}_0 \subset \mathcal{L}$ having no shadows (i.e., intersecting with the non-negative orthant[2]) forms a convex cone [1].

If an object is illuminated by k light sources at infinity, then the image is given by the superposition of the images which would have been produced by the individual light sources, i.e.,

$$\mathbf{x} = \sum_{i=1}^{k} \max(B\mathbf{s}_i, \mathbf{0}) \tag{3}$$

[2]By orthant we mean the high-dimensional analogue to quadrant, i.e., the set $\{\mathbf{x}|\mathbf{x} \in \mathbb{R}^n$, with certain components of $\mathbf{x} \geq 0$ and the remaining components of $\mathbf{x} < 0\}$. By non-negative orthant we mean the set $\{\mathbf{x}|\mathbf{x} \in \mathbb{R}^n$, with all components of $\mathbf{x} \geq 0\}$.

where $\mathbf{s}_i$ is a single light source. Due to this superposition, the set of all possible images $\mathcal{C}$ of a convex Lambertian surface created by varying the direction and strength of an arbitrary number of point light sources at infinity is a convex cone. It is also evident from Equation 3 that this convex cone is completely described by matrix B. (Note that $B^* = BA$, where $A \in GL(3)$ is a member of the general linear group of 3×3 invertible matrices, also describes this cone.)

Furthermore, any image in the illumination cone $\mathcal{C}$ (including the boundary) can be determined as a convex combination of *extreme rays* (images) given by

$$\mathbf{x}_{ij} = \max(B\mathbf{s}_{ij}, \mathbf{0}), \tag{4}$$

where

$$\mathbf{s}_{ij} = \mathbf{b}_i \times \mathbf{b}_j. \tag{5}$$

The vectors $\mathbf{b}_i$ and $\mathbf{b}_j$ are rows of B with $i \neq j$. It is clear that there are at most $m(m-1)$ extreme rays for $m \leq n$ independent surface normals. Since the number of extreme rays is finite, the convex illumination cone is polyhedral.

2.2 Constructing the Illumination Cone

Equations 4 and 5 suggest a way to construct the illumination cone for each face: gather three or more images under differing illumination without shadowing and use these images to estimate a basis for the three-dimensional illumination subspace $\mathcal{L}$. At the core of our approach for generating images with novel lighting viewpoints is a variant of photometric stereo [64, 70, 29, 28, 73] which simultaneously estimates geometry and albedo across the scene. However, the main limitation of classical photometric stereo is that the light source positions must be accurately known, and this necessitates a fixed, calibrated lighting rig (as might be possible in an industrial setting). Instead, the proposed method *does not* require knowledge of light source locations—illumination could be varied by simply waiving a light source around the scene.

A method to estimate a basis for $\mathcal{L}$ is to normalize the images to be of unit length, and then use singular value decomposition (SVD) to find the best (in a least-squares sense) 3-D orthogonal basis in the form of matrix B^*. This task can be cast into a minimization problem given by

$$\min_{B^*,S} ||X - B^*S||^2 \tag{6}$$

where $X = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$ is the data matrix for k images of a face (in vector form), and S is a $3 \times k$ matrix whose columns, $\mathbf{s_i}$, are the light source directions scaled by their corresponding source intensities for all k images.

Note that even if the columns of B^* exactly span the subspace $\mathcal{L}$, they differ from those of B by an unknown linear transformation, i.e., $B^* = BA$ where $A \in GL(3)$; for any light source, $\mathbf{x} = B\mathbf{s} = (BA)(A^{-1}\mathbf{s})$ [28]. Nonetheless, both B^* and B define the same illumination cone $\mathcal{C}$ and represent valid illumination models. From B^*, the extreme rays defining the illumination cone $\mathcal{C}$ can be computed using Equations 4 and 5.

Unfortunately, using SVD in the above procedure leads to an inaccurate estimate of B^*. Even for a convex object, whose occluding contour is visible, there is only one light source direction (the viewing direction) for which no point on the surface is in shadow. For any other light source direction, shadows will be present. If the object is non-convex, such as a face, then shadowing in the modeling images is likely to be more pronounced. When SVD is used to find B^* from images with shadows, these systematic errors can bias its estimate significantly. Therefore, an alternative method is needed to find B^*, one that takes into account the fact that some data values are invalid and should not be used in the estimation. For the purpose of this estimation, any invalid data will be treated as missing measurements. The minimization problem stated in Equation 6 can then be reformulated into:

$$\min_{\mathbf{b}_j^*, \mathbf{s}_i} \sum_{ij} w_{ij} | x_{ij} - < \mathbf{b}_j^*, \mathbf{s}_i > |^2 \tag{7}$$

where x_{ij} is the j-th pixel of the i-th image, $\mathbf{b}_j^*$ is the j-th row of matrix $B^* \in \mathbb{R}^{n \times 3}$, $\mathbf{s}_i$ is the light source direction and strength in the i-th image, and

$$w_{ij} = \begin{cases} 1 & x_{ij} \text{ valid measurement,} \\ 0 & \text{otherwise.} \end{cases}$$

The technique we use to solve this minimization problem is a combination of two algorithms. A variation of [63] (see also [31, 65]) which finds a basis for the 3-D linear subspace $\mathcal{L}$ from image data with missing elements is used together with the method in [16] which enforces integrability in shape from shading. We have modified the latter method to guarantee integrability in the estimates of the basis vectors of subspace $\mathcal{L}$ from multiple images. By enforcing integrability, a surface context is introduced. Namely, the vector field induced by the basis vectors is guaranteed to be a gradient field that corresponds to a surface.

Furthermore, enforcing integrability inherently leads to more accurate estimates because there are fewer parameters (or degrees of freedom) to determine. It also resolves six out of the nine parameters of $A \in GL(3)$. The other three parameters correspond to the generalized bas-relief

(GBR) transformation which cannot be resolved with illumination information alone (i.e. shading and shadows) [2, 74]. (We term this the GBR ambiguity.) This means we cannot recover the true matrix B and its corresponding surface, $z(x, y)$; we can only find their GBR transformed versions $\bar{B}$ and $\bar{z}(x, y)$.

Our estimation algorithm is iterative. To enforce integrability, the possibly non-integrable vector field induced by the current estimate of B^* is, in each iteration, projected down to the space of integrable vector fields, or gradient fields [16]. Let us expand the surface $\bar{z}(x, y)$ using basis surfaces (functions):

$$\bar{z}(x, y; \bar{c}(\mathbf{w})) = \sum \bar{c}(\mathbf{w})\phi(x, y; \mathbf{w}) \tag{8}$$

where $\mathbf{w} = (u, v)$ is a two dimensional index over which the sum is performed, and $\{\phi(x, y; \mathbf{w})\}$ is a finite set of basis functions which are not necessarily orthogonal. We chose the discrete cosine basis so that $\{\bar{c}(\mathbf{w})\}$ is exactly the *full* set of discrete cosine transform (DCT) coefficients of $\bar{z}(x, y)$. Since the partial derivatives of the basis functions, $\phi_x(x, y; \mathbf{w})$ and $\phi_y(x, y; \mathbf{w})$, are integrable, the partial derivatives of $\bar{z}(x, y)$ are guaranteed to be integrable as well; that is $\bar{z}_{xy}(x, y) = \bar{z}_{yx}(x, y)$.

Note that the partial derivatives of $\bar{z}(x, y)$ can also be expressed in terms of this expansion, giving

$$\bar{z}_x(x, y; \bar{c}(\mathbf{w})) = \sum \bar{c}(\mathbf{w})\phi_x(x, y; \mathbf{w}) \tag{9}$$

and

$$\bar{z}_y(x, y; \bar{c}(\mathbf{w})) = \sum \bar{c}(\mathbf{w})\phi_y(x, y; \mathbf{w}). \tag{10}$$

Since the partial derivatives of the basis functions, $\phi_x(x, y; \mathbf{w})$ and $\phi_y(x, y; \mathbf{w})$, are integrable and the expansions of $\bar{z}_x(x, y)$ and $\bar{z}_y(x, y)$ share the same coefficients $\bar{c}(\mathbf{w})$, it is easy to see that $\bar{z}_{xy}(x, y) = \bar{z}_{yx}(x, y)$.

Suppose, now, we have the possibly non-integrable estimate B^* from which we can easily deduce from Equation 1 the possibly non-integrable partial derivatives $z^*_x(x, y)$ and $z^*_y(x, y)$. These partial derivatives can also be expressed as a series, giving

$$z^*_x(x, y; c^*_1(\mathbf{w})) = \sum c^*_1(\mathbf{w})\phi_x(x, y; \mathbf{w}) \tag{11}$$

and

$$z^*_y(x, y; c^*_2(\mathbf{w})) = \sum c^*_2(\mathbf{w})\phi_y(x, y; \mathbf{w}). \tag{12}$$

Note that in general $c^*_1(\mathbf{w}) \neq c^*_2(\mathbf{w})$, which implies that $z^*_{xy}(x, y) \neq z^*_{yx}(x, y)$.

Let us assume that $z_x^*(x,y)$ and $z_y^*(x,y)$ are known from an estimate of B^* and we would like to find $\bar{z}_x(x,y)$ and $\bar{z}_y(x,y)$ (a set of integrable partial derivatives) which are as close as possible to $z_x^*(x,y)$ and $z_y^*(x,y)$, respectively, in a least-squares sense. The goal is to solve the following:

$$\min_{\bar{c}} \sum_{x,y} \quad (\bar{z}_x(x,y;\bar{c}) - z_x^*(x,y;c_1^*))^2 + \\ (\bar{z}_y(x,y;\bar{c}) - z_y^*(x,y;c_2^*))^2. \tag{13}$$

In other words, we take a set of possibly non-integrable partial derivatives, $z_x^*(x,y)$ and $z_y^*(x,y)$, and "enforce" integrability by finding the least-squares fit of integrable partial derivatives $\bar{z}_x(x,y)$ and $\bar{z}_y(x,y)$. Notice that to get the GBR transformed surface $\bar{z}(x,y)$, we need only perform the inverse 2-D DCT on the coefficients $\bar{c}(\mathbf{w})$.

The above procedure is incorporated into the following algorithm. Recall that the data matrix for k images of an individual is defined as $X = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$. If there were no shadowing, X would be rank 3 [61] (assuming no image noise), and we could use SVD to factorize X into $X = B^*S$, where S is a $3 \times k$ matrix whose columns, $\mathbf{s}_i$, are the light source directions scaled by their corresponding source intensities for all k images.

Since the images have shadows (both cast and attached), and possibly saturations, we first have to determine which data values do not satisfy the Lambertian assumption. Unlike saturations, which can be simply determined, finding shadows is more involved. In our implementation, a pixel is labeled as being in shadow if its value divided by its corresponding albedo is below a threshold. As an initial estimate of the albedo, we use the average of the modeling (or training) images. A conservative threshold is then chosen to determine shadows, making it almost certain that no invalid data is included in the estimation process, at the small expense of throwing away a few valid measurements. Any invalid data (both shadows and saturations) are treated as missing measurements by the following estimation method:

1. Find the average of the modeling (or training) images and use it as an initial estimate of the albedo, $\alpha(x,y)$.
2. Without doing any row or column permutations, sift out all the full rows (with no missing measurements) of matrix X to form a full sub-matrix $\tilde{X}$. The number of rows in $\tilde{X}$ is almost always larger than its number of columns, k.
3. Perform SVD on $\tilde{X}$ to find an initial estimate of matrix $S \in \mathbb{R}^{3\times k}$ which best spans the row space of $\tilde{X}$.

4. Find the vectors $\mathbf{b}_j^*$ (the rows of B^*) by performing the minimization in Equation 7, and by using the elements of matrix X for the values of x_{ij} and the columns of the S matrix for the values of $\mathbf{s}_i$. The S matrix is fixed to its current estimate.
5. Estimate a possibly non-integrable set of partial derivatives $z_x^*(x, y)$ and $z_y^*(x, y)$ by using the rows of B^* for the values of $\mathbf{b}(x, y)$ in Equation 1. The value of $\alpha(x, y)$ is fixed to its current estimate.
6. Estimate (as functions of $\bar{c}(\mathbf{w})$) a set of integrable partial derivatives $\bar{z}_x(x, y)$ and $\bar{z}_y(x, y)$ by minimizing the cost functional in Equation 13. (For more details on how to perform this minimization see [16].)
7. Update the albedo $\alpha(x, y)$ by least-squares minimization using the previously estimated matrix S and the partial derivatives $\bar{z}_x(x, y)$ and $\bar{z}_y(x, y)$.
8. Construct $\bar{B}$ by using the newly calculated albedo $\alpha(x, y)$ and the partial derivatives $\bar{z}_x(x, y)$ and $\bar{z}_y(x, y)$ in Equation 1.
9. Update each of the light source directions and strengths $\mathbf{s}_i$ independently using least-squares minimization and the newly constructed $\bar{B}$.
10. Repeat steps 4-9 until the estimates converge.
11. Perform inverse DCT on the coefficients $\bar{c}(\mathbf{w})$ to get the GBR surface $\bar{z}(x, y)$.

In our experiments, the algorithm is well behaved, provided the input data is well conditioned, and converges within 10-15 iterations. With 7 training images of size 192×168 pixels, the algorithm took 3-4 minutes to converge on a Pentium II with a 300 MHz processor and 384MB of memory.

Figure 5 demonstrates the process for constructing the illumination cone. Figure 5.a shows the 7 original single light-source images of a face used in the estimation of $\bar{B}$. Note that the light source in each image moves only a small amount (up to $\pm 12^o$ in either direction) about the viewing axis. Despite this, the images do exhibit some shadowing, e.g., left and right of the nose. In fact, there is a tradeoff in the image acquisition process: the smaller the motion of the light source, meaning fewer shadows present in the images, the worse the conditioning of the estimation problem. On the other hand, if the light source moves excessively, the conditioning improves, however, more extensive shadowing can increase the possibility of having too few (less than three) valid measurements for some parts of the face. Therefore, the light source should move in moderation as it did in the images shown in Figure 5.a.

Figure 5.b shows the basis images of the estimated matrix $\bar{B}$. These basis images encode not only the albedo (reflectance) of the face but also its surface normal field. They can be used to construct images of the face under arbitrary and quite extreme illumination conditions. Figure 5.c shows the reconstructed surface of the face $\bar{z}(x, y)$ up to a GBR transformation. On the left, the surface was rendered with flat shading; on the right, the first basis image of $\bar{B}$ shown in Figure 5.b was texture-mapped on the surface.

Figure 5.d shows images of the face generated using the image formation model in Equation 2 which has been extended to account for cast shadows. To determine cast shadows, we employ ray-tracing that uses the reconstructed GBR surface of the face $\bar{z}(x, y)$. Specifically, a point on the surface is in cast shadow if, for a given light source direction, a ray emanating from that point parallel to the light source direction intersects the surface at some other point. With this extended image formation model, the generated images exhibit realistic shading and have strong attached and cast shadows, unlike the images in Figure 5.a.

2.3 Image Synthesis Under Differing Lighting and Pose

The reconstructed surface and the illumination cones can be combined to synthesize novel images of an object under differing lighting and pose. However, one complication arises because of the GBR ambiguity, in that, the surface and albedo can only be reconstructed up to a 3-parameter GBR transformation. Even though shadows are preserved under a GBR transformation [2], without resolution of this ambiguity, images with non-frontal viewpoints synthesized from a GBR reconstruction will differ from a valid image by an affine warp of image coordinates. (It is affine because the GBR is a 3-D affine transformation and the weak perspective imaging model assumed here is linear.) Since the affine warp is an image transformation, one could perform recognition over variations in viewing direction and affine image transformations. We, instead, resolve the GBR ambiguity to obtain a Euclidean reconstruction using class-specific information.

In our experiments with faces, we use prior knowledge about the shape of faces to resolve the 3 parameters of the GBR ambiguity, namely the scale, the slant, and the tilt of the surface. We take advantage of the left-to-right symmetry of faces (correcting for tilt); we exploit the fact that the forehead and chin of a face are at about the same height (correcting for slant); and we require that the range of height of the surface is about twice the distance between the eyes (correcting for scale). (Another possible method is to use a set of 3-D "eigenheads" to describe the subspace of

typical head shapes in the space of all 3-D shapes [49], and then find the three GBR parameters which minimize the distance of the face reconstruction to this subspace.) Once the GBR parameters are resolved, it is a simple matter using ray-tracing techniques to render synthetic images under variable lighting and pose. Figure 6 shows the shape and albedo reconstructions for the 10 individuals shown in Figure 3. These reconstructions are used in the image synthesis for creating the face representations proposed in Section 3 and used in the experiments reported on in Section 4.

Figure 7 shows synthetic images of a face under novel pose and lighting. Note that these images were generated from the seven images in Figure 5.a where the pose is fixed and where there are only small, unknown variations in illumination. In contrast, the synthetic images exhibit not only large variations in pose but also a wide range in shading and shadowing. The simulated point light source in the images is fixed, therefore, as the face moves around and its gaze direction changes with respect to the light source direction, the shading of the surface changes and both attached and cast shadows are formed, as one would expect.

Figure 8 demonstrates in a more direct way the ability of our generative model to synthesize images under large variations in illumination and pose. The synthesized images are of the same individual shown in Figure 4 with the same illumination and viewpoints. Note that all the images in Figure 8 were generated from the seven images in frontal pose shown in Figure 5.a. Yet, the variability in the synthesized images is as rich as in the original (captured) images shown in Figure 4. This means that synthesized images can be used to form representations of faces useful for recognition under variable lighting and pose.

3 Representations and Algorithms for Face Recognition

While the set of images of a face in fixed pose and under all lighting conditions is a convex cone, there does not appear to be a similar geometric structure in the image space for the variability due to pose. We choose to systematically sample the pose space in order to generate a face representation, $\mathcal{R}_f$. For every sample pose $p \in \{1, \ldots, P\}$ of the face, we generate its illumination cone $\mathcal{C}_p$, and the union of all the cones forms its representation $\mathcal{R}_f = \bigcup_{p=1}^{P} \mathcal{C}_p$, where P is the total number of sample poses. In other words, each face f is represented by a collection of synthesized illumination cones, one for each pose.

However, as noted in Section 2.2, the number of independent surface normals m in matrix B can be large (more than a thousand), hence the number of synthesized extreme rays (images)

needed to completely define the illumination cone for a particular pose can run into the millions. Furthermore, the pose space is six-dimensional, so the complexity of a face representation consisting of one cone per pose can be very large. We therefore use several approximations to make the face representations computationally tractable.

As a first step, we use only a small number of synthesized extreme rays (images) to create each cone, i.e., cones are sub-sampled. The hope is that a sub-sampled cone will provide an approximation that causes a negligible decrease in the recognition performance; in our experiments about 80-120 synthesized images were sufficient, provided that the corresponding light source directions $\mathbf{s}_{ij}$ (from Equation 5) were more or less uniform on the illumination sphere. The resulting cone $\hat{\mathcal{C}}_p$ is a subset of $\mathcal{C}_p$, the true cone of the face in a particular pose. An alternative approximation to $\mathcal{C}_p$ can be obtained by directly sampling the space of light source directions rather than using Equation 5. Again 80-120 source directions were sufficient. While the resulting images from the alternative approximation form the extreme rays of the representation $\hat{\mathcal{C}}_p$ and lie on the boundary of $\mathcal{C}_p$, they are not necessarily extreme rays of $\mathcal{C}_p$. Nevertheless, like before $\hat{\mathcal{C}}_p$ is a subset of $\mathcal{C}_p$.

Another simplifying factor which can reduce the size of the representations is the assumption of a weak perspective imaging model. Under this model, the effect of pose variation can be decoupled into that due to image plane translation, rotation, and scaling (a similarity transformation), and that due to the viewpoint direction. Within a face recognition system, the face detection process generally provides estimates for the image plane transformations. Neglecting the effects of occlusion or appearance of surface points, the variation due to viewpoint can be seen as a non-linear warp of the image coordinates with only two degrees of freedom (one for azimuth and one for elevation). Therefore, in our method, representations of faces contain only variations in illumination and viewpoint; the search over planar transformations is performed during testing. In the recognition experiments described in Section 4, a cone was constructed for every sample of the viewing sphere at 4° intervals over elevation from -24° to $+24^\circ$ and over azimuth from -4° to $+28^\circ$ about the frontal axis. Hence, a face representation consisted a total of ($P = 17 \times 9 =$) 117 sub-sampled illumination cones—one for each sampled viewpoint. The illumination cones for non-frontal viewpoints can be constructed by applying an image warp on the extreme rays defining the frontal illumination cone. This image warp is done in a manner dictated by the 3-D rigid transformations of the reconstructed surface geometry of each face. Finally, the extreme rays in all pose-specific illumination cones are masked using the binary mask shown in Figure 12.

Yet, recognition using a representation consisting of a collection of sub-sampled illumination cones can still be too costly since computing distance to a cone is $O(n\,e^2)$, where n is the number of pixels and e is the number of extreme rays (images). From an empirical study, it was conjectured in [1] that the cone $\hat{\mathcal{C}}_p$ for typical objects is flat (i.e., all points lie near a low-dimensional linear subspace), and this was confirmed for faces in [15]. Hence, for computational efficiency, we perform dimensionality reduction on each sub-sampled cone in the representation of a face. In other words, we model a face in fixed pose but over all lighting conditions by a low-dimensional linear subspace $\hat{\mathcal{I}}_p$ which approximates the sub-sampled cone $\hat{\mathcal{C}}_p$. In our experiments, we chose each subspace $\hat{\mathcal{I}}_p$ to be 11-D since this captured over 99% of the variation in the sample extreme rays of its corresponding cone $\hat{\mathcal{C}}_p$. With this approximation, the representation of a face is then defined as $\mathcal{R}_f = \bigcup_{p=1}^{P} \hat{\mathcal{I}}_p$.

As a final speed-up, the whole face representation $\mathcal{R}_f$ is projected down to $\mathcal{D}_f$, a low dimensional subspace. The basis vectors of this subspace, which is specific to face f, are computed by performing SVD on the 117×11 basis images of the subspaces $\hat{\mathcal{I}}_p$, where those basis images have been scaled by their corresponding singular values from the previous dimensionality reduction. We intended this as an approximation to finding the basis vectors of $\mathcal{D}_f$ by performing SVD directly on all the synthesized images of the face in the collection of sub-sampled illumination cones. In the experiments described in Section 4, each face-specific subspace $\mathcal{D}_f$ had a dimension of 100 as this was enough to capture over 99% of the variability in the 117 illumination cones.

In summary, each face f is represented by a union of the (projected) linear subspaces $\hat{\mathcal{I}}_p$, $p \in \{1, 2, \ldots, 117\}$, within subspace $\mathcal{D}_f$. Recognition of a test image $\mathbf{x}$ is performed by first normalizing $\mathbf{x}$ to unit length and then computing the distance to the representation of each face f in the database. This distance is defined as the Euclidean distance to $\mathcal{D}_f$ plus the Euclidean distance to the closest projected subspace $\hat{\mathcal{I}}_p$ within $\mathcal{D}_f$. The image $\mathbf{x}$ is then assigned the identity of the closest representation.

For three test images of a face in three different poses, Figure 9 shows the closest images in the representation for that individual. Note that these images are not explicitly stored or directly synthesized by the generative model, but instead lie within the closest matching linear subspace. This figure qualitatively demonstrates how well the union of linear subspaces within $\mathcal{D}_f$ approximates the union of the original illumination cones.

4 Recognition Results

We have performed two sets of experiments. In either set, all methods, including ours, were trained on seven acquired images per face in frontal pose. (These experiments were also performed with 19 training images per face, and the results were reported on in [19].)

In the first set, tests were performed under variable illumination but fixed pose, and the goal was, first, to compare the illumination cones representation with three other popular methods, and second, to test the accuracy of the subspace approximation of illumination cones. The second set of experiments was performed under variable illumination *and* pose. The primary goal of this set was to test the performance of the face representation proposed in Section 3. Secondarily, affine image transformations are often used to model modest variations in viewpoint, and so another goal was to determine when this would break down for faces; i.e., recognition methods using affine image transformations were trained on frontal-pose images, but tested on non-frontal images. As demonstrated, our proposed face representations more effectively handle both large variations in lighting and viewpoint.

The rest of this section is divided into three parts. In the following section, we describe the face image database used in our experiments. In Section 4.2, we present the first set of experiments under variable illumination but fixed pose, while in Section 4.3, we describe the second set under variable illumination and pose.

4.1 Face Image Database

The experimentation reported on here was performed on the Yale Face Database B. To capture the images in this database, we have constructed the geodesic lighting rig shown in Figure 10 with 64 computer controlled xenon strobes whose positions in spherical coordinates are shown in Figure 11. With this rig, we can modify the illumination at frame rate and capture images under variable illumination and pose. Images of ten individuals (shown in Figure 3) were acquired under 64 different lighting conditions in 9 poses (a frontal pose, five poses at 12°, and three poses at 24° from the camera axis). The 64 images of a face in a particular pose are acquired in about 2 seconds. Therefore, there is only minimal change in head position and facial expression in those 64 images.

Of the 64 images per person in each pose, 45 were used in our experiments, for a total of 4050 images (9 poses $\times$ 45 illumination conditions $\times$ 10 faces). The images from each pose were divided into 4 subsets (12°, 25°, 50°, and 77°) according to the angle the light source direction

makes with the camera's axis; see Figure 4. Subset 1 (respectively 2, 3, 4) contains 70 (respectively 120, 120, 140) images per pose.

The original size of the images is 640×480 pixels. In our experiments, all images were manually cropped to include only the face with as little hair and background as possible. The images from the frontal pose were aligned (scaled and rotated) so that the eyes in each image fell on the same positions lying on a horizontal line. This alignment was performed in order to remove any bias from the recognition results due to the association of a particular scale, position, or orientation to a particular face. Only the frontal-pose images were aligned because these were the only ones used for training purposes in all methods, including ours. The images in the other 8 poses were only loosely cropped—the position of each face could vary by as much as $\pm 5\%$ of the width of the cropped window, while the scale could vary between 95% to 105% from the average. All cropped images, used for both training and testing, were finally sub-sampled by 4 down to a resolution of 36×42 pixels, and then masked using the binary mask shown in Figure 12.

4.2 Extrapolation in Illumination

The first set of recognition experiments was performed under fixed pose using 450 images (45 per face) for both training and testing. This experimental framework, where only illumination varies while pose is kept fixed, was designed to compare three other recognition methods to the illumination cone representation. Another goal of these experiments was to test the accuracy of the subspace approximation of illumination cones.

From a set of face images labeled with the person's identity (*the training set*) and an unlabeled set of face images from the same group of people (*the test set*), each algorithm was used to identify the person in the test images. For more details about the comparison algorithms, see [3] and [20]. Here, we only present short descriptions:

Correlation: The simplest recognition scheme is a nearest neighbor classifier in the image space [6]. An image in the test set is recognized (classified) by assigning to it the label of the closest point in the learning set, where distances are measured in the image space. When all of the images are normalized to have zero mean and unit variance, this procedure is also known as Correlation. As correlation techniques are computationally expensive and require great amounts of storage, it is natural to pursue dimensionality reduction schemes.

Eigenfaces: A technique commonly used in computer vision—particularly in face recognition

—is principal components analysis (PCA), which is popularly known as *Eigenfaces* [23, 36, 47, 66]. Given a collection of training images $\mathbf{x}_i \in \mathbb{R}^n$, a linear projection of each image $\mathbf{y}_i = W\mathbf{x}_i$ to a l-dimensional feature space is performed, where the projection matrix $W \in \mathbb{R}^{l \times n}$ is chosen to maximize the scatter of all projected samples. A face in a test image $\mathbf{x}$ is recognized by projecting $\mathbf{x}$ into the feature space, followed by nearest neighbor classification in $\mathbb{R}^l$. One proposed method for handling illumination variation in PCA is to discard from W the three most significant principal components; in practice, this yields better recognition performance [3]. In Eigenfaces, like Correlation, the images were normalized to have zero mean and unit variance, as this improved its performance. This also made the results independent of light source intensity. In our implementations of Eigenfaces, the dimensionality of the feature space was chosen to be 20, that is, we used 20 principal components. (Recall that performance approaches correlation as the dimensionality of the feature space is increased [3, 47].) Error rates are also presented when principal components four through twenty-three were used.

Linear Subspace: A third approach is to model the illumination variation of each face with the three-dimensional linear subspace $\mathcal{L}$ described in Section 2.1. To perform recognition, we simply compute the distance of the test image to each linear subspace $\mathcal{L}$ and choose the face identity corresponding to the shortest distance. We call this recognition scheme the *Linear Subspace* method [2]; it is a variant of the photometric alignment method proposed in [62] and is related to [24, 48]. While this method models the variation in shading when the surface is completely illuminated, it does not model shadowing.

Illumination Cones: Finally, recognition is performed using the illumination cone representation. We have used 121 generated images (extreme rays) to form the illumination cone for each face. The illumination sphere was sampled at 15° intervals in azimuth and elevation from -75° to 75° in both directions, hence a total of ($11 \times 11 =$) 121 sample light-source directions. We have tested on three variations of the illumination cones:

1. **Cones-attached**: The cone representation was constructed without cast shadows, so the 121 extreme rays were generated directly from Equation 4. That is, the representation contained only attached shadows along with shading.
2. **Cones-cast**: The representation was constructed as described in Section 2.2 where we employed ray-tracing and used the reconstructed surface of the face to determine cast shadows.
3. **Cones-cast Subspace Approximation**: The illumination cone of each face with cast shad-

ows $\hat{\mathcal{C}}_p$ is approximated by an 11-D linear subspace $\hat{\mathcal{I}}_p$. It was empirically determined [1, 19] that 11 dimensions capture over 99% of the variance in the sample extreme rays (images). The basis vectors for this subspace are determined by performing SVD on the 121 extreme rays in $\hat{\mathcal{C}}_p$ and then selecting the 11 eigenvectors associated with the largest singular values.

In the first two variations of the illumination cone representation, recognition is performed by computing the distance of the test image to each cone and then choosing the face identity corresponding to the shortest distance. Since each cone is convex, the distance can be found by solving a convex optimization problem (see [20]). A modified version of the Matlab non-negative linear least-squares (`nnls`) function was used. This modified algorithm has a computational complexity of $O(n\,e^2)$, where n is the number of pixels and e is the number of extreme rays. In the third variation, recognition is performed by computing the distance between the test image and each linear subspace, and then choosing the face corresponding to the shortest distance. Using the cone subspace approximation method significantly reduces the computational time and storage (as compared to the original illumination cone method). Since the basis vectors of each subspace are orthogonal, the computational complexity of using the subspace approximation is only $O(n\,q)$, where n is the number of pixels and q is the number of basis vectors (11 in this case).

Similar to the extrapolation experiments described in [3], each method was trained on images from Subset 1 (seven images per face in frontal pose with near-frontal illumination), and then tested on all 450 images from the frontal pose. Figure 13 shows the results from these experiments. (This test was also performed on the Harvard Robotics Lab face database [23, 24] and was reported on in [20].) Notice that the cone subspace approximation performed as well as the original illumination cone representation with no mistakes in 450 images. This supports the use of low-dimensional subspaces to approximate the full illumination cones in the face representation described in Section 3.

4.3 Recognition Under Variable Lighting and Pose

Next, we conducted recognition experiments under variable pose and illumination using images from all nine poses in the database. The primary goal of these experiments were to test the performance of the face representation proposed in Section 3. Secondarily, affine image transformations are often used to model modest variations in viewpoint, and so another goal was to determine when this would break down for faces.

Five recognition methods (in part different from those in the previous section) were compared on 4050 images. Each method was trained on seven images per face with near-frontal illumination and frontal pose, and then tested on all images from all nine poses—an extrapolation in both pose and illumination.

Like before, we have used 121 generated images to form the illumination cone $\hat{\mathcal{C}}_p$ of a face in a particular viewpoint. The illumination sphere was sampled at 15° intervals in azimuth and elevation from -75° to 75° in both directions, hence a total of ($11 \times 11 =$) 121 sample light-source directions. Furthermore, each cone was approximated by an 11-D linear subspace, $\hat{\mathcal{I}}_p$. The basis vectors for this subspace were determined by performing SVD on the 121 extreme rays in $\hat{\mathcal{C}}_p$ and then selecting the 11 eigenvectors associated with the largest singular values.

As mentioned in Section 3, the effect of pose variation can be decoupled into that due to image plane translation, rotation, and scaling, and that due to viewpoint. While variations due to viewpoint have been incorporated into the face representations introduced in Section 3, the search over planar transformations is performed during testing. Note that in the implementation of the last three methods presented here, the planar transformations did not include image rotations (only translations and scaling). This was done to reduce computational time. The search in translations was in both directions from -6 to $+6$ pixels in 2-pixel increments (at the 42×36 resolution). This is equivalent to -24 to $+24$ pixels in 8-pixel increments in the original resolution. The search in scale was from 0.92 to 1.04 at increments of 0.04.

The five methods are:

1. **Correlation**: (As described in the previous Section.)
2. **Cones-cast Subspace Approximation**: Each face is represented by an 11-D subspace approximation of the cone (with cast shadows) corresponding to the frontal pose. In this method, no effort was made to accommodate for pose during recognition, not even a search in image plane transformations.
3. **Correlation + planar transformations**: This is like the first method except that during recognition a search in planar transformations was allowed.
4. **Cones-cast Subspace Approximation + planar transformations**: This is an extension of the second method where recognition is also performed over a search in planar transformations.

5. **Face Representation proposed in Section 3**: Each face is represented by the union of 117, projected, 11-D linear subspaces $\hat{\mathcal{I}}_p$, $p \in \{1, 2, \ldots, 117\}$, within the 100-D subspace $\mathcal{D}_f$. Each 11-D subspace approximates its corresponding illumination cone which models the variation due to illumination at each sampled viewpoint. Recognition of a test image $\mathbf{x}$ is performed by computing the distance to the representation of each face in the database. This distance is defined as the Euclidean distance to $\mathcal{D}_f$ plus the Euclidean distance to the closest projected subspace $\hat{\mathcal{I}}_p$ within $\mathcal{D}_f$. The image $\mathbf{x}$ is then assigned the identity of the closest representation. Note that as with the third and fourth methods, recognition is performed over variations in planar transformations. In our tests, it takes about 25 seconds to recognize an image with 10 faces in the database using a Pentium II with a 300 MHz processor and 384MB of main memory.

The recognition results are shown in Figure 14. Note that each reported error rate is for *all* illumination Subsets (1 through 4). Figure 15, on the other hand, shows the break-down of the results of the last method (using our proposed face representations) for different poses against variable illumination.

As shown in Figure 14, the method of Cones-cast Subspace Approximation with planar transformations performs reasonably well for poses up to 12° from the viewing axis, but breaks down when the viewpoint becomes more extreme. This demonstrates the need for representations, like the one introduced here, that explicitly capture large variations in viewpoint (i.e., out-of-plane rotations) as well as illumination. Figure 14 shows that our proposed representations more effectively handle image variability due to large changes in lighting and viewpoint. This is in spite of the fact that they were created using only a handful of training images per face in frontal pose and with small, unknown changes in the lighting direction and intensity. Figure 15 shows that these face representations perform almost without error for all poses, except on the most extreme lighting directions.

Note that in the frontal pose there was a marginal increase in the error rates for the cones approximation method when, at first, it was extended to include planar transformations and, then, when variations in viewpoint were allowed. Remember that for test images with faces in frontal pose, the cone subspace approximation without any pose variations (both viewpoint and planar transformations) is an adequate model of the image variability. Hence, any additional degrees

of freedom, such as image-plane transformations, or variations in viewpoint, may provide more ways for a mismatch. Nevertheless, the increase in errors is insignificant compared to the gain in performance when the viewpoint in a test image is non-frontal.

5 Discussion

We draw the following conclusions from the experimental results:

- A small number of images of a face in fixed pose and illuminated by a single point light source at unknown positions can provide enough information to generate a rich representation of the face useful for recognition under variable pose and illumination. Figure 15 demonstrates the effectiveness of our representation in face recognition on a database of 4050 images of 10 individuals viewed under large variations in pose and illumination.
- Figure 14 specifically shows the effectiveness of our representation in recognizing faces under large variations in viewpoint. This is because it *explicitly* captures the image variability due to viewpoint. Even though methods that only allow planar transformations can perform reasonably well for viewpoints up to 12° from the camera axis, they break down when the viewpoint becomes more extreme.
- In the experiments under variable illumination but fixed pose, the illumination cone method outperforms all other methods, as shown in Figure 13. In fact, both Correlation and Eigenfaces methods break down under extreme illumination conditions.
- Including cast shadows in the illumination cones improves recognition rates. See Figure 13.
- Since the illumination cone of an object lies near a low-dimensional subspace in the image space, the images of a face under variable illumination (but fixed pose) can be well approximated by a low-dimensional subspace. Figure 13 demonstrates the effectiveness of using low-dimensional subspaces. This agrees with their use in the full representations proposed in Section 3.

The central point of our work was to show that, from a small number of exemplars, it is possible to extrapolate to extreme viewing conditions. Recall that the seven images of the face in Figure 5.a were the only information used to synthesize the 405 images in Figure 8. Not only do these synthesized images contain large variations in lighting and pose, they can also be used in recognition, as demonstrated by the experimental results.

We believe that our method is applicable to the more general problem of object recognition where similar representations could be used. In this paper, we have assumed that faces exhibit Lambertian reflectance which, as demonstrated by the recognition results, is a good approximation. Nevertheless, applying our method to the recognition of other object classes will require to relax the Lambertian assumption allowing for more complex bi-directional reflectance distribution functions (BRDFs).

Future work will concentrate on determining the BRDFs of object surfaces and incorporating them in object representations. Furthermore, we plan to expand our representations to include larger variations in viewpoint. Other exciting research domains include facial expression recognition, aging, and object recognition with occlusions.

Acknowledgments

P. N. Belhumeur and A. S. Georghiades were supported by a Presidential Early Career Award, NSF Career Award IRI-9703134, ARO grant DAAH04-95-1-0494, and National Institute of Health grant R01-EY-12691-01. D. J. Kriegman was supported by NSF under NYI, IRI-9257990, and by ARO grant DAAG55-98-1-0168. The authors would like to thank Alan Yuille and David Jacobs for many useful discussions, Jonas August for his very helpful and careful comments, Saul Nadata for his help in implementing the code for resolving the GBR ambiguity in the surface reconstruction of faces, and Melissa Koudelka and Todd Zickler for reviewing the manuscript.

References

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

[2] P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 1040–1046, 1997.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):711–720, 1997. Special Issue on Face Recognition.

[4] D. Beymer. Face recognition under varying pose. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 756–761, 1994.

[5] D. Beymer and T. Poggio. Face recognition from one example view. In *Int. Conf. on Computer Vision*, pages 500–507, 1995.

[6] R. Brunelli and T. Poggio. Face recognition: Features vs templates. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 15(10):1042–1053, 1993.

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

[8] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 254–261, 2000.

[9] Q. Chen, H. Wu, and M. Yachida. Face detection by fuzzy pattern matching. In *Int. Conf. on Computer Vision*, pages 591–596, 1995.

[10] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *Proc. European Conf. on Computer Vision*, volume 2, pages 484–498, 1998.

[11] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[12] I. Cox, J. Ghosn, and P. Yianilos. Feature-based face recognition using mixture distance. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 209–216, 1996.

[13] I. Craw, D. Tock, and A. Bennet. Finding face features. pages 92–96, 1992.

[14] G. Edwards, T. Cootes, and C. Taylor. Advances in active appearance models. In *Int. Conf. on Computer Vision*, pages 137–142, 1999.

[15] R. Epstein, P. Hallinan, and A. Yuille. 5+/-2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *Physics Based Modeling Workshop in Computer Vision*, Session 4, 1995.

[16] R. T. Frankot and R. Chellapa. A method for enforcing integrabilty in shape from shading algorithms. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 10(4):439–451, 1988.

[17] T. Fromherz. Face recognition: A summary of 1995-1997. International Computer Science Institute ICSI TR-98-027, University of California, Berkeley, 1998.

[18] A. Georghiades, P. Belhumeur, and D. 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, Fort Collins, Colorado, 1999.

[19] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Generative models for recognition under variable pose and illumination. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 277–284, 2000.

[20] A. Georghiades, D. Kriegman, and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 52–59, 1998.

[21] A. Goldstein, L. Harmon, and A. Lesk. Identification of human faces. *Proceedings of the IEEE*, 59(5):748–760, May 1971.

[22] V. Govindaraju. Locating human faces in photographs. *Int. Journal of Computer Vision*, 19(2):129–146, August 1996.

[23] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 995–999, 1994.

[24] P. Hallinan. *A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions*. PhD thesis, Harvard University, 1995.

[25] P. Hallinan, G. Gordon, A. Yuille, and D. Mumford. *Two- and Three-Dimensional Patterns of the Face*. A.K.Peters, 1999.

[26] L. Harmon, M. Kaun, R. Lasch, and P. Ramig. Machine identification of human faces. *Pattern Recognition*, 13(2):97–110, 1981.

[27] L. Harmon, S. Kuo, P. Ramig, and U. Raudkivi. Identification of human face profiles by computer. *Pattern Recognition*, 10:301–312, 1978.

[28] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *J. Opt. Soc. Am. A*, 11(11):3079–3089, Nov. 1994.

[29] B. Horn. *Computer Vision*. MIT Press, Cambridge, Mass., 1986.

[30] F. Huang, Z. Zhou, H. Zhang, and T. Chen. Pose invariant face recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 245–250, 2000.

[31] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and characterizing intensity images. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1997.

[32] P. Juell and R. Marsh. A hierarchical neural-network for human face detection. *Pattern Recognition*, 29(5):781–787, May 1996.

[33] T. Kanade. *Picture Processing by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973.

[34] T. Kanade. *Computer Recognition of Human Faces*. Birkhauser Verlag, Stuttgart, Germany, 1977.

[35] G. Kaufman and K. Breeding. The automatic recognition of human faces from profile silhouettes. *IEEE Trans. on Systems, Man and Cybernetics*, 6:113–121, February 1976.

[36] L. Sirovitch and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Optical Soc. of America A*, 2:519–524, 1987.

[37] J. Lambert. *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae*. Eberhard Klett, 1760.

[38] A. Lanitis, C. Taylor, and T. Cootes. A unified approach to coding and interpreting face images. In *Int. Conf. on Computer Vision*, pages 368–373, 1995.

[39] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):743–756, July 1997.

[40] C. Lee, J. Kim, and K. Park. Automatic human face location in a complex background using motion and color information. *Pattern Recognition*, 29(11):1877–1889, November 1996.

[41] T. Leung, M. Burl, and P. Perona. Finding faces in cluttered scenes using labeled random graph matching. In *Int. Conf. on Computer Vision*, pages 637–644, 1995.

[42] S. Li and J. Lu. Face recognition using nearest feature line. *IEEE Trans. on Neural Networks*, 10(2):439–443, March 1999.

[43] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 300–305, 2000.

[44] Moghaddam and Pentland. Probabilistic visual learning for object detection. In *Int. Conf. on Computer Vision*, pages 786–793, 1995.

[45] B. Moghaddam and A. Pentland. Probabilistic Visual Learning for Object Representation. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):696–710, 1997.

[46] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In *European Conf. on Computer Vision*, pages 286–296, 1994.

[47] H. Murase and S. Nayar. Visual learning and recognition of 3-D objects from appearence. *Int. J. Computer Vision*, 14(5–24), 1995.

[48] S. Nayar and H. Murase. Dimensionality of illumination manifolds in appearance matching. In *Int. Workshop on Object Representations for Computer Vision*, page 165, 1996.

[49] A. O'Toole, T. Vetter, N. Toje, and H. Bulthoff. Sex classification is better with three-dimensional head structure than with texture. *Perception*, 26:75–84, 1997.

[50] A. Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 22:107–119, 2000.

[51] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 84–91, 1994.

[52] P. Phillips, H. Moon, P. Rauss, and S. Risvi. The FERET evaluation methodology for face-recognition algorithms. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 137–143, 1997.

[53] P. Phillips, H. Moon, P. Rauss, and S. Risvi. The FERET September 1996 database evaluation procedure. In *Audio and Video-Based Biometric Person Authentication*, 1997.

[54] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Visual Computing*, 16(5), 1998.

[55] T. Poggio and K. Sung. Example-based learning for view-based human face detection. In *Proc. Image Understanding Workshop*, pages II:843–850, 1994.

[56] T. Riklin-Raviv and A. Shashua. The quotient image: Class based recognition and synthesis under varying illumination conditions. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages II:566–571, 1999.

[57] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 20(1):23–38, January 1998.

[58] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 38–44, 1998.

[59] A. Samil and P. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25:65–77, 1992.

[60] A. Samil and P. Iyengar. Human face detection using silhouettes. *Pattern Recognition and Artificial Intelligence*, 9:845–867, 1995.

[61] A. Shashua. *Geometry and Photometry in 3D Visual Recognition*. PhD thesis, MIT, 1992.

[62] A. Shashua. On photometric issues to feature-based object recognition. *Int. J. Computer Vision*, 21:99–122, 1997.

[63] H. Shum, K. Ikeuchi, and R. Reddy. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 17(9):854–867, September 1995.

[64] W. Silver. *Determining Shape and Reflectance Using Multiple Images*. PhD thesis, MIT, Cambridge, MA, 1980.

[65] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. J. Computer Vision*, 9(2):137–154, 1992.

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

[67] T. Vetter. Synthesis of novel views from a single face image. *Int. Journal of Computer Vision*, 28(2):103–116, June 1998.

[68] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):733–742, July 1997.

[69] L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):775–779, July 1997.

[70] R. Woodham. Analysing images of curved surfaces. *Artificial Intelligence*, 17:117–140, 1981.

[71] M. Yang, N. Ahuja, and D. Kriegman. Mixture of linear subspaces for face detection. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 196–201, 2000.

[72] K. Yow and R. Cipolla. Feature-based human face detection. *Image Visual Computing*, 15(9):713–735, September 1997.

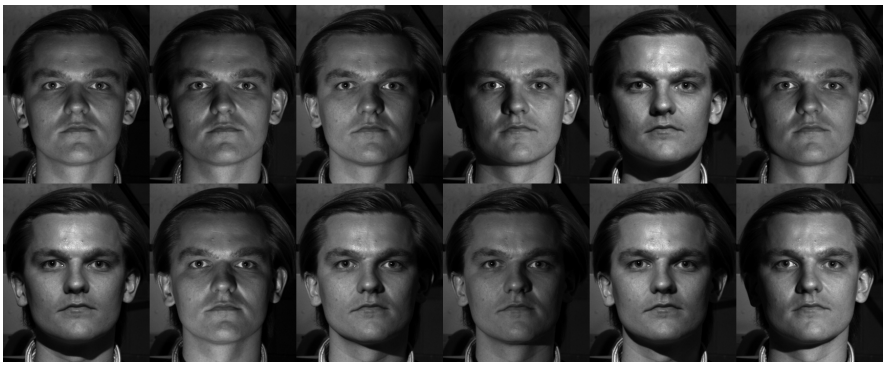
[73] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *Computer Graphics (SIGGRAPH)*, pages 207–218, 1998.

[74] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 158–164, 1997.

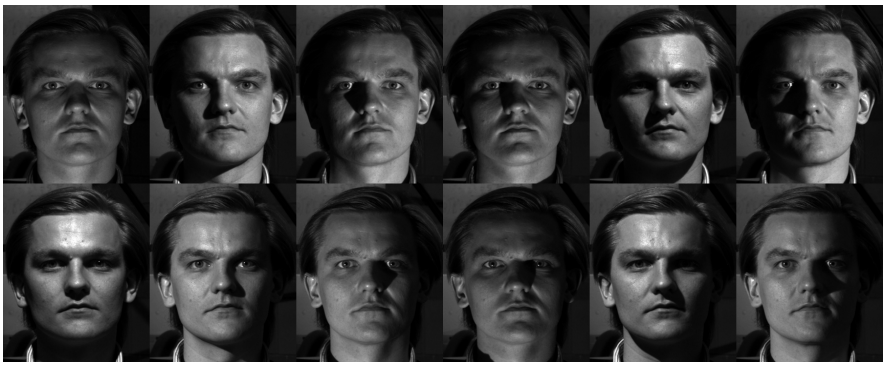
[75] W. Zhao and R. Chellapa. SFS based view synthesis for robust face recognition. In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 285–292, 2000.

[76] W. Zhao, R. Chellappa, and P. Phillips. Subspace linear discriminant analysis for face recognition. Center for Automation Research CAR-TR-914, University of Maryland, College Park, 1999.

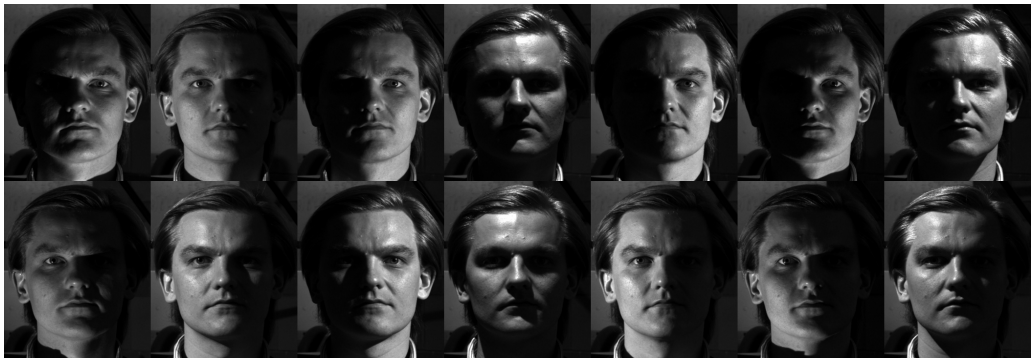
Subset 1.



Subset 2.



Subset 3.



Subset 4.

Figure 1: Example images of a single individual in frontal pose from the Yale Face Database B, showing the variability due to illumination. The images have been divided into four subsets according to the angle the light source direction makes with the camera axis—Subset 1 (up to 12°), Subset 2 (up to 25°), Subset 3 (up to 50°), and Subset 4 (up to 77°).

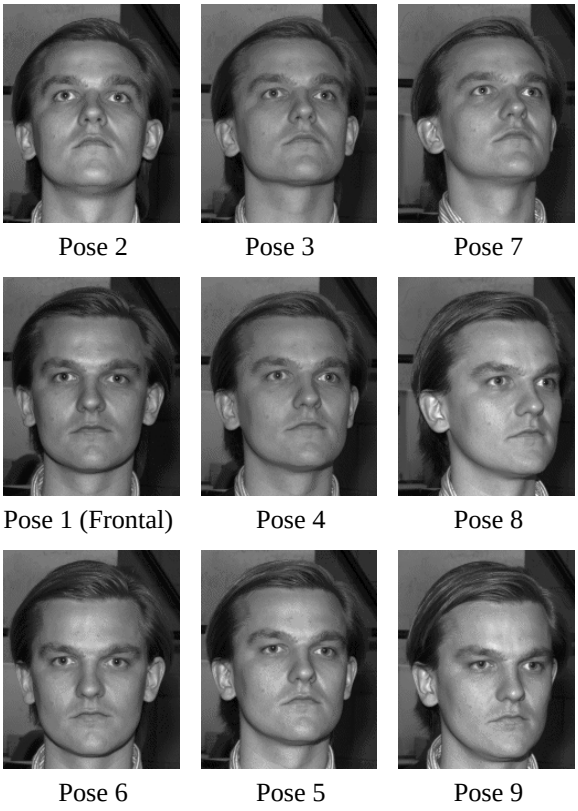


Figure 2: Example images of a single individual, one from each of the nine different poses in the Yale Face Database B.

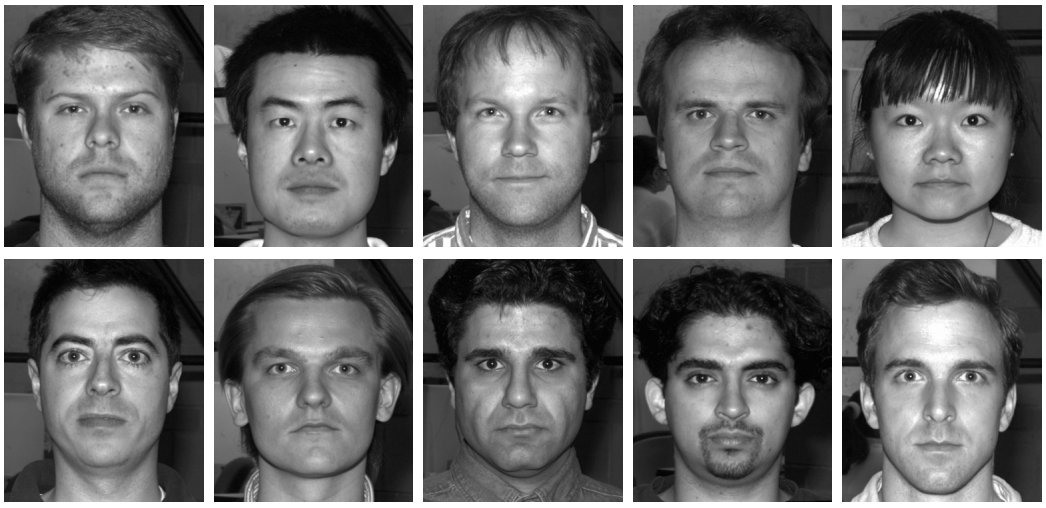


Figure 3: The ten individuals in the Yale Face Database B.

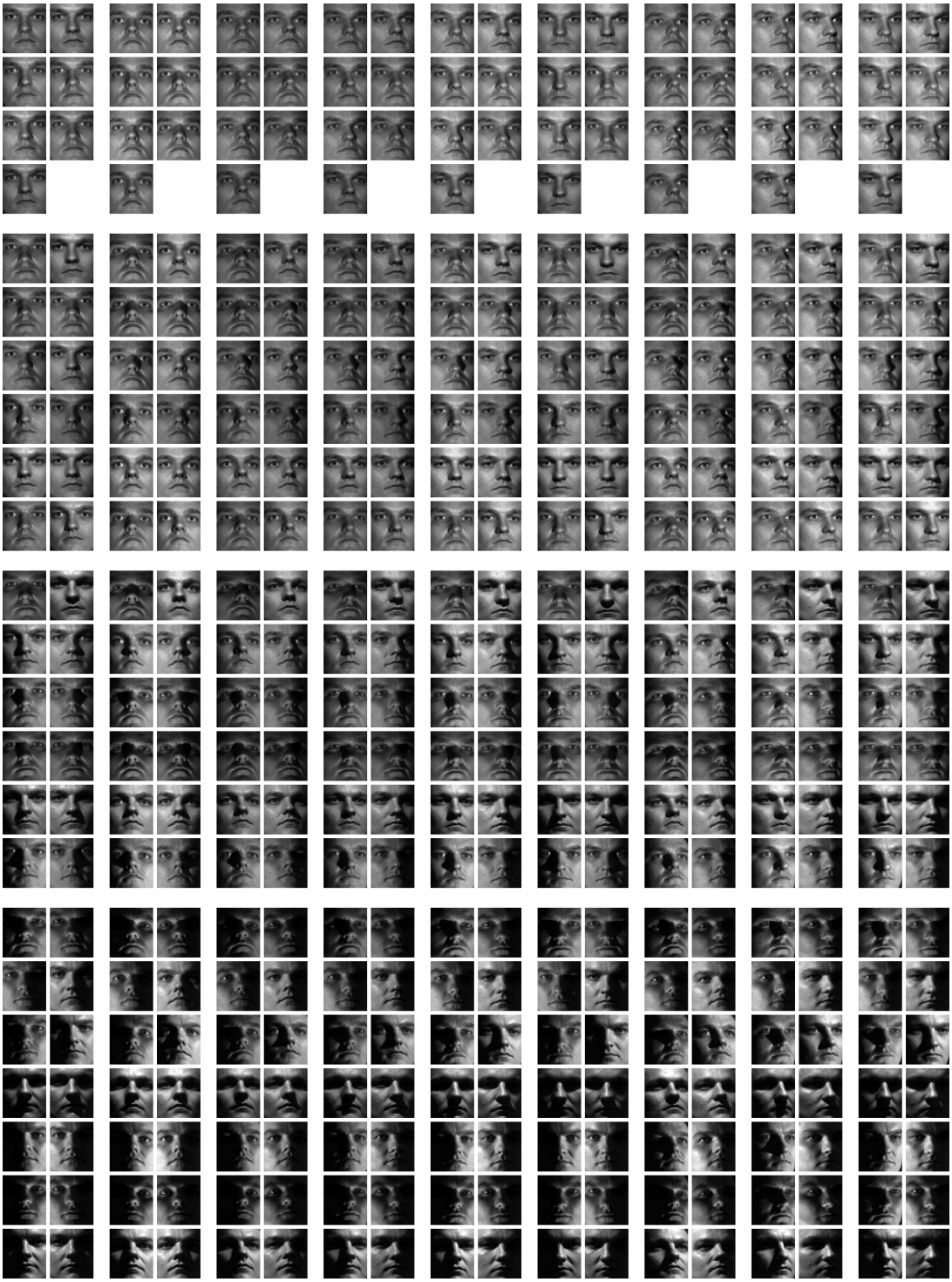
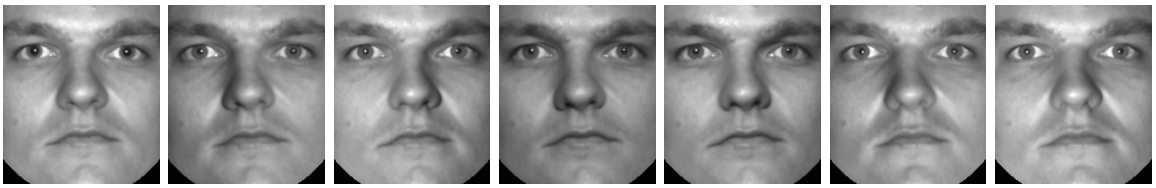


Figure 4: Original (captured) images of a single individual from the Yale Face Database B, showing the variability due to illumination and pose. The images have been divided into four subsets (1 through 4 from top to bottom) according to the angle the light source direction makes with the camera axis. Every pair of columns shows the images of a particular pose (1 through 9 from left to right).



a.

b.

c.

d.

Figure 5: The process of constructing the cone $\hat{\mathcal{C}}$. a. The seven training images from Subset 1 (near frontal illumination) in frontal pose; b. Images corresponding to the columns of $\bar{B}$; c. Reconstruction up to a GBR transformation. On the left, the surface was rendered with flat shading, i.e., the albedo was assumed to be constant across the surface, while on the right the surface was texture-mapped with the first basis image of $\bar{B}$ shown in Figure 5.b; d. Synthesized images from the illumination cone of the face with novel lighting conditions but fixed pose. Note the large variations in shading and shadowing as compared to the seven training images.

Figure 6: The surface reconstructions of the 10 faces shown in Figure 3. These reconstructions are used in the image synthesis for creating the face representations proposed in Section 3 and used in the experiments reported on in Section 4.

Figure 7: Synthesized images under variable pose and lighting generated from the training images shown in Figure 5.a.

Figure 8: Synthesized images of the same individual under the same illumination and viewpoints as in Figure 4. As before, the synthesized images have been divided into four subsets (1 through 4 from top to bottom) according to the angle the light source direction makes with the camera axis. Every pair of columns shows the images from a particular pose (1 through 9 from left to right). Note that all the images were generated from the seven acquired images of Subset 1, Pose 1 shown in Figure 5.a.

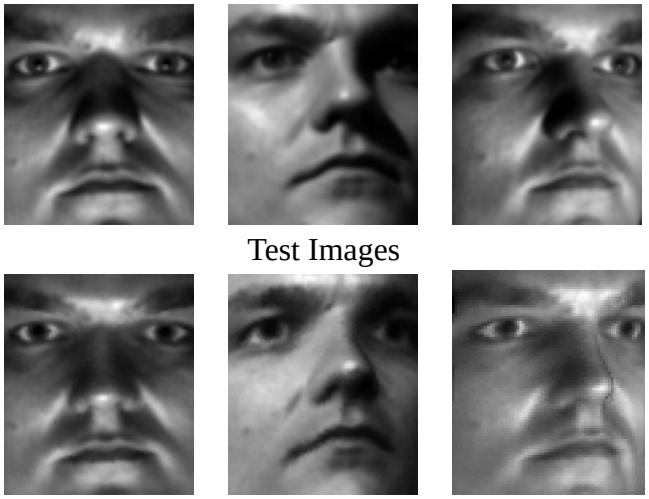


Figure 9: TOP ROW: Three images of a face from the test set. BOTTOM ROW: The closest reconstructed image from the representation proposed in Section 3. Note that these images are not explicitly stored or directly synthesized by the generative model, but instead lie within the closest matching linear subspace.

Figure 10: A geodesic dome with 64 strobes used to gather images under variable illumination and pose.

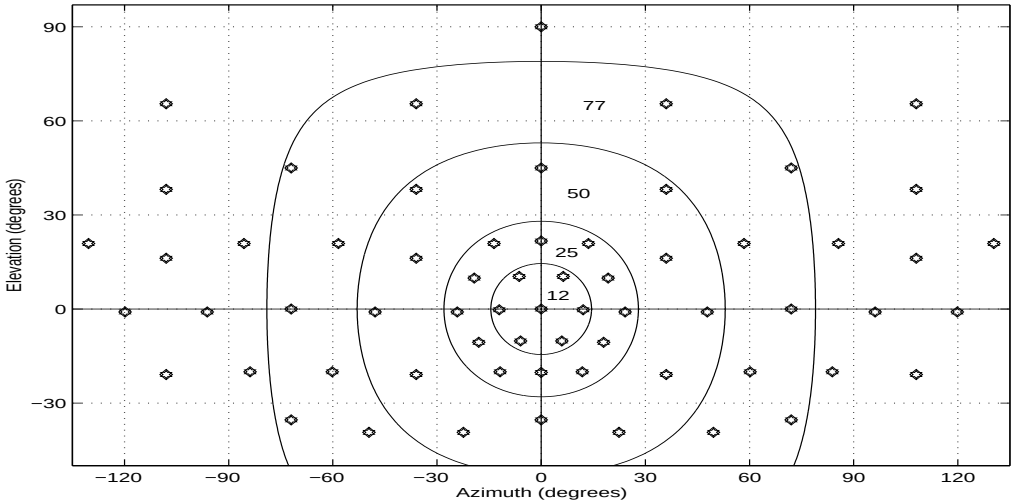


Figure 11: The azimuth and elevation of the 64 strobes. Each annulus contains the positions of the strobes corresponding to the images of each illumination subset—Subset 1 ($12°$), Subset 2 ($25°$), Subset 3 ($50°$), Subset 4 ($77°$). Note that the sparsity at the top of the figure is due to the distortion during the projection from the sphere to the plane.

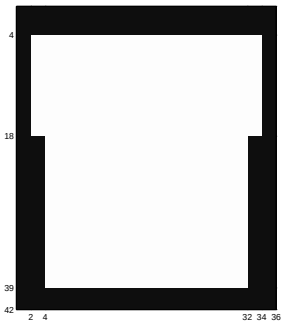
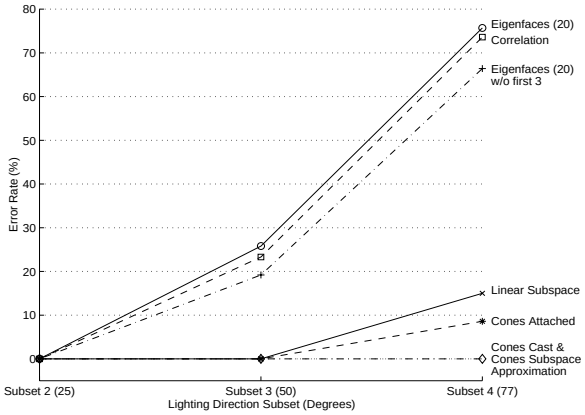
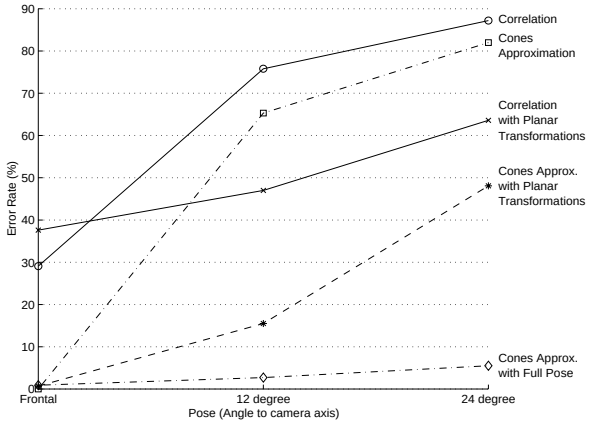


Figure 12: The 42×36 binary mask used in the image normalization process.



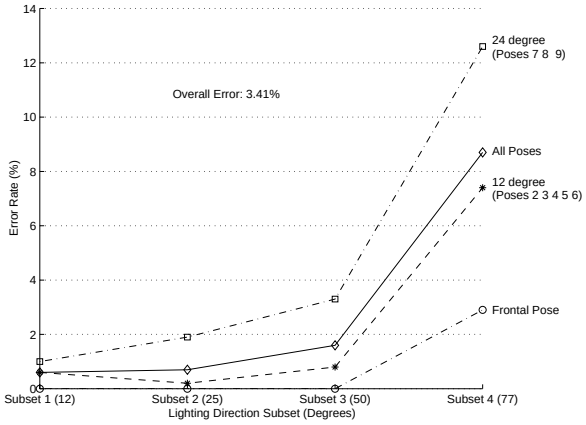
EXTRAPOLATION IN ILLUMINATION			
Method	Error Rate (%) vs. Illum.		
	Subset 2	Subset 3	Subset 4
Correlation	0.0	23.3	73.6
Eigenfaces	0.0	25.8	75.7
Eigenfaces w/o 1st 3	0.0	19.2	66.4
Linear subspace	0.0	0.0	15.0
Cones-attached	0.0	0.0	8.6
Cones-cast (Subspace Approx.)	0.0	0.0	0.0
Cones-cast	0.0	0.0	0.0

Figure 13: **Extrapolation in Illumination:** Each method was trained on seven images per person from Subset 1 (near-frontal illumination), Pose 1 (frontal pose). The plots show the error rates under more extreme lighting conditions while the pose was kept fixed. The experiments were conducted on 450 images from Pose 1.



EXTRAPOLATION IN POSE			
Method	Error Rate (%) vs. Pose		
	Frontal (Pose 1)	12° (Poses 2 3 4 5 6)	24° (Poses 7 8 9)
Correlation	29.1	75.8	87.2
Cones Approximation	0.0	65.3	82.0
Correlation with Planar Transformations	37.6	47.0	63.6
Cones Approx. with Planar Transformations	0.7	15.5	48.1
Cones Approx. with Full Pose	0.9	2.7	5.5

Figure 14: **Extrapolation in Pose:** Error rates as the viewing direction becomes more extreme. The five methods have been trained on seven images per person from Subset 1 (near frontal illumination), Pose 1 (frontal pose). Note that each reported error rate is for *all* illumination subsets (1 through 4). The "Frontal Pose" includes 450 images, the "12 degree" (Poses 2, 3, 4, 5, 6) includes 2250 images, and the "24 degree" (Poses 7, 8, 9) includes 1350 images.



CONES APPROXIMATION WITH FULL POSE				
	Lighting Variation			
Pose	Subset 1	Subset 2	Subset 3	Subset 4
Frontal (Pose 1)	0.0	0.0	0.0	2.9
12° (Poses 2 3 4 5 6)	0.6	0.2	0.8	7.4
24° (Poses 7 8 9)	1.0	1.9	3.3	12.6
All Poses	0.6	0.7	1.6	8.7

Figure 15: Error rates (%) for different poses against variable lighting using our face representations proposed in Section 3. The training was performed on seven frontal-pose images per face with near-frontal illumination. The tests were conducted on a total of 4050 images from all nine poses of the Yale Face Database B.