

Dimensionality of Illumination Manifolds in Eigenspace*

Shree K. Nayar[†] and Hiroshi Murase[‡]

[†]Department of Computer Science
Columbia University, New York, NY 10027, USA
nayar@cs.columbia.edu

[‡]NTT Basic Research Labs
3-1, Morinosato Wakamiya, Atsugi-shi
Kanagawa 243-01, Japan
murase@siva.ntt.jp

CUCS-021-94

August, 1994

Revised: September 1995

Abstract

In appearance matching, a 3D object is represented as a continuous appearance manifold in a low-dimensional subspace parametrized by object pose and illumination direction. In this correspondence, the structural properties of appearance manifolds are analyzed with the aim of making appearance representation more efficient in off-line computation, storage requirements, and on-line recognition time. In particular, the effect of illumination on the structure of the appearance manifold is studied. It is shown that for an ideal diffuse surface of arbitrary texture, the appearance manifold is linear and of dimensionality 3. This enables the construction of the entire illumination manifold from just three images of the object taken using linearly independent light sources. This result is shown to hold even for illumination by multiple light sources and for concave surfaces that exhibit interreflections. Finally, a simple but efficient algorithm is presented that uses just three manifold points for recognizing images taken under novel illuminations. Experiments are conducted to demonstrate the use of these results for efficient appearance learning and matching.

1 Introduction

Appearance matching techniques are fast becoming popular in machine vision. Recent applications include face recognition [17] and the real-time recognition of complex 3D objects [4]. A representation of object appearance called the parametric eigenspace has resulted from this work [4]. For a given vision application, a visual workspace is first defined as the range of visual appearances that result from varying the parameters of the task. This workspace is sampled to obtain an image set that is used to compute a low-dimensional linear subspace [13], called the eigenspace, in which the visual workspace is represented by one or more parametrized manifolds. During recognition, novel images are projected into the eigenspace. The closest manifold and the exact location of the closest point on the manifold reveal the task parameters.

The parametric eigenspace representation has found several important applications. These include learning [4] and real-time recognition of 3D objects [3], positioning and tracking of 3D objects by a robot manipulator [9], and illumination planning for robust object recognition [5]. Recently, a recognition system with 100 complex objects in its database was developed that is solely based on appearance matching [3]. The sheer efficiency of appearance matching enables the system to accomplish both recognition and pose estimation in real time using nothing more than a standard workstation equipped with an image sensor (see Figure 1).

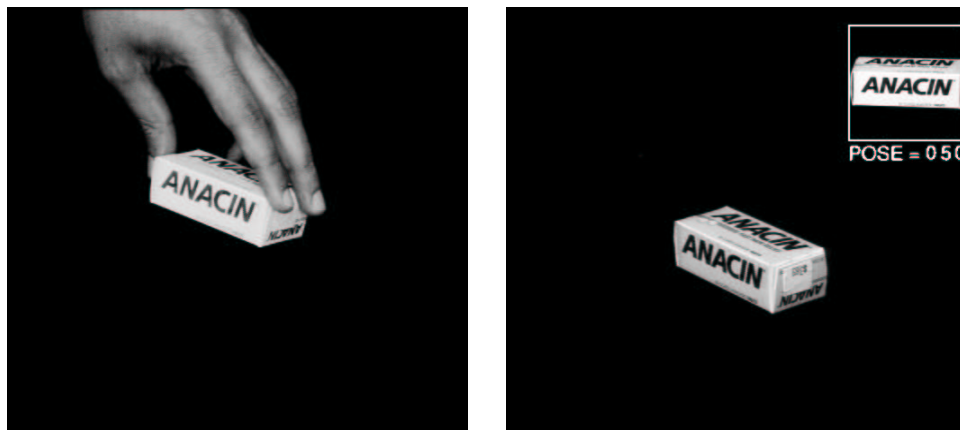


Figure 1: An automated recognition system with 100 objects in its database. A complete recognition and pose estimation cycle takes less than 1 second on a Sun SPARC workstation [6].

In the context of large systems, the primary bottleneck in appearance matching has turned out to be the learning stage which includes the acquisition of large image sets, the

computation of eigenspaces from large covariance matrices, and the construction of parametric appearance manifolds. In object recognition, each object is represented as a separate manifold in eigenspace that is parametrized by pose and illumination parameters. The efficiency of the learning stage is determined by the number of sample images needed to compute an accurate appearance manifold. This brings us to the following question: What is the smallest number of images needed for constructing the appearance manifold for any given object?

The answer to the above question lies in the structural properties of appearance manifolds. The structure of an object's manifold is closely related to its geometric and reflectance properties. In special cases, such as solids of high symmetry and solids of revolution, one can make concrete statements regarding the dimensionality of the manifold. For instance, given a fixed illumination direction and viewpoint, the manifold for a sphere of uniform reflectance is simply a point since the sphere appears the same in all its poses. This unfortunately is an extreme instance of little practical value. Under perspective projection, the relation between object shape and manifold structure is complex to say the least. A general expression that relates object pose to manifold structure would be much to hope for.

In contrast, the function space associated with object reflectance is more concise and hence conducive to analysis. It is possible to establish, under certain reflectance assumptions, a closed-form relationship between illumination parameters and manifold structure [8]. Given that the eigenspaces we use are linear subspaces, the class of linear reflectance functions [15][16] is of particular interest to us. We show that for this reflectance class the structure of the illumination manifold is completely determined from a small number of samples of the manifold. In particular, for Lambertian surfaces of arbitrary texture, the entire illumination manifold can be constructed from just three images taken using known illuminants. Alternatively, the dimensionality of the illumination manifold is exactly 3. This result is supported by a detailed empirical investigation reported recently by Epstien et al. [1]. We use the above bound on the manifold dimensionality to show that novel images of the object can be recognized from just three projections on the illumination manifold without the explicit construction of the manifold. In addition, the validity of the above results for illumination by multiple sources and in the presence of interreflections caused by concave surfaces is demonstrated.

For ideal diffuse objects, these results have direct implications on the efficiency of both learning and recognition, as they dramatically reduce the number of images needed for appearance representation. Our results for diffuse surfaces stem from the observation that the image of a diffuse object under any illumination can be expressed as a linear combination of

images taken using three independent basis illuminants. Such a linear combination does not generally exist for objects with nonlinear reflectance functions. For instance, a pure specular object would produce only strong highlights for each of the basis illuminants. The highlights produced by a novel source cannot in general be expressed as a linear combination of basis images. In fact, it is hard to envision non-trivial upper bounds on the dimensionality of a vector space containing illumination manifolds for the class of nonlinear reflectance functions.

2 Linear Reflectance Models

We will assume throughout our presentation that all surface properties and image brightness values correspond to a single wavelength, λ . The dependence of our results on λ vanishes in the case of gray textured surfaces illuminated by white-light sources. However, to accommodate the general case of colored surfaces and colored illuminants, we assume that all brightness values are measured using narrow-band filters, say, narrow-band red, green, and blue filters. This ensures that the wavelength of light is in effect fixed for any given color band.

We begin with a brief discussion on the class of linear reflectance models [15][16]. Our interest in this class stems from the fact that it allows the analysis of illumination effects in the context of linear subspaces. A linear reflectance function may be written as:

$$e(\mathbf{x}) = \mathbf{p}(\mathbf{x}) \cdot \mathbf{q} \tag{1}$$

where, $e(\mathbf{x})$ is the image irradiance or intensity at point \mathbf{x} , $\mathbf{p}(\mathbf{x})$ represents local surface properties, and \mathbf{q} is an arbitrary vector that could depend on the illumination and the viewpoint of the observer. If $Dim(\mathbf{p}) = Dim(\mathbf{q}) = k$, we have a k -order linear reflectance model (see [15][16]).

When is the linear reflectance model valid in practice? In general, reflectance functions can be viewed as the combination of surface (specular) and body (diffuse) components [11]. Surface reflection is a nonlinear function of viewpoint. In contrast, the body component is relatively less viewpoint dependent. Though the dependence can be significant in the case of surfaces with high macroscopic roughness (see [14] [7]), many man-made objects (for instance, those with matte paints) as well as some natural surfaces can be approximated by a linear model. The most popular and widely used approximation is the Lambertian model where surface radiance, and hence also image brightness, depend only on the irradiance of the surface and not the observer's viewpoint. The image brightness of a Lambertian surface

element illuminated by a point light source is:

$$e(\mathbf{x}) = \mathbf{n}(\mathbf{x}) \cdot \mathbf{s} \quad (2)$$

Here, $\mathbf{n}(\mathbf{x}) = \rho(\mathbf{x}) \hat{\mathbf{n}}(\mathbf{x})$ is the normal vector, where ρ is the local surface albedo and $\hat{\mathbf{n}}$ is the local unit surface normal. Similarly, $\mathbf{s} = b \hat{\mathbf{s}}$ is the source vector, where b represents the intensity of the source and $\hat{\mathbf{s}}$ is a unit vector in the direction of the source. We are assuming here that the source is distant and hence its direction is independent of the location of the surface point in the scene. From the above expression, we see that the Lambertian model is a linear one of order 3.

3 Images as Linear Combinations

Given that \mathbf{s} is a constant three-dimensional vector, it is clear that three non-coplanar source vectors, say $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, can be used as a basis to represent any source vector. Any arbitrary source vector is simply a linear combination of the three basis vectors:

$$\mathbf{s} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3 \quad (3)$$

Let us define $\mathbf{\Lambda} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3]$ as the basis source matrix. Then, the coefficient vector $\mathbf{\Phi} = [\alpha_1 \ \alpha_2 \ \alpha_3]^T$ for any given source can be determined as:

$$\mathbf{\Phi} = \mathbf{\Lambda}^{-1} \mathbf{s} \quad (4)$$

The image brightness values of a Lambertian surface point due to the three basis illuminants are:

$$\begin{aligned} e_1(\mathbf{x}) &= \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_1 \\ e_2(\mathbf{x}) &= \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_2 \\ e_3(\mathbf{x}) &= \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_3 \end{aligned} \quad (5)$$

From (3) and (5), the brightness of the surface point due to a novel source is:

$$\begin{aligned} e(\mathbf{x}) &= \mathbf{n}(\mathbf{x}) \cdot \mathbf{s} \\ &= \alpha_1 \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_1 + \alpha_2 \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_2 + \alpha_3 \mathbf{n}(\mathbf{x}) \cdot \mathbf{a}_3 \end{aligned} \quad (6)$$

If we define $\mathbf{\Gamma} = [e_1, e_2, e_3]^T$, we have:

$$e(\mathbf{x}) = \mathbf{\Gamma}^T(\mathbf{x}) \mathbf{\Phi} \quad (7)$$

Note the similarity between the linear combination in the above expression and the one in (3). The basis illuminants result in basis brightness values. The brightness due to any novel source is the same linear combination of the basis brightness values as the novel source is of the basis illuminants. This result may not seem obvious at first glance. However, it turns intuitive when one notes that the three brightness values in $\mathbf{\Gamma}(\mathbf{x})$ and the three corresponding sources $\mathbf{\Lambda}$ contain all the information required to estimate the albedo $\rho(\mathbf{x})$ and the unit normal vector $\hat{\mathbf{n}}(\mathbf{x})$ as done in the case of photometric stereo [18]. It is therefore not surprising that the three brightness values corresponding to the basis illuminants can be used to predict brightness for any desired source vector. The linear combination of (7) has been used in its explicit form for the analysis of surface color [15] as well as specular detection and photometric recognition [16].

It is immediate that expression (7) holds independently for all points on the imaged object. Hence, if the basis illuminants and the novel source are distant and are visible to all observed points on the object, the image \mathbf{I} of the object under the novel illumination can be expressed as a linear combination of its three images $\mathbf{\Pi} = [\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3]$ due to the basis illuminants.

$$\mathbf{I} = \mathbf{\Pi} \mathbf{\Phi} \quad (8)$$

The above linear combination holds true irrespective of the texture (albedo variation) of the surface. The main assumptions used thus far are; (a) the surface is illuminated by a single distant point source that is visible to all imaged points, and (b) the surface is convex and consequently image brightness is due to only direct source illumination and not interreflections. In subsequent sections, we will see that these two assumptions can be substantially weakened, and that expression (8) does in fact hold true for multiple sources and concave surfaces with interreflections.

4 Multiple Novel Sources

Consider a Lambertian surface simultaneously illuminated by R point sources. In this case, the image brightness of a surface point will be the sum of the contributions of individual sources:

$$e(\mathbf{x}) = \sum_{r=1}^R \mathbf{n}(\mathbf{x}) \cdot \mathbf{s}_r$$

$$= \mathbf{n}(\mathbf{x}) \cdot \sum_{r=1}^R \mathbf{s}_r = \mathbf{n}(\mathbf{x}) \cdot \tilde{\mathbf{s}} \quad (9)$$

Note that $\tilde{\mathbf{s}}$ serves as a single *effective source* and is simply the average, or center of mass, of the set of individual source vectors. It should be evident that in the above derivation the individual sources need not be point sources. They could be extended sources with arbitrary radiance functions and still each be replaced by an effective point source. The above result is well-known and has surfaced in various guises in previous work. Its implication is that, for Lambertian reflectance, any number of novel sources that are all visible to the entire imaged surface can be viewed as a single effective point source $\tilde{\mathbf{s}}$. Hence, the linear combination of (8) holds true for multiple sources; the coefficient vector Φ is simply that of $\tilde{\mathbf{s}}$.

5 Interreflections

Next, let us consider the case of a concave Lambertian surface with arbitrary texture. We assume that each of the basis illuminants and subsequent novel sources are visible to all points on the observed surface. The image brightness of each surface point in this case is due not only to the source but also the contributions of other points on the surface that are visible to it. Though, in general, an infinite number of interreflections occur between any two mutually visible surface points, the brightness of a surface point can be expressed as the sum of the brightness due to direct source illumination and contributions due to the *final* radiance values of all surface points visible to it. Thus, for any given wavelength λ of incident light, the brightness image of the surface is:

$$\mathbf{I} = \mathbf{I}_s + \mathbf{P} \mathbf{K} \mathbf{I} \quad (10)$$

where, \mathbf{I}_s is the image due to direct source illumination, \mathbf{P} is the albedo matrix whose diagonal elements are the albedo values of individual infinitesimal elements on the surface, and \mathbf{K} is the interreflection kernel that captures the relative geometric configurations of pairs of surface elements.

In [12], the above brightness equation was analyzed to show that the concave surface behaves exactly like a Lambertian one without interreflections, but with a different set of surface normals and albedo values. The underlying assumption is that all points on the surface are visible and illuminated, i.e. no self-occlusions and self-shadows. The above apparent surface is called the *pseudo surface*. The relation between the pseudo surface and

the actual surface was found to be:

$$\mathbf{F}_p = [\mathcal{I} - \mathbf{P} \mathbf{K}]^{-1} \mathbf{F} \quad (11)$$

Here, the matrix $\mathbf{F}_p = [\mathbf{n}_p^1, \mathbf{n}_p^2, \dots, \mathbf{n}_p^n]$ represents the pseudo surface and is composed of the pseudo normal vectors of all n visible surface elements. The true surface is given by the matrix $\mathbf{F} = [\mathbf{n}^1, \mathbf{n}^2, \dots, \mathbf{n}^n]$.

This result implies that, under any given illumination that satisfies the assumptions stated above, the image of a concave surface \mathbf{F} with all its interreflections exactly equals the image of its corresponding pseudo surface \mathbf{F}_p without interreflections. As a result, all the linear combinations derived in the previous sections hold true for concave Lambertian surfaces of arbitrary texture. In general, the pseudo surface varies with the wavelength of incident light. Once again, this dependence on wavelength vanishes in the cases on gray textured surfaces illuminated by white-light sources. In the general case of colored surfaces and colored illuminants, the linear combination of (8) remains valid when the images are taken using narrow-band filters, say, narrow-band red, green, and blue filters [10][2].

6 Multispectral Images

As stated earlier, if the application involves the use of a color image sensor, it is assumed that each color band is obtained using a narrow-band spectral filter. Then, the intensities of the three basis illuminants, irrespective of their spectral distributions, are fixed for each of the narrow bands. In other words, all the results derived thus far remain valid in each band. Appearance manifolds of an object can then be constructed independently for each band (as in [6]) and recognition is deemed successful if all bands of a novel image are found to match the same object in the appearance database.

Alternatively, an image vector can be constructed by concatenating the multiple bands of the color image. Appearance representation in subspaces is invariant to the order of concatenation since this order only alters the order of values in the principle vectors (dimensions) of the subspace [13]. However, since the linear combinations of the previous sections can be expected to differ between bands, a concatenated vector cannot be assumed to represent any single linear combination. Here, some of the results related to color-rank in [15] could lead to interesting results.

7 Illumination Manifold in Eigenspace

We are now equipped to analyze the dimensionality of illumination manifolds in eigenspaces. An eigenspace \mathcal{E} is an image subspace that is typically computed using the Karhunen-Lo eve transform [13]. The bases of \mathcal{E} are then the normalized eigenvectors of the covariance matrix computed from an image set, one that typically includes images of a large number of objects taken at different poses and illumination conditions during a learning (or training) stage [4]. Suppose the dimensionality of the eigenspace is d . The eigenvectors $\mathbf{e}_s, s = 1, 2, \dots, d$ are those with the largest eigenvalues λ_s of the covariance matrix, such that, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. Our main concern here is the analysis of the illumination manifold of an object, i.e. the projections in eigenspace \mathcal{E} of images of an object taken under different source directions, for a fixed pose.

Scale and brightness normalizations are applied to all object images before they are either used to construct eigenspace representations or to recognize novel object images [4]. The scale normalization ensures that both appearance representation and recognition are invariant to the magnification of the imaging system under weak-perspective projection. The brightness normalization is used to achieve invariance to the intensity of illumination. As a result of brightness normalization, all images lie on a unit ball in a high but finite dimensional Hilbert space.

We show here that the linear combination of (8) enables us to use the projections of the three basis images ($\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3$) in a d -dimensional eigenspace to construct the entire illumination manifold (projections due to all other source directions) without the use of any additional measurements. First, let us define the normalized basis images as $\mathbf{i}_1 = \mathbf{I}_1/m_1$, $\mathbf{i}_2 = \mathbf{I}_2/m_2$, and $\mathbf{i}_3 = \mathbf{I}_3/m_3$, where, $m_i = \|\mathbf{I}_i\|$. Further, we define the magnitude matrix \mathbf{M} to be a 3×3 diagonal matrix with the m_i as its three diagonal elements, and the normalized image matrix as $\mathbf{N} = [\mathbf{i}_1 \mathbf{i}_2 \mathbf{i}_3]$. Using (8), a novel image can be written as:

$$\begin{aligned} \mathbf{I} &= \alpha_1 m_1 \mathbf{i}_1 + \alpha_2 m_2 \mathbf{i}_2 + \alpha_3 m_3 \mathbf{i}_3 \\ &= \mathbf{N} \mathbf{M} \boldsymbol{\Phi} \end{aligned} \tag{12}$$

The magnitude of the novel image is related to its source coefficient vector as:

$$m = \sqrt{\boldsymbol{\Phi}^T \mathbf{N}^T \mathbf{M}^T \mathbf{M} \mathbf{N} \boldsymbol{\Phi}} \tag{13}$$

The normalized novel image can now be expressed as a combination of the three normalized basis images:

$$\mathbf{i} = \frac{1}{m} \mathbf{N} \mathbf{M} \boldsymbol{\Phi} \tag{14}$$

Now, the projections of the three normalized basis images in a d -dimensional eigenspace are:

$$\begin{aligned}\mathbf{g}_1 &= [\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_d]^T \mathbf{i}_1 \\ \mathbf{g}_2 &= [\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_d]^T \mathbf{i}_2 \\ \mathbf{g}_3 &= [\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_d]^T \mathbf{i}_3\end{aligned}\tag{15}$$

Likewise, the projection of a novel image is:

$$\mathbf{g} = [\mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_d]^T \mathbf{i}\tag{16}$$

If we define the basis projection matrix as $\mathbf{G} = [\mathbf{g}_1 \mathbf{g}_2 \mathbf{g}_3]$, the expressions (14), (15), and (16) give:

$$\mathbf{g} = \frac{1}{m} \mathbf{G} \mathbf{M} \Phi\tag{17}$$

Therefore, given the three eigenspace projections corresponding to the basis illuminants, we can determine the projection for any novel source \mathbf{s} from its coefficient vector Φ . If the basis and novel images are not normalized, the illumination manifold spans a linear subspace $\mathcal{G} = \mathbf{g}(\mathbf{s})$ whose dimensionality is 3 for a Lambertian surface of arbitrary texture, irrespective of the dimensionality d of the eigenspace used. When the basis and novel images are brightness normalized, the illumination manifold spans a nonlinear subspace of dimensionality 2. In this case, the nonlinearity arises from the normalization procedure, and the reduction in dimensionality by one results from all normalized images being constrained to lie on a unit ball in the Hilbert space.

The above results imply that we do not need to take a large number of images by sampling the entire illumination space for each pose of each object as done in [4]. For any given object pose, the entire illumination manifold can be constructed from the projections of just the three basis images.

It is worth reiterating that the above results are valid only for Lambertian surfaces and not much can be stated regarding the dimensionality of the illumination manifold for surfaces with nonlinear reflectance functions. This becomes intuitive when one considers the extreme case of a pure specular object that only produces highlights for each of the basis sources. In this case, the image produced by any novel source cannot in general be expressed as a linear combination of any number of basis images. Fortunately, there does exist a class of real-world objects that closely approximate Lambertian reflectance. For such objects, the above results prove useful as shown in the following sections.

8 Recognition of Novel Object Images

We have assumed thus far that the eigenspace \mathcal{E} is known a-priori. In practice, such an eigenspace is computed from a large image set obtained by varying pose and illumination in small increments [4]. Given the above results, it is possible to dramatically reduce the number of images that are required during learning. We now need to vary object pose in small increments and take only three images for each pose corresponding to the independent basis illuminants¹. All of the acquired images are used to compute \mathcal{E} . It is assumed here that \mathcal{E} is less sensitive to illumination variations than pose variations. This is typically the case when dealing with objects of complex shape and textural properties (see [3]).

Next, all images of an object are projected to eigenspace. The expression in (17) can be used to compute eigenspace projections corresponding to any desired number of source directions. In [4], the projections due to both pose and illumination variations are interpolated and the resulting manifold is densely resampled. The resulting points are stored in a database and serve as a discrete appearance representation of the object. Given a novel image, a segmentation algorithm is used to extract object regions. Each object region is normalized in scale and brightness and projected to eigenspace. A nearest neighbor algorithm is then used to identify the object. The exact location of the closest point on the object's appearance model determines the pose and illumination of the object in the image.

The idea of storing each object as a large number of densely resampled manifold points is practical only when the number of objects is small. In the case of large object databases, this approach proves inefficient both in terms of storage requirements as well as search time. Below, we present an algorithm that can be used to find the closest point on an illumination manifold directly from the three basis projections without ever constructing the illumination manifold. The algorithm is based on the observation that since the illumination manifold can be expressed in terms of three basis projections, this expression can in turn be used to determine if a novel eigenspace projection lies on the manifold. A variety of optimization schemes can be used to implement the above test. The approach outlined below is the one we have used in our experiments.

Given a scale normalized novel object image \mathbf{I}' , it is first normalized in brightness to get $\mathbf{i}' = \mathbf{I}'/m'$, where $m' = \|\mathbf{I}'\|$. If the novel image does lie on a particular illumination

¹The basis illuminants must be chosen with some care to ensure that they are representative of appearance variations due to illumination. For instance, three sources that form a compact cluster in the physical world, even if independent, may not in practice be able to accurately predict appearances due to sources distant from the cluster.

manifold, its projection \mathbf{g}' in eigenspace must satisfy:

$$\mathbf{g}' = \frac{1}{m'} \mathbf{G} \mathbf{M} \Phi \quad (18)$$

This expression gives us d equation with just 3 unknowns, namely, the source coefficients in Φ . Given that $10 \leq d \leq 30$ in most previous applications of parametric eigenspaces [4] [9] [5], what we have above is an overdetermined linear system that is easily solved to obtain an estimate $\tilde{\Phi}$ of the source coefficients.

Since the object in the novel image is unknown, the estimate $\tilde{\Phi}$ may or may not correspond to a point on the illumination manifold. A simple test can be employed to verify the validity of $\tilde{\Phi}$ by checking if the eigenspace projection of the novel image matches the projection corresponding to $\tilde{\Phi}$. To this end, we define the error measure:

$$\varepsilon = \left\| \mathbf{g}' - \frac{1}{\tilde{m}} \mathbf{G} \mathbf{M} \tilde{\Phi} \right\| \quad (19)$$

where, the magnitude \tilde{m} is computed by using $\tilde{\Phi}$ in equation (13). In theory, if the novel image does belong to the illumination manifold in question, we have $\varepsilon = 0$. In practice, a threshold is applied to ε to determine if the novel projection is close enough to the illumination manifold to be assumed to belong to it. We therefore have a simple and efficient algorithm that uses only three basis projections on an illumination manifold to check if a novel projection belongs to the manifold.

How do we use this algorithm when both pose and illumination are unknown in the recognition task? In such cases, one of two approaches may be taken. In one, object pose is first determined by finding the closest pose manifold. The resulting pose estimate would correspond to one of the three basis illuminants. Then, the illumination of the object is determined by fixing pose and using the above algorithm. Since a natural decoupling of pose and illumination effects does not generally exist in the case of complex objects, this approach can only provide approximate pose and illumination estimates. A precise way of solving the problem is to construct the illumination manifold for each object pose using the three corresponding basis projections and expression (17). The entire manifold, that represents both pose and illumination variations, is then densely resampled and stored as a discrete one. During on-line recognition, a nearest neighbor algorithm is used to find the object and its pose and illumination parameters as described in [3].

9 Experiments

Two sets of experiments were conducted to verify the theoretical results presented above. The first is related to the efficiency of learning; the construction of the illumination manifold of an object in given pose. The second pertains to real-time recognition; the test that determines whether a novel eigenspace projection lies on a given illumination manifold.

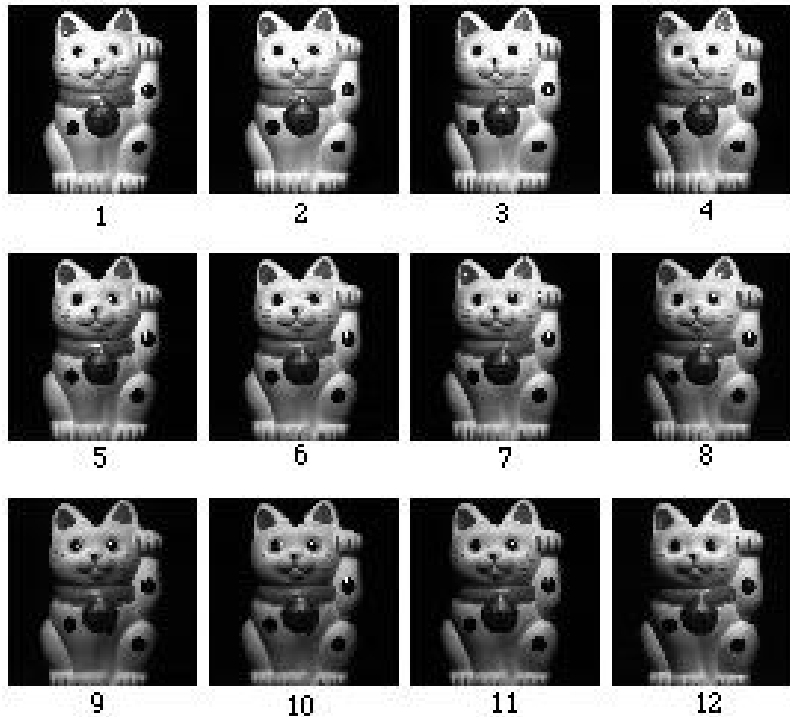


Figure 2: The image set used in the experiments. The object is primarily diffuse in reflectance, has patches with unknown albedo values, and includes concavities that produce interreflections. Image 2, 9, and 12 correspond to the basis illuminants and were used to determine the structure of the illumination manifold. The remaining 9 images were used to test the accuracies of the theoretical manifold and recognition.

Figure 2 shows 12 images of a complex object taken under known illumination directions. The object is more or less diffuse in reflectance, has surface patches with different albedo values, and includes concavities that cause interreflections. All the experiments were conducted in an eigenspace that was precomputed for a larger set of objects. The source directions are expressed as (θ, ϕ) , the azimuth and polar angles subtended by the source in an object centered coordinate frame with its z -axis pointing towards the image sensor.

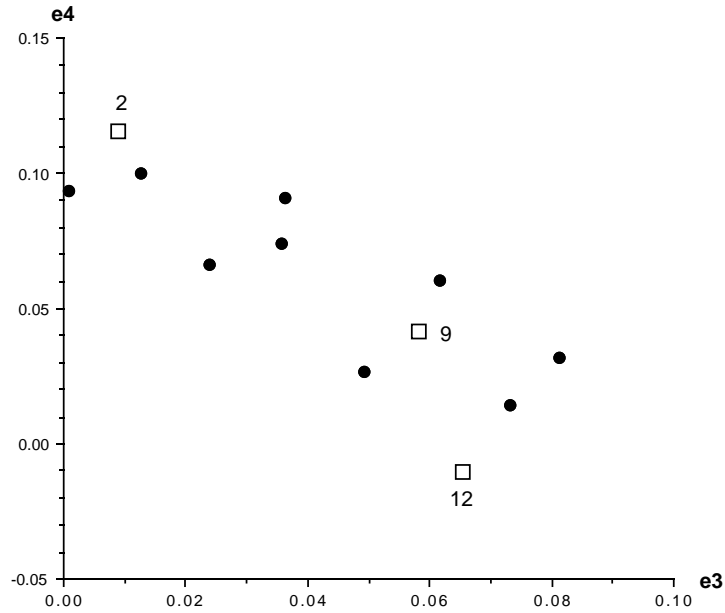


Figure 3: Projections of the basis images 2, 9, and 12 (shown as boxes) and the remaining 9 test images (shown as dots) in a 2D subspace of a 10D eigenspace.

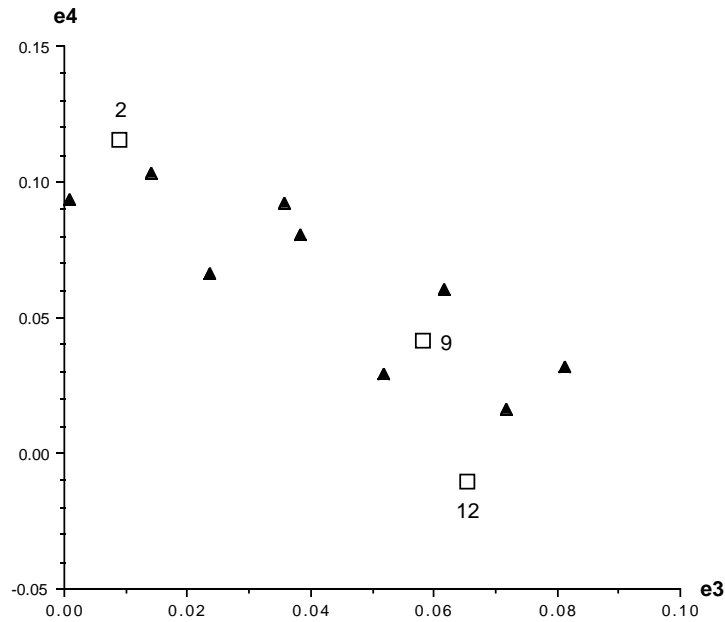


Figure 4: Projections of the three basis images 2, 9, and 12 (shown as boxes) and theoretical projections of the 9 test images (shown as triangles) determined directly from their known source directions (using expression 17). These projections can be compared with the actual image projections in Figure 3 to evaluate the accuracy of the theoretical illumination manifold. Corresponding projections in the two plots appear in similar positions with respect to the basis projections.

Three of the 12 images, namely, images 2, 9, and 12, were used to determine the structure of the illumination manifold. The directions of the basis illuminants are $(-9.1^\circ, 81.0^\circ)$, $(17.7^\circ, 81.3^\circ)$, and $(0.0^\circ, 64.4^\circ)$, respectively. The remaining 9 images in Figure 2 were used to test the accuracy of the illumination manifold. Figure 3 shows projections of the 9 images in eigenspace. For display, the projections are shown in a 2D subspace of a 10D eigenspace. The known source directions for the 9 test images were used in expression (17) to determine theoretical predictions of their locations on the illumination manifold. These projections are shown in Figure 4. We see that the actual projections of the 9 images and their theoretical predictions are in strong agreement, demonstrating the accuracy of the illumination manifold derived from just three basis projections. The slight discrepancies in the plots could have resulted from several factors; the object is not purely Lambertian, includes a few specular patches, and produces self-shadows that are not accounted for in the theory.

IMAGE NUMBER	ACTUAL SOURCE DIRECTION (θ, ϕ)	ESTIMATED SOURCE DIRECTION (θ, ϕ)	ERROR IN EIGENSPACE PROJECTION
1	(-9.1, 64.6)	(-8.2, 63.3)	0.0057
3	(9.1, 64.6)	(9.6, 66.3)	0.0181
4	(17.7, 65.4)	(19.5, 69.3)	0.0486
5	(-9.1, 72.5)	(-7.5, 72.2)	0.0106
6	(0.0, 72.3)	(-0.8, 72.6)	0.0169
7	(9.1, 72.5)	(7.8, 73.6)	0.0147
8	(17.7, 73.1)	(15.5, 75.8)	0.0098
10	(0.0, 80.9)	(-0.8, 78.3)	0.0103
11	(9.1, 81.0)	(9.5, 78.9)	0.0069

Table 1: Results of illumination direction estimation using just three basis projections in eigenspace.

The second experiment involves the recognition of novel images and the estimation of illumination direction using just three basis projections. Table 1 compares the actual illumination directions used to take the test images with the directions estimated using the algorithm described in section 8. We see that the estimates are very accurate. Also shown in the table are the distances in eigenspace between predicted and actual projections.

Acknowledgements

The authors would like to thank Brian Funt of Simon Fraser University for references and discussions on the use of illumination spaces in color analysis, and Simon Baker of Columbia University for his detailed comments on an early draft that have helped improve the paper. This research was conducted at the NTT Basic Research Laboratory, Atsugi, Japan, and at the Center for Research in Intelligent Systems, Department of Computer Science, Columbia University. It was supported in part by the NTT Basic Research Laboratory and in part by an NSF National Investigator Award. The authors thank Ken-ichiro Ishii of the NTT Basic Research Lab. for his support of this work.

References

- [1] R. Epstein, P. W. Hallinan, and A. L. Yuille, “ 5 ± 2 Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models,” *Proc. of IEEE Workshop on Physics Based Modeling in Computer Vision*, pp. 108-116, Boston, June 1995.
- [2] B. V. Funt and M. S. Drew, “Color Space Analysis of Mutual Illumination,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 12, pp. 1319-1325, December 1993.
- [3] H. Murase and S. K. Nayar, “Visual Learning and Recognition of 3D Objects from Appearance,” *International Journal of Computer Vision*, Vol. 14, No. 1, pp. 5-24, January, 1995.
- [4] H. Murase and S. K. Nayar, “Learning and Recognition of 3D Objects from Appearance,” *Proc. of IEEE Workshop on Qualitative Vision*, pp. 39-50, June 1993.
- [5] H. Murase and S. K. Nayar, “Illumination Planning for Object Recognition in Structured Environments,” *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition*, Seattle, pp. 31-38, June 1994.
- [6] S. K. Nayar, S. A. Nene, and H. Murase, “Real-Time 100 Object Recognition System,” CUUCS-025-95, Technical Report, Department of Computer Science, Columbia University, New York, September, 1995.
- [7] S. K. Nayar and M. Oren, “Visual Appearance of Matte Surfaces,” *SCIENCE*, Vol. 267, pp. 1153-1156, February 1995.

- [8] S. K. Nayar and H. Murase, "On the Dimensionality of Illumination Manifolds in Eigenspace," CUCS-021-94, Technical Report, Department of Computer Science, Columbia University, New York, August 1994.
- [9] S. K. Nayar, H. Murase, and S. A. Nene, "Learning, Positioning, and Tracking Visual Appearance," *Proc. of IEEE Intl. Conf. on Robotics and Automation*, San Diego, May 1994.
- [10] S. K. Nayar and Y. Gong, "Colored Interreflections and Shape Recovery," *Proc. of DARPA Image Understanding Workshop*, San Diego, CA, 1992.
- [11] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Surface Reflection: Geometrical and Physical Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp. 611-634, July 1991.
- [12] S. K. Nayar, K. Ikeuchi, and T. Kanade, "Shape from Interreflections," *International Journal of Computer Vision*, Vol. 2, No. 3, pp. 173-195, 1991.
- [13] E. Oja, *Subspace methods of Pattern Recognition*, Research Studies Press, Hertfordshire, 1983.
- [14] M. Oren and S. K. Nayar, "Generalization of the Lambertian Model and Implications for Machine Vision," *International Journal of Computer Vision*, Vol. 14, No. 2-3, pp. 227-251, April, 1995.
- [15] A. P. Petrov, "Color and Grassman-Cayley coordinates of shape," in *Human Vision, Visual Processing and Digital Display II*, SPIE Proc., Vol. 1453, pp. 342-352, 1991.
- [16] A. Sashua, "On Photometric Issues in 3D Visual Recognition from a Single 2D Image," Technical Report, Artificial Intelligence Lab., MIT, 1993.
- [17] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.
- [18] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, Vol. 19, pp. 139-144, 1980.