1999 International Symposium on Robotics Research, Snowbird.

Shape and Enlightenment: Reconstruction and Recognition under Variable Illumination

David J. Kriegman[1], **Peter N. Belhumeur**[2], **and Athinodoros S. Georghiades**[2]

[1]Beckman Institute, Dept. of Computer Science,
University of Illinois Urbana, IL 61801, USA, kriegman@uiuc.edu

[2]Center for Computational Vision & Control, Dept. of Electrical Engineering,
Yale University, New Haven, CT 06520-8267, USA, {belhumeur, athinodoros.georghiades}@yale.edu

Abstract

Due to illumination variability, the same object can appear dramatically different even when viewed in fixed pose, and this variability can confound recognition systems. This paper summarizes recent work on developing appearance-based methods for modeling the variability due to illumination in the images of objects. They differ from past appearance-based methods, however, in that a small set of training images is used to generate a representation – the illumination cone – which models the *complete set of images* of an object with Lambertian reflectance map under an arbitrary combination of point light sources at infinity. From a few images of an object in fixed pose but varying and unknown lighting, a surface and albedo map are reconstructed up to a family of affine (generalized bas-relief or GBR) deformations, and the cone representation is derived from this GBR surface. The methods have been tested within the domain of face recognition on two databases, one with 660 images of 10 faces in fixed pose but variable lighting, and one with 1350 images of 10 faces with variable pose and lighting; the results exceed those of popular existing methods.

1. Introduction

The images of objects and scenes can appear dramatically different even when observed from the same viewpoint because of changes in lighting (See Fig. 1). On the one hand, vision-based robot systems can exploit this naturally occurring variability to better understand scene structure. On the other, without suitable models of the effects of illumination, recognition systems can be confounded by this variation. Most vision systems address this issue by (a) controlling illumination, (b) employing a representation that is invariant to the variability, or (c) directly modeling this variability. For example, there is a long tradition of performing edge detection at an early stage since the presence of an edge at an image location is thought to be largely independent of lighting.

Here, we consider modeling the effects of illumination variability rather than trying to achieve illumination invariance, and show how these models can be exploited for reconstructing the 3-D geometry of scenes and used to significantly increase the performance of appearance-based recognition systems. We demonstrate the use of these models within the context of face recognition, but believe that they have much broader applicability. For example, within industrial and service robotics, it is not always possible to control the illumination since the light cast through windows will vary with the time of day, season and weather.

Methods have recently been introduced which use low-dimensional representations of images of objects to perform recognition; see for example [9, 18, 22]. These methods, often termed appearance-based methods, differ from feature-based methods in that their low-dimensional representation is, in a least-squared sense, faithful to the original image. Systems such as SLAM [18] and Eigenfaces [22] have demonstrated the power of appearance-based methods both in ease of implementation and in accuracy.

Beyond recognition, these methods have been shown to be useful for inspection, visual tracking, visual control, and mobile robot navigation [20]. Yet these methods suffer from an important drawback: recognition of an object (or face) under a particular pose and lighting can be performed reliably *provided that the object has been previously seen under similar circumstances*. In other words, these methods in their original form have no way of extrapolating to novel viewing conditions. Here, we consider the construction of a generative appearance model and demonstrate its usefulness for image-based rendering and recognition.

Arbitrary illumination can be modeled as a scalar function on a four-dimensional manifold of light

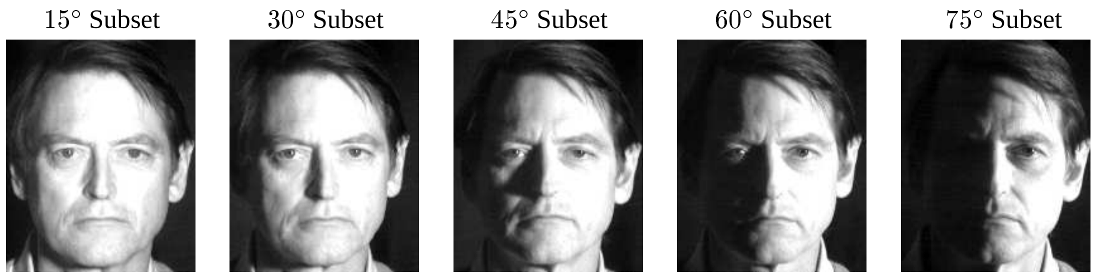


Figure 1: An example image from each subset of the Harvard Database.

rays [17] . It is also easy to show that the set of all possible n-pixel monochrome images of an object observed in fixed pose, but over all possible illumination conditions is a convex cone in $\mathbb{R}^n$. However, without limiting assumptions about the possible light sources, the bidirectional reflectance density functions (BRDF), or object geometry, it is difficult to draw limiting conclusions about the set of images. For example, when the object is restricted to being convex and Lambertian and when this object is illuminated by an arbitrary number of point light sources at infinity, this *illumination cone* is polyhedral, it can be characterized by a finite number of extreme rays, it can be constructed from a minimum of three images, and its dimension equals the number of distinct surface normals [3].

Now, as a point light source (nearby or at infinity) assumes different locations, both the shading and shadowing in an image change. For any finite set of point light sources illuminating an object viewed under either orthographic or perspective projection, there is an equivalence class of object shapes having the same set of shadows [14]. Members of this equivalence class differ by a four parameter family of projective transformations, and the shadows of a transformed object are identical when the same transformation is applied to the light source locations. Under orthographic projection, this family is the generalized bas-relief (GBR) transformation, a subgroup of affine transformations [1]. Furthermore, for objects with Lambertian surfaces illuminated by distant light sources, the equivalence class of object shapes which preserves shadows also preserves surface shading. Hence, two objects differing by a GBR transformation have identical illumination cones. A natural implication is that two members of an equivalence class cannot be distinguished from a fixed viewpoint, and any reconstruction algorithm can only recover structure up to this family of transformations [1, 14].

Nevertheless, this leads to a method for image-based rendering and reconstruction. In particular, given three or more images of a Lambertian surface with a continuous depth function illuminated by unknown light sources at infinity, surface geometry and an albedo map can be estimated up to a generalized bas-relief transformation. Nothing in the shading or shadows can be used to further resolve this ambiguity. Synthetic images of the surface can be rendered using ray-tracing for multiple point or extended light sources. Even though the surface is only recovered up to a GBR transformation, every image will be physically valid. Class specific information can be used to select a representative member of the equivalence class, and images can be rendered from arbitrary viewpoints as well [6].

Surface reconstruction up to a GBR transformation and the illumination cone representation have been married to produce a representation useful for object recognition. From three or more images of a continuous (possibly non-convex) Lambertian surface, the GBR surface can be reconstructed, and extreme rays of its illumination cone can be constructed. For a collection of objects, each object is represented by a cone, and recognition is performed through nearest neighbor classification by measuring the minimal distance of an image to each cone. We demonstrate the utility of this approach to the problem of face recognition (a class of non-convex and non-Lambertian objects with similar geometry). The method has been tested on a database with 660 images of 10 faces in fixed pose but variable lighting. The object representations was also extended to include variable viewpoints (see Sec. 5.) and our method was also tested on a database with 1350 images of 10 faces with variable pose and lighting.

In the following section, we summarize some properties of the illumination cones of convex objects while in Section. 3. we discuss issues related to the construction of cone representations for non-convex Lambertian objects. In Sec. 4.1., we provide experimental results for recognizing faces in fixed pose, but under variable lighting while in Sec. 5. we consider issues of recognition under variation in both pose and lighting (The methods and experimental results presented in Section 5. should be considered preliminary). Much of the material in this paper was presented in [1, 3, 6, 7, 14].

2. The Illumination Cone

In earlier work, it was shown that for an object with convex shape and Lambertian reflectance, the set of all n-pixel images under an arbitrary combination of point light sources forms a convex polyhedral cone in the image space $\mathbb{R}^n$. This cone can be constructed from as few as three images [3]. Here we summarize the relevant results.

To begin, consider a convex object with a Lambertian reflectance function which is illuminated by a single point source at infinity. Let $\mathbf{x} \in \mathbb{R}^n$ denote an image of this object with n pixels. Let $B \in \mathbb{R}^{n\times 3}$ be a matrix where each row of B is the product of the albedo with the inward pointing unit normal for a point on the surface projecting to a particular pixel in the image. A point light source at infinity can be represented 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. A convex Lambertian surface with normals and albedo given by B, illuminated by $\mathbf{s}$, produces an image $\mathbf{x}$ given by

$$\mathbf{x} = \max(B\mathbf{s}, \mathbf{0}), \tag{1}$$

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 an *attached shadow*. Convexity of the object's shape is assumed at this point to avoid *cast shadows* (shadows that the object casts on itself). While attached shadows are defined by a simple local geometric conditions, cast shadows must satisfy a global condition. When no part of the surface is shadowed, $\mathbf{x}$ lies in the 3-D subspace $\mathcal{L}$ given by the span of the matrix B [9, 19, 21]; the subset $\mathcal{L}_0 \subset \mathcal{L}$ having no shadows (i.e., intersecting with the non-negative orthant) forms a convex cone [3].

The illumination subspace $\mathcal{L}$ slices through other orthants as well as the non-negative orthant. Let $\mathcal{L}_i$ be the intersection of the illumination subspace $\mathcal{L}$ with an orthant i in $\mathbb{R}^n$ through which $\mathcal{L}$ passes. Certain components of $\mathbf{x} \in \mathcal{L}_i$ are always negative and others always greater than or equal to zero. Since image intensity is always non-negative, the image corresponding to points in $\mathcal{L}_i$ is formed by a projection P_i determined by Equation 1. The projection P_i is such that it leaves the non-negative components of $\mathbf{x} \in \mathcal{L}_i$ untouched, while the negative components of $\mathbf{x}$ become zero. The projected set $P_i(\mathcal{L}_i)$ is also a convex cone. $\mathcal{L}$ intersects at most $n(n-1)+2$ orthants [3], and so the set of images created by varying the direction and strength of a *single* light source at infinity is given by the union of at most $n(n-1)+2$ convex cones, each of which is at most three dimensional.

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.,

$$\mathbf{x} = \sum_{i=1}^{k} \max(B\mathbf{s}_i, \mathbf{0}) \tag{2}$$

where $\mathbf{s}_i$ is a single light source. It follows that 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.

The cone can be constructed as the convex hull of the $n(n-1)+2$ single light source convex cones. Alternatively, any image in the cone $\mathcal{C}$ (including the boundary) can be found as a convex combination of *extreme rays* (extreme images) given by

$$\mathbf{x}_{ij} = \max(B\mathbf{s}_{ij}, \mathbf{0}), \tag{3}$$

where

$$\mathbf{s}_{ij} = \mathbf{b}_i \times \mathbf{b}_j. \tag{4}$$

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

3. Constructing the Illumination Cone

Equations 3 and 4 suggest a way to construct the illumination cone for each object: gather three or more images under varying illumination without shadowing and use these images to estimate the three-dimensional illumination subspace $\mathcal{L}$. One way of estimating this is to normalize the images to be of unit length, and then use singular value decomposition (SVD) to estimate the best three-dimensional orthogonal basis B^* in a least-squares sense. Note that the basis B^* differs from B by an unknown linear transformation, i.e., $B = B^*A$ where $A \in GL(3)$; for any light source, $\mathbf{x} = B\mathbf{s} = (B^*A)(A^{-1}s)$ [11]. Nonetheless from B^*, the extreme rays defining the illumination cone $\mathcal{C}$ can be computed using Eqs. 3 and 4. We now consider three issues or problems that arise in using this method for nonconvex objects such as faces – See [7] for detailed solutions.

The first problem that arises with the above procedure is with the estimation of B^*. For even a convex object whose Gaussian image covers the Gauss sphere (i.e., its 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. For non-convex objects, shadowing in the modeling images is likely to be more pronounced. When SVD is used to estimate B^* from images with shadows, these systematic errors can bias its estimate significantly.

The next problem is that usually m, the number of independent normals in B, can be large (more than a thousand) hence the number of extreme rays needed to completely define the illumination cone can run in the millions. Therefore, we must approximate the cone in some fashion; in this work, we choose to use a small number of extreme rays (images). The hope is that a sub-sampled cone will provide an adequate approximation that negligibly decreases recognition performance; in our experience, around 60-80 images are sufficient, provided that the corresponding light source directions $\mathbf{s}_{ij}$ are more or less uniform on the illumination sphere. The resulting cone $\mathcal{C}^*$ is a subset of the object's true cone $\mathcal{C}$. An alternative approximation to $\mathcal{C}$ can be obtained by directly sampling the space of light source directions rather than generating the samples through Eq. 4. While the resulting images form the extreme rays of the representation $\mathcal{C}^*$ and lie on the boundary of $\mathcal{C}$, they are not necessarily extreme rays of $\mathcal{C}$. Again $\mathcal{C}^*$ is a subset of $\mathcal{C}$.

The last issue arises because objects such as faces are non-convex, and so cast shadows can cover significant portions of the face under extreme illumination (See the images from Subsets 4 and 5 in Figure 1.) However, the image formation model given by Eq. 1 does not account for cast shadows. For the light source direction associated with each extreme ray given by Equation 4, we need to determine which pixels (elements) of $\mathbf{x}_{ij}$ will fall in a cast shadow.

It has been shown [1, 23] that from multiple images where the light source directions are unknown, one can only recover a Lambertian surface up to a three-parameter family given by the generalized bas-relief (GBR) transformation. This family of affine transformations scales the relief (flattens or extrudes) and introduces an additive plane. Consequently, when computing $\mathbf{s}^*_{ij}$ from B^*, the light source direction differs from the true light source by a GBR transformation. Since shadows are preserved under these transformations [1], images synthesized from a surface whose normal field is given by B^* and illuminated by light source $\mathbf{s}^*_{ij}$ will have correct shadowing. Thus, in constructing the extreme rays of the cone, we first reconstruct a surface and then use ray-tracing techniques to determine which points lie in a cast shadow. It should be noted that the vector field B^* estimated via SVD may not be integrable, so prior to reconstructing the surface up to GBR, integrability of B^* is enforced.

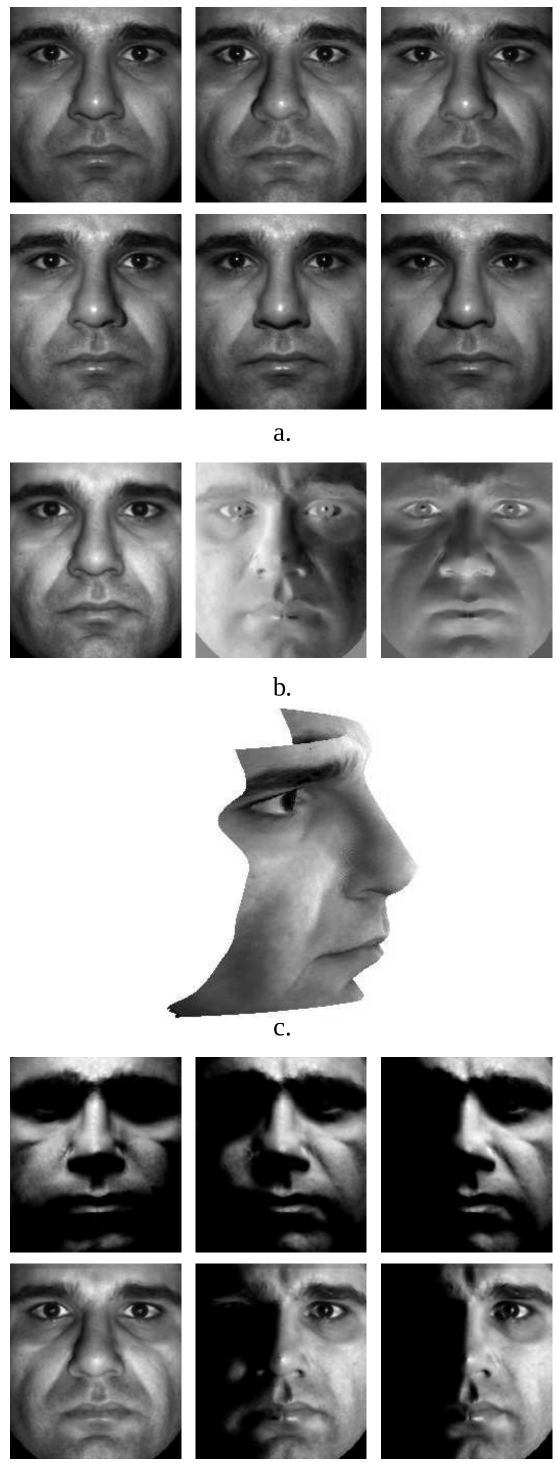


Figure 2: The process of constructing the cone $\mathcal{C}^*$: a. The training images; b. Images corresponding to columns of B^*; c. Reconstruction up to a GBR transformation; d. Sample images from the illumination cone under novel lighting conditions, but fixed pose.

This leads to the following steps for constructing a representation of the illumination cone $\mathcal{C}^*$ from a set of images taken under unknown lighting.

1. Estimate B^* from training images.
2. Enforce integrability of B^*.
3. Reconstruct the surface up to GBR.
4. For a set of light source directions that uniformly sample the sphere, synthesize extreme rays (images) of the cone that account for cast and attached shadows.

Details of the steps and the entire method can be found in [7].

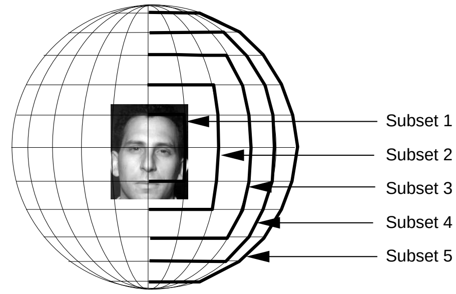


Figure 3: Images in the Harvard face database were acquired by sampling half the illumination sphere at 15° increments in longitude and latitude. The highlighted lines indicate the light source directions for Subsets 1 through 5.

4. Recognition

The cone $\mathcal{C}^*$ can be used in a natural way for object recognition, and we empirically evaluate it within the context of face recognition. In experiments described below, we compare three recognition algorithms to the proposed method. From a set of face images labeled with the person's identity (*the learning set*) and an unlabeled set of face images from the same group of people (*the test set*), each algorithm is used to identify the person in the test images. For more details of the comparison algorithms, see [2]. We assume that the face has been located and aligned within the image.

The simplest recognition scheme is a nearest neighbor classifier in the image space [4]. 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. If all of the images are normalized to have zero mean and unit variance, this procedure is equivalent to choosing the image in the learning set that best *correlates* with the test image.

As correlation methods are computationally expensive and require great amounts of storage, it is natural to pursue dimensionality reduction schemes. A technique now commonly used in computer vision – particularly in face recognition – is principal components analysis (PCA) which is popularly known as *Eigenfaces* [9, 16, 18, 22]. 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 an f-dimensional feature space is performed. A face in a test image $\mathbf{x}$ is recognized by projecting $\mathbf{x}$ into the feature space and performing nearest neighbor classification in $\mathbb{R}^f$. The projection matrix $W \in \mathbb{R}^{f\times n}$ is chosen to maximize the scatter of all projected samples. 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 [2].

A third approach is to model the illumination variation of each face as a three-dimensional linear subspace $\mathcal{L}$ as described in Section 2. To perform recognition, we simply compute the distance of the test image to each linear subspace and choose the face 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 [21] and is related to [10, 19]. While this models the variation in intensity when the surface is completely illuminated, it does not model shadowing.

Finally, given a test image $\mathbf{x}$, recognition using *illumination cones* is performed by first computing the distance of the test image to each cone, and then choosing the face that corresponds to the shortest distance. Since each cone is convex, the distance can be found by solving a convex optimization problem. In particular, a modified version of the non-negative linear least squares technique contained in Matlab was used in our implementation, and this 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.

4.1. Experimental Results

To test the effectiveness of these recognition algorithms, we performed a series of experiments on a database from the Harvard Robotics Laboratory in which lighting had been systematically varied [9, 10]. In each image, a subject held his/her head steady while being illuminated by a dominant light source. The space of light source directions, which can be parameterized by spherical angles, was then sampled in 15° increments. See Figure 3. From this database, we used 660 images of 10 persons (66 of each). We extracted five subsets to quantify the effects of varying lighting.

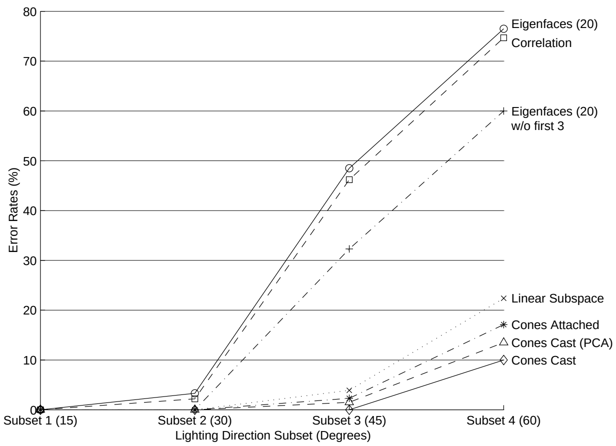


Figure 4: **Extrapolation:** When each of the methods is trained on images with near frontal illumination (Subset 1), this graph shows the error rates under more extreme light source conditions.

Figure 5: Synthesized images under variable pose but with fixed lighting. The representation was constructed from the images in Figure 2.a.

A sample image from each subset is shown in Fig. 1. Subset 1 (respectively 2, 3, 4, 5) contains 60 (respectively 90, 130, 170, 210) images for which both the longitudinal and latitudinal angles of light source direction are within $15°$ (respectively $30°, 45°, 60°, 75°$) of the camera axis.

Mirroring the extrapolation experiment described in [2], each method was trained on samples from Subset 1 and then tested using samples from Subsets 2, 3, 4 and 5. (Note that when tested on Subset 1, all methods performed without error). Figure 4 shows the result from this experiment. Error rates for subset 5 are not shown since they approached chance (90%) for the correlation and Eigenfaces methods, and at their best they were 37% when cones were used.

5. Image-Based Rendering and Recognition Under Variable Pose

So far, we have considered the issue of recognizing an object (particularly faces) under a wide range of illumination conditions. Here we consider pose variation along with illumination variation. Clearly, for every pose of the object, the set of images under all lighting conditions is a convex cone. Under a weak perspective projection imaging 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 viewing direction. Within a face recognition system, the face detection process generally provides estimates for the image plane transformation. 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 two degrees of freedom.

Three possible approaches to handle viewing direction variation are: 1) Estimate the viewpoint, perhaps iteratively, during the recognition process much like in the alignment approach [13]; 2) Find some attribute of the image which is invariant to viewpoint variation; or 3) Model the set of images under viewing direction variation. Here, we choose the third approach since, as shall be shown, the recovered surface and albedo pattern provide a suitable generative model.

However, one complication arises because of the generalized bas-relief (GBR) ambiguity. Without resolution of this ambiguity, images synthesized from a GBR reconstruction will differ from a valid image by an affine warp of image coordinates since GBR is a 3-D affine transformation and weak perspective is a linear camera model. Since this is an image transformation, one could perform recognition over variation in viewing direction and affine image transformations rather than similarity transformations. Alternatively, one can attempt to resolve the GBR ambiguity to obtain a Euclidean reconstruction. For faces, one can use bilateral symmetry and class information to resolve the ambiguity. Once resolved, it is a simple matter to use ray-tracing techniques to determine shadows and render synthetic images under pose and lighting variation. See Figure 5.

With the GBR resolved, an object can be represented

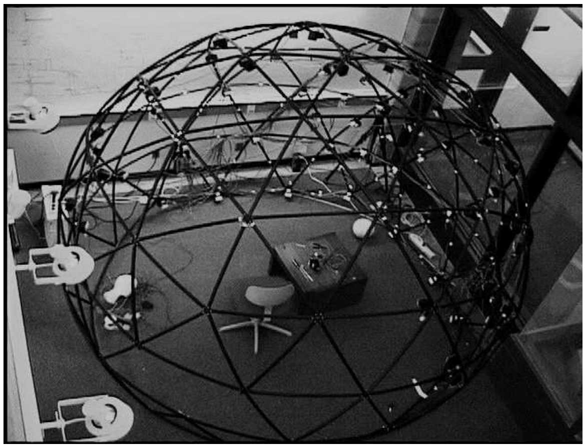


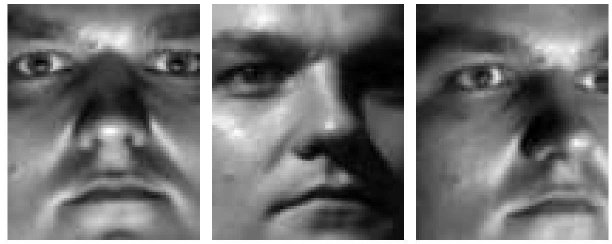
Figure 6: A geodesic dome with 64 strobes used to gather images reported in Table 1.

by first sampling the space of viewing directions and then constructing a cone for each viewing direction. Recognition using this raw representation is going to be costly since computing distance to cone is $O(ne^2)$, where e is the number of extreme rays; for a convex object, the cone has $O(n^2)$ extreme rays [3]. From an empirical study, it was conjectured in [3] that the cone for a typical object is flat (i.e., all points lie near a low-dimensional linear subspace), and this was confirmed for faces in [5]. Hence, an alternative is to model a face in fixed pose but over all lighting conditions by a low-dimensional linear subspace. Finally, for a set of sample viewing directions, we construct subspaces which approximate the corresponding cones. Recognition of a test image $\mathbf{x}$ is then performed by finding the closest linear subspace to $\mathbf{x}$.

For the experimental results reported below, subspaces were constructed by sampling the viewing sphere at 4° intervals over the elevation from -20° to $+24^\circ$ and the azimuth from -4° to $+28^\circ$ about frontal. We chose to use 11-D linear subspaces for each pose since eleven dimensions captured over 99% of the variation in the sample extreme rays. Recognition was performed by computing the distance of a test image to each 11-D subspace.

5.1. Experimental Results

The experiment described in Section 4.1. was limited to the available dataset from the Harvard Robotics Laboratory. To perform more extensive experimentation, we have constructed the geodesic lighting rig with 64 computer controlled xenon strobes shown in Fig. 6. Using this rig, we can modify the illumination at frame rates. Here we report on some preliminary results us-



Test Images

Closest image in cone representation

Figure 7: The top row shows three images from the test set, and the bottom row shows the closest reconstructed image from the representation. Note that these images are not explicitly stored, but lie within the closest matching linear subspace.

Error Rate (%)				
	Lighting Variation			
Pose	12°	25°	50°	77°
Frontal	0.0	0.0	0.0	0.7
12°	0.0	0.0	0.8	1.4
24°	0.0	0.0	0.0	9.2

Table 1: Error rates over variable pose and lighting for a 1350 image subset of the Yale Face Database B.

ing this rig in the following experiment; images of ten individuals were acquired in three poses (frontal and rotations of approximately 12° and 24° degrees about the vertical) under 64 different lighting conditions. The method was tested using a subset of 1350 images. As in the previous experiment, the images were divided into subsets (12°, 25°, 50° and 77°) according to the angle of the light source with the camera's axis.

The representation described in Section 5. was constructed using images with near frontal lighting from the frontal pose. For ten individuals, there were 108 11-D subspaces. Given a test image, the point on the nearest 11-D subspace is computed. Figure 7 shows the closest match for images of an individual in three poses. This figure qualitatively shows how well the union of 11-D subspaces approximates the true cones. Over the entire dataset of 1350 images, Table 1 summarizes the recognition results for each of the subsets and poses. Essentially, recognition is perfect for all poses with lighting directions to $25°$, and then performance begins to degrade. However, when lighting is within $50°$, the cumulative error rates over all three poses was about 0.1% while over the whole database including those at $77°$, the error rate was 1.2%.

6. Conclusions

We have exploited the fact that images of an object in fixed pose but under variable illumination form a convex cone in the image space, the fact that for Lambertian surfaces this cone can be constructed through an estimate of the corresponding 3-D linear subspace, and the empirical observation that most of the cone lies near a low-dimensional linear subspace. Even though structure can only be recovered up to a generalized bas-relief (GBR) transformation when the 3-D subspace is estimated with unknown light source directions, one can still construct a valid cone representation. We have demonstrated the use of this generative model and cone representation within the context of recognizing faces under extremes of illumination. When the GBR ambiguity is resolved, realistic images can be rendered under pose lighting variation. Furthermore, these synthetic images can be used to construct representations for recognition under variable pose and lighting.

The algorithm and recognition results reported in Sec. 5. should be considered preliminary. More extensive experimentation over a larger database is underway. Note that this database will be publicly available upon completion of this study. Though preliminary, the reported results do support the power of this generative model for recognition and inspection. There is much to be done to develop efficient and effective classifiers that exploit this generative model, yet are less computationally and memory expensive.

While our experiments have focused on the domain of face recognition, we believe that the underlying concepts can be applied to many vision and robotics problems where appearance-based approaches have been successful such as robot navigation, inspection, and visual guidance [20]. Some of these concepts have already been exploited to make visual tracking more robust to dynamic lighting changes [8], and clearly visual tracking is a critical element of visual servoing [12, 15].

Acknowledgments

D. J. Kriegman and A.S. Georghiades were supported under NSF NYI IRI-9257990 and ARO DAAG55-98-1-0168. P. N. Belhumeur was supported by a Presidential Early Career Award, an NSF Career Award IRI-9703134, and ARO grant DAAH04-95-1-0494.

References

[1] P. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. *Int. J. Computer Vision*, 35(1):33–44, 1999.

[2] 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.

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

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

[5] R. Epstein, P. Hallinan, and A. Yuille. 5+/-2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *PBMCV*, 1995.

[6] 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, 1999.

[7] 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.

[8] G. D. Hager and P. N. Belhumeur. Real-time tracking of image regions with changes in geometry and illumination. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 403–410, 1996.

[9] 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.

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

[11] H. Hayakawa. Photometric stereo under a light-source with arbitrary motion. *J. Optical Society of America A*, 11(11):3079–3089, Nov. 1994.

[12] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Trans. on Robotics and Automation*, 12(5):651–670, 1996.

[13] D. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *IJCV*, 5(2):195–212, November 1990.

[14] D. Kriegman and P. Belhumeur. What shadows reveal about object structure. In *Proc. European Conf. on Computer Vision*, pages 399–414, 1998.

[15] D. Kriegman, G. Hager, and A. Morse. *The Confluence of Vision and Control*. Springer-Verlag, 1998.

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

[17] M. Langer and S. Zucker. A ray-based computational model of light sources and illumination. In *Physics Based Modeling Workshop in Computer Vision*, 1995.

[18] H. Murase and S. Nayar. Visual learning and recognition of 3-D objects from appearance. *Int. J. Computer Vision*, 14(1):5–24, 1995.

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

[20] S. Nayar, S. Nene, and H. Murase. Subspace methods for robot vision. *IEEE Trans. on Robotics and Automation*, 12(5):750–758, October 1996.

[21] A. Shashua. On photometric issues in 3D visual recognition form a single image. *Int. J. Computer Vision*, 21:99–122, 1997.

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1), 1991.

[23] 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.