Weizmann Institute of Science, Technical Report MCS00-21
NEC Research Institute Technical Report 2000-172R

Lambertian Reflectance and Linear Subspaces

Ronen Basri*

Dept. of Computer Science
The Weizmann Institute of Science
Rehovot, 76100 Israel

David Jacobs

NEC Research Institute
4 Independence Way
Princeton, NJ 08540

Abstract

We prove that the set of all reflectance functions (the mapping from surface normals to intensities) produced by Lambertian objects under distant, isotropic lighting lies close to a 9D linear subspace. This implies that, in general, the set of images of a convex Lambertian object obtained under a wide variety of lighting conditions can be approximated accurately by a low-dimensional linear subspace, explaining prior empirical results. We also provide a simple analytic characterization of this linear space. We obtain these results by representing lighting using spherical harmonics and describing the effects of Lambertian materials as the analog of a convolution. These results allow us to construct algorithms for object recognition based on linear methods as well as algorithms that use convex optimization to enforce non-negative lighting functions. Finally, we show a simple way to enforce non-negative lighting when the images of an object lie near a 4D linear space.

*Research conducted while at NEC Research Institute, Princeton, NJ. Ronen Basri is an incumbent of Arye Dissentshik Career Development Chair at the Weizmann Institute.

1 Introduction

One of the most basic problems in vision is to understand how variability in lighting affects the images that an object can produce. Even when lights are isotropic and distant, smooth Lambertian objects can produce infinite-dimensional sets of images (Belhumeur and Kriegman [1]). But recent *experimental* work ([7, 12, 30]) has indicated that the set of images produced by an object under a wide range of lighting conditions lies near a low dimensional linear subspace in the space of all possible images. This can be used to construct efficient recognition algorithms that handle lighting variations. In this paper we explain these empirical results analytically and use this understanding to produce new recognition algorithms.

When light is isotropic and distant from an object, we can describe its intensity as a function of direction. Light, then, is a non-negative function on the surface of a sphere. Our approach begins by representing these functions using *spherical harmonics*. This is analogous to Fourier analysis, but on the surface of the sphere. To model the way surfaces turn light into an image we look at reflectance as a function of the surface normal (assuming unit albedo). We show that reflectance functions are produced through the analog of a convolution of the lighting function using a kernel that represents Lambert's reflection. This kernel acts as a low-pass filter with 99.2% of its energy in the first nine components. We use this and the non-negativity of light to prove that under *any* lighting conditions, a nine-dimensional linear subspace, for example, accounts for 98% of the variability in the reflectance function. This suggests that in general the set of images of a convex, Lambertian object can be approximated accurately by a low dimensional linear space. We further show how to analytically derive this subspace from an object model.

This allows us to better understand several existing methods. For example, we show that the linear subspace methods of Shashua [25] and Moses [20] use a linear space spanned by the three first order harmonics, but that they omit the significant DC component. Also, it leads us to new methods of recognizing objects with unknown pose and lighting conditions. In particular, we discuss how the harmonic basis can be used in a linear-based object recognition algorithm, replacing bases derived by performing SVD on large collections of rendered images. Furthermore, we show how we can enforce non-negative light by projecting this constraint to the space spanned by the harmonic basis. With this constraint recognition is expressed as a non-negative least-squares problem that can be solved using convex optimization. This leads to an algorithm for recognizing objects under varying pose and illumination that resembles Georghides et al. [9], but works in an analytically derived low-dimensional space. The use of the harmonic basis, in this case, allows us to rapidly produce a representation to the images of an object in poses determined at runtime. Finally, we discuss the case in which a first order approximation provides an adequate approximation to the images of an object. The set of images then lies near a 4D linear subspace. In this case we can express the non-negative lighting constraint analytically. We use this expression to perform recognition in a particularly efficient way, without complex, iterative optimization techniques.

It has been very popular in object recognition to represent the set of images that an object can produce using low dimensional linear subspaces of the space of all images. Ullman and Basri [28] analytically derive such a representation for sets of 3D points undergoing scaled orthographic projection.

Shashua [25] and Moses [20] (and later also [22, 31]) derive a 3D linear representation of the set of images produced by a Lambertian object as lighting changes, but ignoring attached shadows. Hayakawa [13] uses factorization to build 3D models using this linear representation. Koenderink and van Doorn [18] extend this to a 4D space by allowing the light to include a diffuse component. Researchers have collected large sets of images and performed PCA to build representations that capture within class variations [16, 27, 4] and variations due to pose and lighting [21, 12, 30]. Hallinan [12], Epstein et al. [7] and Yuille et al. [30] perform experiments that show that large numbers of images of Lambertian objects, taken with varied lighting conditions, do lie near a low-dimensional linear space, justifying this representation. More recently, analytically derived, convex representations have been used by Belhumeur and Kriegman [1] to model attached shadows. Georghides et al. [8, 9] use this representation for object recognition.

Spherical harmonics have been used in graphics to efficiently represent the bidirectional reflection distribution function (BRDF) of different materials by, e.g., Cabral [3] and Westin et al. [29] (Koenderink and van Doorn [17] proposed replacing the spherical harmonics basis with the Zernike polynomials, since BRDFs are defined over a half sphere.) Nimeroff et al. [23]. Dobashi et al. [5] and Teo et al. [26] explore specific lighting configurations that can be represented efficiently as a linear combination of basis lightings (e.g., daylight). Dobashi et al. [5] in particular use spherical harmonics to form such a basis. D'Zmura [6] was first to point out that the process of turning incoming light into reflection can be described in terms of spherical harmonics. With this representation, after truncating high order components, the reflection process can be written as a linear transformation, and so the low order components of the lighting can be recovered by inverting the transformation. He used this analysis to explore ambiguities in lighting. We extend this work by deriving subspace results for the reflectance function, providing analytic descriptions of the basis images, and constructing new recognition algorithms that use this analysis while enforcing non-negative lighting. Independent of and contemporaneous with our work, Ramamoorthi and Hanrahan [24] have described the effect of Lambertian reflectance as a convolution. Like D'Zmura they use this analysis to explore the problem of recovering lighting from reflectances. Also, preliminary comments on this topic can be found in Jacobs, Belhumeur and Basri[15].

In summary, the main contribution of our paper is to show how to analytically find low dimensional linear subspaces that accurately approximate the set of images that an object can produce. We can then carve out portions of these subspaces corresponding to non-negative lighting conditions, and use these descriptions for recognition.

2 Modeling Image Formation

Consider a convex object illuminated by distant isotropic light sources. Assume further that the surface of the object reflects light according to Lambert's law [19]. This relatively simple model has been analyzed and used effectively in a number of vision applications. The set of images of a Lambertian object obtained with arbitrary light has been termed the "*illumination cone*" by Belhumeur and Kriegman [1]. Our objective is to analyze properties of the illumination cone. For the analysis it will be useful to consider the set of reflectance functions obtained under different illumination conditions.

A *reflectance function* (also called *reflectance map*, see Horn [14], Chapters 10-11) associated with a specific lighting configuration is defined as the light reflected by a sphere of unit albedo as a function of the surface normal. A reflectance function is related to an image of a convex object illuminated by the same lighting configuration by the following mapping. Every visible point on the object's surface inherits its intensity from the point on the sphere with the same normal, and this intensity is further scaled by the albedo at the point. We will discuss the effect of this mapping later on in this section.

2.1 Image Formation as the Analog of a Convolution

Let S denote a unit sphere centered at the origin. Let $p = (x, y, z)$ denote a point on the surface of S, and let $N_p = (x, y, z)$ denote the surface normal at p. p can also be expressed as a unit vector using the following notation:

$$(x, y, z) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi), \tag{1}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. In this coordinate frame the poles are set at $(0, 0, \pm 1)$, θ denotes the solid angle between p and $(0, 0, 1)$, and it varies with latitude, and ϕ varies with longitude. Since we assume that the sphere is illuminated by a distant and isotropic set of lights all points on the sphere see these lights coming from the same directions, and they are illuminated by identical lighting conditions. Consequently, the configuration of lights that illuminate the sphere can be expressed as a non-negative function $\ell(\theta, \phi)$, expressing the intensity of the light reaching the sphere from each direction (θ, ϕ). Furthermore, according to Lambert's law the difference in the light reflected by the points is entirely due to the difference in their surface normals. Thus, we can express the light reflected by the sphere as a function $r(\theta, \phi)$ whose domain is the set of surface normals of the sphere.

According to Lambert's law, if a light ray of intensity l reaches a surface point with albedo λ forming an angle θ with the surface normal at the point, then the intensity reflected by the point due to this light is given by

$$l\lambda \max(\cos\theta, 0). \tag{2}$$

In a reflectance function we use $\lambda = 1$. If light reaches a point from a multitude of directions then the light reflected by the point would be the sum of (or in the continuous case the integral over) the contribution for each direction. Denote by $k(\theta) = \max(\cos\theta, 0)$, then, for example, the intensity of the point $(0, 0, 1)$ is given by:

$$r(0, 0) = \int_0^{2\pi} \int_0^{\pi} k(\theta)\ell(\theta, \phi) \sin\theta d\theta d\phi. \tag{3}$$

Similarly, the intensity $r(\theta, \phi)$ reflected by a point $p = (\theta, \phi)$ is obtained by centering k about p and integrating its inner product with ℓ over the sphere. Thus, the operation that produces $r(\theta, \phi)$ is the analog of a convolution on the sphere. We will refer to this as a convolution, and write:

$$r(\theta, \phi) = k * \ell. \tag{4}$$

The kernel of this convolution, k, is the circularly symmetric, "half-cosine" function. The convolution is obtained by rotating k so that its center is aligned with the surface normal at p. This still leaves one degree of freedom in the rotation of the kernel undefined, but since k is rotationally symmetric this ambiguity disappears.

2.2 Properties of the Convolution Kernel

Just as the Fourier basis is convenient for examining the results of convolutions in the plane, similar tools exist for understanding the results of the analog of convolutions on the sphere. The *surface spherical harmonics* are a set of functions that form an orthonormal basis for the set of all functions on the surface of the sphere. We denote these functions by h_{nm}, with $n = 0, 1, 2, ...$ and $-n \leq m \leq n$:

$$h_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_{nm}(\cos\theta) e^{im\phi}, \tag{5}$$

where P_{nm} are the *associated Legendre functions*, defined as

$$P_{nm}(z) = \frac{(1-z^2)^{m/2}}{2^n n!} \frac{d^{n+m}}{dz^{n+m}} (z^2 - 1)^n. \tag{6}$$

In the course of this paper it will sometimes be convenient to parameterize h_{nm} as a function of space coordinates (x, y, z) rather than angles. The spherical harmonics, written $h_{nm}(x, y, z)$, then become polynomials of degree n in (x, y, z).

We may express the kernel, k, and the lighting function, ℓ, as harmonic series, that is, as linear combinations of the surface harmonics. We do this primarily so that we can take advantage of the analog to the convolution theorem for surface harmonics. An immediate consequence of the Funk-Hecke theorem (see, e.g., [10], Theorem 3.4.1, page 98) is that "convolution" in the function domain is equivalent to multiplication in the harmonic domain. In the rest of this section we derive a representation of k as a harmonic series. We use this derivation to show that k is nearly a low-pass filter. Specifically, almost all of the energy of k resides in the first few harmonics. This will allow us to show that the possible reflectances of a sphere all lie near a low dimensional linear subspace of the space of all functions defined on the sphere.

In Appendix A we derive a representation of k as a harmonic series. In short, since k is rotationally symmetric about the pole, under an appropriate choice of a coordinate frame its energy concentrates exclusively in the *zonal harmonics* (the harmonics with $m = 0$), while the coefficients of all the harmonics with $m \neq 0$ vanish. Thus, we can express k as:

$$k = \sum_{n=0}^{\infty} k_n h_{n0}, \tag{7}$$

with

$$k_n = \int_0^{2\pi} \int_0^{\pi} k(\theta) h_{n0}(\theta, \phi) \sin\theta d\theta d\phi. \tag{8}$$

After some tedious manipulation (detailed in Appendix A) we obtain that

$$k_n = \begin{cases} \frac{\sqrt{\pi}}{2} & n = 0 \\ \sqrt{\frac{\pi}{3}} & n = 1 \\ (-1)^{\frac{n}{2}+1} \frac{(n-2)!\sqrt{(2n+1)\pi}}{2^n (\frac{n}{2}-1)!(\frac{n}{2}+1)!} & n \geq 2, \text{ even} \\ 0 & n \geq 2, \text{ odd} \end{cases} \tag{9}$$

n	0	1	2	4	6	8
Energy	37.5	50	11.72	0.59	0.12	0.04
Cumulative energy	37.5	87.5	99.22	99.81	99.93	99.97
Lower bound	37.5	75	97.96	99.48	99.80	99.90

Table 1: The top row shows the energy captured by the n'th zonal harmonic for the Lambertian kernel ($0 \leq n \leq 8$). The middle row shows the energy accumulated up to order n. This energy represents the quality of the n'th order approximation of $r(\theta, \phi)$ (measured in relative squared error). The bottom row shows a lower bound on the quality of this approximation due to the non-negativity of the light. The $n = 3$, 5, and 7 are omitted because they contribute no energy. Relative energies are given in percents.

The first few coefficients, for example, are

$$\begin{array}{lllllllll} k_0 & = & \frac{\sqrt{\pi}}{2} \approx 0.8862 & \quad k_2 & = & \frac{\sqrt{5\pi}}{8} \approx 0.4954 & \quad k_6 & = & \frac{\sqrt{13\pi}}{128} \approx 0.0499 \\ k_1 & = & \sqrt{\frac{\pi}{3}} \approx 1.0233 & \quad k_4 & = & -\frac{\sqrt{\pi}}{16} \approx -0.1108 & \quad k_8 & = & \frac{\sqrt{17\pi}}{256} \approx -0.0285. \end{array} \tag{10}$$

($k_3 = k_5 = k_7 = 0$) A graph representation of the coefficients is shown in Figure 1.

The energy captured by every harmonic term is measured commonly by the square of its respective coefficient divided by the total squared energy of the transformed function. The total squared energy in the half cosine function is given by

$$\int_0^{2\pi} \int_0^{\pi} k^2(\theta) \sin\theta d\theta d\phi = 2\pi \int_0^{\frac{\pi}{2}} \cos^2\theta \sin\theta d\theta = \frac{2\pi}{3}. \tag{11}$$

Table 1 shows the relative energy captured by each of the first several coefficients. It can be seen that the kernel is dominated by the first three coefficients. Thus, a second order approximation already accounts for 99.22% of the energy. With this approximation the half cosine function can be written as:

$$k(\theta) \approx \frac{1}{4} + \frac{1}{2}\cos\theta + \frac{5}{16}\cos 2\theta. \tag{12}$$

The quality of the approximation improves somewhat with the addition of the fourth order term (99.81%) and deteriorates to 87.5% when a first order approximation is used. Figure 2 shows a 1D slice of the Lambertian kernel and its various approximations.

2.3 Approximating the Reflectance Function

The fact that the Lambertian kernel has most of its energy concentrated in the low order terms implies that the set of Lambertian reflectance functions can be well approximated by a low dimensional linear space. This space is spanned by a small set of what we call *harmonic reflectances*. The harmonic reflectance $r_{nm}(\theta, \phi)$ denotes the reflectance of the sphere when it is illuminated by the harmonic "light" h_{nm}. Note that harmonic lights generally are not positive everywhere, so they do not correspond to real, physical lighting conditions; they are abstractions. As is explained below every reflectance function

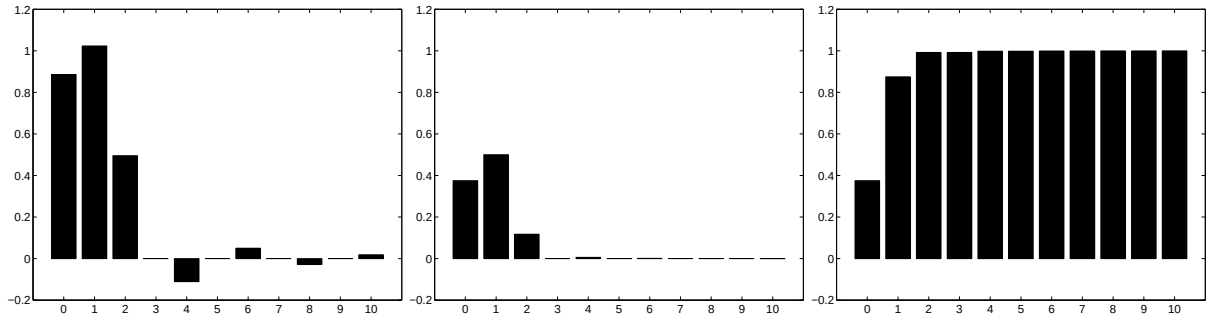

Figure 1: From left to right: a graph representation of the first 11 coefficients of the Lambertian kernel, the relative energy captured by each of the coefficients, and the accumulated energy

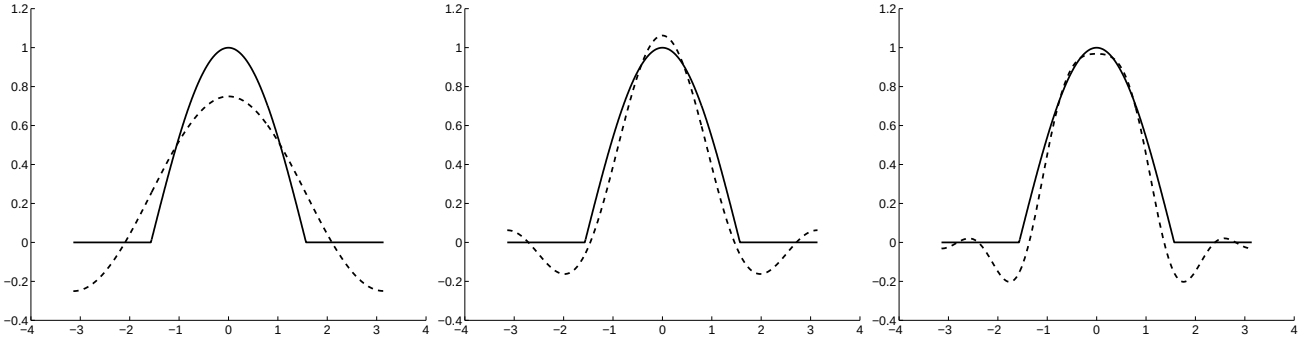

Figure 2: A slice of the Lambertian kernel (solid) and its approximations of first (left, dotted), second (middle), and fourth (right) order.

$r(\theta, \phi)$ will be approximated to an excellent accuracy by a linear combination of a small number of harmonic reflectances.

To evaluate the quality of the approximation consider first, as an example, lighting generated by a point source at the z direction ($\theta = \phi = 0$). A point source is a delta function. The reflectance of a sphere illuminated by a point source is obtained by a convolution of the delta function with the kernel, which results in the kernel itself. Due to the linearity of the convolution, if we approximate the reflectance due to this point source by a linear combination of the first three zonal harmonics, r_{00}, r_{10}, and r_{20}, we account for 99.22% of the energy. In other words

$$\min_{(a_0,a_1,a_2)} \frac{\|a_0 r_{00} + a_1 r_{10} + a_2 r_{20} - k\|^2}{\|k\|^2} = 0.9922, \tag{13}$$

where k, the Lambertian kernel, is also the reflectance of the sphere when it is illuminated by a point source at the z direction. Similarly, first and fourth order approximations yield respectively 87.5% and 99.81% accuracy.

If the sphere is illuminated by a single point source in a direction other than the z direction the reflectance obtained would be identical to the kernel, but shifted in phase. Shifting the phase of a function distributes its energy between the harmonics of the same order n (varying m), but the overall energy in each n is maintained. The quality of the approximation, therefore, remains the same, but now for an N'th order approximation we need to use all the harmonics with $n \leq N$ for all m. Recall that there are $2n + 1$ harmonics in every order n. Consequently, a first order approximation requires four harmonics. A second order approximation adds five more harmonics yielding a 9D space. The third order harmonics are eliminated by the kernel, and so they do not need to be included. Finally, a fourth order approximation adds nine more harmonics yielding an 18D space.

We have seen that the energy captured by the first few coefficients k_i ($1 \leq i \leq N$) directly indicates the accuracy of the approximation of the reflectance function when the light includes a single point source. Other light configurations may lead to different accuracy. Better approximations are obtained when the light includes enhanced diffuse components of low-frequency. Worse approximations are anticipated if the light includes mainly high frequency patterns.

However, even if the light includes mostly high frequency patterns the accuracy of the approximation is still very high. This is a consequence of the non-negativity of light. A lower bound on the accuracy of the approximation for *any* light function can be derived as follows. It is simple to show that for any non-negative function the amplitude of the DC component must be at least as high as the amplitude of any of the other components.[1] One way to see this is by representing such a function as a non-negative sum of delta functions. In such a sum the amplitude of the DC component is the weighted sum of the amplitudes of all the DC components of the different delta functions. The amplitude of any other frequency may at most reach the same level, but often will be lower due to interference. Consequently, in an N'th order approximation the worst scenario is obtained when the amplitudes in all frequencies

[1]Note that to obtain the amplitude of the n'th component we must normalize its coefficient, multiplying it by $\sqrt{\frac{4\pi}{2n+1}}$. Consequently the coefficient of the DC component may be smaller than that of other components, while the amplitude may not. The Funk-Hecke theorem applies to the amplitudes.

higher than N saturate to the same amplitude as the DC component, while the amplitude of orders $1 \leq n \leq N$ are set to zero. In this case the relative squared energy becomes

$$\frac{k_0^2}{k_0^2 + \sum_{n=N+1}^{\infty} k_n^2} = \frac{k_0^2}{\frac{2\pi}{3} - \sum_{n=1}^{N} k_n^2}. \tag{14}$$

Table 1 shows the bound obtained for several different approximations. It can be seen that using a second order approximation (involving nine harmonics) the accuracy of the approximation for any light function exceeds 97.96%. With a fourth order approximation (involving 18 harmonics) the accuracy exceeds 99.48%. Note that the bound computed in (14) is not tight, since the case that all the higher order terms are saturated yields a function with negative values. Consequently, the worst case accuracy may even be higher than the bound.

2.4 Generating Harmonic Reflectances

Constructing a basis to the space that approximates the reflectance functions is straightforward and can be done analytically. To construct the basis we can simply invoke the Funk-Hecke theorem. Recall that this space is spanned by the harmonic reflectances, i.e., the reflectances obtained when a unit albedo sphere is illuminated by harmonic lights. These reflectances are the result of convolving the half cosine kernel with single harmonics. Due to the orthonormality of the spherical harmonics such a convolution cannot produce energy in any of the other harmonics. Consequently, denote the harmonic light by h_{nm}, then the reflectance due to this harmonic is the same harmonic, but scaled. Formally,

$$r_{nm} = k * h_{nm} = c_n h_{nm}. \tag{15}$$

(It can be readily verified that the harmonics of the same order n but different phase m share the same scale factor c_n.) It is therefore left to determine c_n.

To determine c_n (which is important when we enforce non-negative lighting in Sections 3.2 and 3.3) we can use the fact that the half-cosine kernel k is an image obtained when the light is a delta function centered in the z direction. The transform of the delta function is given by

$$\delta = \sum_{n=0}^{\infty} \sqrt{\frac{2n+1}{4\pi}} h_{n0}, \tag{16}$$

and the image it produces is

$$k = \sum_{n=0}^{\infty} k_n h_{n0}, \tag{17}$$

where the coefficients k_n are given in (9). c_n determines by how much the harmonic is scaled following the convolution; therefore, it is the ratio between k_n and the respective coefficient of the delta function, that is,

$$c_n = \sqrt{\frac{4\pi}{2n+1}} k_n. \tag{18}$$

The first few harmonic reflectances are given by

$$\begin{array}{lllllllll} r_{00} & = & \pi h_{00} & & r_{2m} & = & \frac{\pi}{4} h_{2m} & & r_{6m} & = & \frac{\pi}{64} h_{6m} \\ r_{1m} & = & \frac{2\pi}{3} h_{1m} & & r_{4m} & = & \frac{\pi}{24} h_{4m} & & r_{8m} & = & \frac{\pi}{128} h_{8m} \end{array} \tag{19}$$

for $-n \leq m \leq n$ (and $r_{3m} = r_{5m} = r_{7m} = 0$).

For the construction of the harmonic reflectances it is useful to express the harmonics using space coordinates (x, y, z) rather than angles (θ, ϕ). This can be done by substituting the following equations for the angles:

$$\begin{array}{lll} \theta & = & \cos^{-1} z \\ \phi & = & \tan^{-1} \frac{y}{x}. \end{array} \tag{20}$$

The first nine harmonics then become

$$\begin{array}{lllllllll} h_{00} & = & \frac{1}{\sqrt{4\pi}} & & h^o_{11} & = & \sqrt{\frac{3}{4\pi}} y & & h^o_{21} & = & 3\sqrt{\frac{5}{12\pi}} yz \\ h_{10} & = & \sqrt{\frac{3}{4\pi}} z & & h_{20} & = & \frac{1}{2}\sqrt{\frac{5}{4\pi}}(2z^2 - x^2 - y^2) & & h^e_{22} & = & \frac{3}{2}\sqrt{\frac{5}{12\pi}}(x^2 - y^2) \\ h^e_{11} & = & \sqrt{\frac{3}{4\pi}} x & & h^e_{21} & = & 3\sqrt{\frac{5}{12\pi}} xz & & h^o_{22} & = & 3\sqrt{\frac{5}{12\pi}} xy, \end{array} \tag{21}$$

where the superscripts e and o denote the even and the odd components of the harmonics respectively (so $h_{nm} = h^e_{n|m|} \pm i h^o_{n|m|}$, according to the sign of m; in fact the even and odd versions of the harmonics are more convenient to use in practice since the reflectance function is real). Notice that the harmonics are simply polynomials in these space coordinates. Below we invariably use $h_{nm}(\theta, \phi)$ and $h_{nm}(x, y, z)$ to denote the harmonics expressed in angular and space coordinates respectively.

2.5 From Reflectances to Images

Up to this point we have analyzed the reflectance functions obtained by illuminating a unit albedo sphere by arbitrary light. Our objective is to use this analysis to efficiently represent the set of images of objects seen under varying illumination. An image of an object under certain illumination conditions can be constructed from the respective reflectance function in a simple way: each point of the object inherits its intensity from the point on the sphere whose normal is the same. This intensity is further scaled by its albedo. In other words, given a reflectance function $r(x, y, z)$, the image of a point p with surface normal $n = (n_x, n_y, n_z)$ and albedo λ is given by

$$I(p) = \lambda r(n_x, n_y, n_z). \tag{22}$$

We now wish to discuss how the accuracy of our low dimensional linear approximation to a model's images can be affected by the mapping from the reflectance function to images. We will make two points. First, in the worst case, this can make our approximation arbitrarily bad. Second, in typical cases it will not make our approximation less accurate.

There are two components to turning a reflectance function into an image. One is that there is a rearrangement in the x, y position of points. That is, a particular surface normal appears in one

location on the unit sphere, and may appear in a completely different location in the image. This rearrangement has no effect on our approximation. We represent images in a linear subspace in which each coordinate represents the intensity of a pixel. The decision as to which pixel to represent with which coordinate is arbitrary, and changing this decision by rearranging the mapping from (x, y) to a surface normal merely reorders the coordinates of the space.

The second and more significant difference between images and reflectance functions is that occlusion, shape variation and albedo variations affect the extent to which each surface normal on the sphere helps determine the image. For example, occlusion ensures that half the surface normals on the sphere will be facing away from the camera, and will not produce any visible intensities. A discontinuous surface may not contain some surface normals, and a surface with planar patches will contain a single normal over an extended region. In between these extremes, the curvature at a point will determine the extent to which its surface normal contributes to the image. Albedo has a similar effect. If a point is black (zero albedo) its surface normal has no effect on the image. In terms of energy, darker pixels contribute less to the image than brighter pixels. Overall, these effects are captured by noticing that the extent to which the reflectance of each point on the unit sphere influences the image can range from zero to the entire image.

We will give an example to show that in the worst case this can make our approximation arbitrarily bad. First, one should notice that at any single point, a low-order harmonic approximation to a function can be arbitrarily bad (this can be related to the Gibbs phenomenon in the Fourier domain). Consider the case of an object that is a sphere of constant albedo (this example is adapted from Belhumeur and Kriegman [1]). If the light is coming from a direction opposite the viewing direction, it will not illuminate any visible pixels. We can then shift the light slightly, so that it illuminates just one pixel on the boundary of the object; by varying the intensity of the light we can give this pixel any desired intensity. A series of lights can do this for every pixel on the rim of the sphere. If there are n such pixels, the set of images we get fully occupies the positive orthant of an n-dimensional space. Obviously, points in this space can be arbitrarily far from any 9D space. What is happening is that all the energy in the image is concentrated in those surface normals for which our approximation happens to be poor.

However, generally, things will not be so bad. In general, occlusion will render an arbitrary half of the normals on the unit sphere invisible. Albedo variations and curvature will emphasize some normals, and deemphasize others. But in general, the normals whose reflectances are poorly approximated will not be emphasized more than any other reflectances, and we can expect our approximation of reflectances on the entire unit sphere to be about as good over those pixels that produce the intensities visible in the image.

Therefore, we assume that the subspace results for the reflectance functions carry on to the images of objects. Thus we approximate the set of images of an object by a linear space spanned by what we call *harmonic images*, denoted b_{nm}. These are images of the object seen under harmonic light. These images are constructed as in (22) as follows:

$$b_{nm}(p) = \lambda r_{nm}(n_x, n_y, n_z). \tag{23}$$

Note that b_{00} is an image obtained under constant, ambient light, and so it contains for every point simply the surface albedo at the point (scaled by a constant factor). The first order harmonic images

Figure 3: We show the first nine harmonic images for a model of a face. The top row contains the zero'th harmonic (left) and the three first order harmonic images (right). The second row shows the images derived from the second harmonics. Negative values are shown in black, positive values in white.

b_{1m} are images obtained under cosine lighting centered at the three main axes. These images are, for every point, the three components of the surface normals scaled by the albedos, and an additional constant. (See a discussion of past use of these images in Section 3.) The higher order harmonic images contain polynomials of the surface normals scaled by the albedo. Figure 3 shows the first nine harmonic images derived from a 3D model of a face.

We can write this more explicitly, combining Equations 21 and 23. Let p_i denote the i'th object point. Let λ denote a vector of the object's albedos, that is, λ_i is the albedo of p_i. Similarly, let $\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z}$ denote three vectors of the same length that contain the x, y and z components of the surface normal, so that, for example, $n_{x,i}$ (the i'th component of $\mathbf{n_x}$) is the x component of the surface normal of p_i. Further, let $\mathbf{n_{x^2}}$ denote a vector such that $n_{x^2,i} = n_{x,i}n_{x,i}$. We define $\mathbf{n_{y^2}}, \mathbf{n_{z^2}}, \mathbf{n_{xz}}, \mathbf{n_{yz}}, \mathbf{n_{xy}}$ similarly, where, for example, $n_{xy,i} = n_{x,i}n_{y,i}$. Finally, we will write $\lambda .* \mathbf{v}$ to denote the component-wise product of λ with any vector $\mathbf{v}$ (this is MATLAB's notation). That is, this product scales the components of a vector by the albedo associated with the point that produced that component. So $\lambda .* \mathbf{n_x}$ is just the x components of the surface normals scaled by their albedos. Using this notation, the first nine harmonic images become:

$$
\begin{array}{lll}
b_{00} = \frac{1}{\sqrt{4\pi}}\lambda & b^o_{11} = \sqrt{\frac{3}{4\pi}}\lambda .* \mathbf{n_y} & b^o_{21} = 3\sqrt{\frac{5}{12\pi}}\lambda .* \mathbf{n_{yz}} \\
b_{10} = \sqrt{\frac{3}{4\pi}}\lambda .* \mathbf{n_z} & b_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda .* (2\mathbf{n_{z^2}} - \mathbf{n_{x^2}} - \mathbf{n_{y^2}}) & b^e_{22} = \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda .* (\mathbf{n_{x^2}} - \mathbf{n_{y^2}}) \\
b^e_{11} = \sqrt{\frac{3}{4\pi}}\lambda .* \mathbf{n_x} & b^e_{21} = 3\sqrt{\frac{5}{12\pi}}\lambda .* \mathbf{n_{xz}} & b^o_{22} = 3\sqrt{\frac{5}{12\pi}}\lambda .* \mathbf{n_{xy}}.
\end{array}
\tag{24}
$$

3 Recognition

We have developed an analytic description of the linear subspace that lies near the set of images that an object can produce. We now show how to use this description to recognize objects. Although our method is suitable for general objects, we will give examples related to the problem of face recognition.

We assume that an image must be compared to a data base of models of 3D objects. We will assume that the pose of the object is already known, but that its identity and lighting conditions are not. For example, we may wish to identify a face that is known to be facing the camera. Or we may assume that either a human or an automatic system have identified features, such as the eyes and the tip of the nose, that allow us to determine pose for each face in the data base, but that the data base is too big to allow a human to select the best match.

Recognition proceeds by comparing a new image to each model in turn. To compare to a model we compute the distance between the image and the nearest image that the model can produce. We present two classes of algorithms that vary in their representation of a model's images. The linear subspace can be used directly for recognition, or we can restrict ourselves to a subset of the linear subspace that corresponds to physically realizable lighting conditions.

We will stress the advantages we gain by having an *analytic* description of the subspace available, in contrast to previous methods in which PCA could be used to derive a subspace from a sample of an object's images. One advantage of an analytic description is that we know this provides an accurate representation of an object's images, not subject to the vagaries of a particular sample of images. A second advantage is efficiency; we can produce a description of this subspace much more rapidly than PCA would allow. The importance of this advantage will depend on the type of recognition problem that we tackle. In particular, we are interested in recognition problems in which the position of an object is not known in advance, but can be computed at run-time using feature correspondences. In this case, the linear subspace must also be computed at run-time, and the cost of doing this is important. Finally, we will show that when we use a 4D linear subspace, an analytic description of this subspace allows us to incorporate the constraint that the lighting be physically realizable in an especially simple and efficient way.

3.1 Linear Methods

The most straightforward way to use our prior results for recognition is to compare a novel image to the linear subspace of images that correspond to a model (D'Zmura [6] also makes this suggestion). To do this, we produce the harmonic basis images of each model, as described in Section 2.5. Given an image I we seek a vector a that minimizes $\|Ba - I\|$, where B denotes the basis images, B is $p \times r$, p is the number of points in the image, and r is the number of basis images used. As discussed above, nine is a natural value to use for r, but $r = 4$ provides greater efficiency while $r = 18$ offers even better potential accuracy. Every column of B contains one harmonic image b_{nm}. These images form a basis for the linear subspace, though not an orthonormal one. So we apply a QR decomposition to B to obtain such a basis. We compute Q, a $p \times r$ matrix with orthonormal columns, and R, an $r \times r$ matrix so that $QR = B$ and $Q^T Q$ is an $r \times r$ identity matrix. We can then compute the distance from the image, I, and the space spanned by B as $\|QQ^T I - I\|$. The cost of the QR decomposition is $O(pr^2)$, assuming $p >> r$.

In contrast to this, prior methods have sometimes performed PCA on a sample of images to find a linear subspace representing an object. Hallinan [12] performed experiments indicating that PCA can produce a five or six dimensional subspace that accurately models a face. Epstein et al. [7] and Yuille

et al. [30] describe experiments on a wider range of objects that indicate that images of Lambertian objects can be approximated by a linear subspace of between three and seven dimensions. Specifically, the set of images of a basketball were approximated to 94.4% by a 3D space and to 99.1% by a 7D space, while the images of a face were approximated to 90.2% by a 3D space and to 95.3% by a 7D space. Georghides et al. [9] render the images of an object and find an 11D subspace that approximates these images.

These numbers are roughly comparable to the 9D space that, according to our analysis, approximates the images of a Lambertian object. Additionally, we note that the basis images of an object will not generally be orthogonal, and can in some cases be quite similar. For example, if the z components of the surface normals of an object do not vary much, then some of the harmonic images will be quite similar, such as λ vs. λz. This may cause some components to be less significant, so that a lower-dimensional approximation can be fairly accurate.

When s sampled images are used (typically $s >> r$), with $s << p$ PCA requires $O(ps^2)$. Also, in MATLAB, PCA of a thin, rectangular matrix seems to take exactly twice as long as its QR decomposition. Therefore, in practice, PCA on the matrix constructed by Georghides et al. would take about 150 times as long as using our method to build a 9D linear approximation to a model's images (this is for $s = 100$ and $r = 9$. One might expect p to be about 10,000, but this does not effect the relative costs of the methods). This may not be too significant if pose is known ahead of time and this computation takes place off line. But when pose is computed at run time, the advantages of our method can become very great.

It is also interesting to compare our method to another linear method, due to Shashua [25] and Moses [20]. Shashua points out that in the absence of attached shadows, every possible image of an object is a linear combination of the x, y and z components of the surface normals, scaled by the albedo. He therefore proposes using these three components to produce a 3D linear subspace to represent a model's images. Notice that these three vectors are identical, up to a scale factor, to the basis images produced by the first order harmonics in our method.

While this equivalence is clear algebraicly, we can also explain it as follows. The first order harmonic images are images of any object subjected to a lighting condition described by a single harmonic. The Funk-Hecke theorem ensures that all components of the kernel describing the reflectance function will be irrelevant to this image except for the first order components. In Shashua's work, the basis images are generated by using a point source as the lighting function, which contains all harmonics. But the kernel used is the full cosine function of the angle between the light and the surface normal. This kernel has components only in the first harmonic. So all other components of the lighting are irrelevant to the image. In either case, the basis images are due only to the first set of harmonics.

We can therefore interpret Shashua's method as also making an analytic approximation to a model's images, using low order harmonics. However, our previous analysis tells us that the images of the first harmonic account for only 50% percent of the energy passed by the half-cosine kernel. Furthermore, in the worst case it is possible for the lighting to contain *no* component in the first harmonic. Most notably, Shashua's method does not make use of the DC component of the images, i.e., of the zero'th harmonic. These are the images produced by a perfectly diffuse light source. Non-negative lighting must

always have a significant DC component. Koenderink and van Doorn [18] have suggested augmenting Shashua's method with this diffuse component. This results in a linear method that uses the four most significant harmonic basis images, although Koenderink and van Doorn propose this as apparently an heuristic suggestion, without analysis or reference to a harmonic representation of lighting.

3.2 Enforcing Positive Light

When we take arbitrary linear combinations of the harmonic basis images, we may obtain images that are not physically realizable. This is because the corresponding linear combination of the harmonics representing lighting may contain negative values. That is, rendering these images may require negative "light", which of course is physically impossible. In this section we show how to use the basis images while enforcing the constraint of non-negative light. Belhumeur and Kriegman [1] have shown that the set of images of an object produced by non-negative lighting is a convex cone in the space of all possible images. They call this the *illumination cone*. We show how to compute approximations to this cone in the space spanned by the harmonic basis images.

Specifically, given an image I we attempt to minimize $\|Ba - I\|$ subject to the constraint that the light is non-negative everywhere along the sphere. A straightforward method to enforce positive light is to infer the light from the images by inverting the convolution. This would yield linear constraints in the components of a, $Ha \geq 0$, where the columns of H contain the spherical harmonics h_{nm}. Unfortunately, this naive method is problematic since the light may contain higher order terms that cannot be recovered from a low order approximation of the images of the object. In addition, the harmonic approximation of non-negative light may at times include negative values. Forcing these values to be non-negative will lead to an incorrect recovery of the light. Below we describe a different method in which we project the illumination cone [1] onto the low dimensional space and use this projection to enforce non-negative lighting.

We first present a method that can use any number of harmonic basis images. A non-negative lighting function can be written as a non-negative combination of delta functions, each representing a point source. Denote by $\delta_{\theta_0\phi_0}$ the function returning a non-zero value at (θ_0, ϕ_0), 0 elsewhere, and integrating to 1. This lighting function represents a point source at direction (θ_0, ϕ_0). To project the delta function onto the first few harmonics we need to look at the harmonic transform of the delta function. Since the inner product of $\delta_{\theta_0\phi_0}$ with a function f returns simply $f(\theta_0, \phi_0)$, we can conclude that the harmonic transform of the delta function is given by

$$\delta_{\theta_0\phi_0} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm}(\theta_0, \phi_0) h_{nm}. \tag{25}$$

The projection of the delta function onto the first few harmonics, therefore, is obtained by taking the sum only over the first few terms.

Suppose now that a non-negative lighting function $\ell(\theta, \phi)$ is expressed as a non-negative combination of delta functions

$$\ell = \sum_{j=1}^{J} a_j \delta_{\theta_j\phi_j}, \tag{26}$$

for some J. Obviously, due to the linearity of the harmonic transform, the transform of ℓ is a non-negative combination of the transforms of the delta functions with the same coefficients. That is,

$$\ell = \sum_{j=1}^{J} a_j \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm}(\theta_j, \phi_j) h_{nm}. \tag{27}$$

Likewise, the image of an object illuminated by ℓ can be expressed as a non-negative combination as follows

$$I = \sum_{j=1}^{J} a_j \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm}(\theta_j, \phi_j) b_{nm}, \tag{28}$$

where b_{nm} are the harmonic images defined in the previous section.

Given an image our objective is to recover the non-negative coefficients a_j. Assume we consider an approximation of order N, and denote the number of harmonics required for spanning the space by $r = r(N)$ (e.g., if $N = 2$ then $r = 9$). In matrix notation, denote the harmonic functions by H, H is $s \times r$, where s is the number of sample points on the sphere. The columns of H contain a sampling of the harmonic functions, while its rows contain the transform of the delta functions. Further, denote by B the basis images, B is $p \times r$, where p is the number of points in the image. Every column of B contains one harmonic image b_{nm}. Finally, denote $a^T = (a_1, ..., a_s)$. Then, our objective is to solve the non-negative least squares problem:

$$\min_a \|BH^T a - I\| \quad \text{s.t.} \quad a \geq 0. \tag{29}$$

We can further project the image to the r-dimensional space spanned by the harmonic images and solve the optimization problem in this smaller space. To do so we apply a QR decomposition to B, as described previously. We obtain:

$$\min_a \|RH^T a - Q^T I\| \quad \text{s.t.} \quad a \geq 0. \tag{30}$$

Now R is $r \times r$ and $Q^T I$ is an r-vector.

Note that this method is similar to that presented in Georghides et al. [8]. The primary difference is that we work in a low dimensional space constructed for each model using its harmonic basis images. Georghides et al. perform a similar computation after projecting all images into a 100-dimensional space constructed using PCA on images rendered from models in a ten-model data base. Also we do not need to explicitly render images using a point source, and project them into a low-dimensional space. In our representation the projection of these images is simply H^T.

3.3 Recognition with Four Harmonics

A further simplification can be obtained if the set of images of an object is approximated only up to first order. Four harmonics are required in this case. One is the DC component, representing the appearance of the object under uniform ambient light, and three are the basis images also used by

Shashua. Again, we attempt to minimize $\|Ba - I\|$ (now B is $p \times 4$) subject to the constraint that the light is non-negative everywhere along the sphere.

As before, we determine the constraints by projecting the delta functions onto the space spanned by the first four harmonics. However, now this projection takes a particularly simple form. Consider a delta function $\delta_{\theta_0\phi_0}$. Its first order approximation is given by

$$\delta_{\theta_0\phi_0} \approx \sum_{n=0}^{1} \sum_{m=-n}^{n} h_{nm}(\theta_0, \phi_0) h_{nm}. \tag{31}$$

Using space coordinates this approximation becomes

$$\delta_{\theta_0\phi_0}(x, y, z) \approx \frac{1}{4\pi} + \frac{3}{4\pi}(x \sin\theta_0 \cos\phi_0 + y \sin\theta_0 \sin\phi_0 + z \cos\theta_0). \tag{32}$$

Let

$$\ell \approx a_0 + a_1 x + a_2 y + a_3 z \tag{33}$$

be the first order approximation of a non-negative lighting function ℓ. ℓ is a non-negative combination of delta functions. It can be readily verified that such a combination cannot decrease the zero order coefficient relative to the first order ones. Consequently, any non-negative combination of delta functions must satisfy

$$9a_0^2 \geq a_1^2 + a_2^2 + a_3^2. \tag{34}$$

(Equality is obtained when the light is a delta function, see (32).) Therefore, we can express the problem of recognizing an object with a 4D harmonic space as minimizing $\|Ba - I\|$ subject to (34).

In the four harmonic case the harmonic images are just the albedos and the components of the surface normals scaled by the albedos, each scaled by some factor. It is therefore natural to use those directly and hide the scaling coefficients within the constraints. Let I be an image of the object illuminated by ℓ, then, using (19) and (23),

$$I \approx \pi a_0 \lambda + \frac{2\pi}{3}(a_1 \lambda n_x + a_2 \lambda n_y + a_3 \lambda n_z). \tag{35}$$

where λ and (n_x, n_y, n_z) are respectively the albedo and the surface normal of an object point. Using the unscaled basis images, λ, λn_x, λn_y, and λn_z, this equation can be written as:

$$I \approx b_0 \lambda + b_1 \lambda n_x + b_2 \lambda n_y + b_3 \lambda n_z, \tag{36}$$

with $b_0 = \pi a_0$ and $b_i = \frac{2\pi}{3} a_i$ $(1 \leq i \leq 3)$. Substituting for the a_i's we obtain

$$\frac{9b_0^2}{\pi^2} \geq \frac{9}{4\pi^2}(b_1^2 + b_2^2 + b_3^2), \tag{37}$$

which simplifies to

$$4b_0^2 \geq b_1^2 + b_2^2 + b_3^2. \tag{38}$$

Figure 4: Test images used in the experiments.

Consequently, to find the nearest image in the space spanned by the first four harmonic images with non-negative light we may minimize the difference between the two sides of (36) subject to (38). This problem has the general form:

$$\min_x \|Ax - b\| \quad \text{s.t.} \quad x^T B x = 0. \tag{39}$$

We show in Appendix B that by diagonalizing A and B simultaneously and introducing a Lagrange multiplier the problem can be solved by finding the roots of a six degree polynomial with a single variable, the Lagrange multiplier. With this manipulation solving the minimization problem is straightforward.

3.4 Experiments

We have experimented with these recognition methods using a database of faces collected at NEC, Japan. The database contains 3D models of 42 faces, including models of their albedos in the red, green and blue color channels. As query images we use 42 images each of ten individuals, taken across seven different poses and six different lighting conditions (shown in Figure 4). In our experiment, each of the query images is compared to each model.

In all methods, we first obtain a 3D alignment between the model and the image, using the algorithm

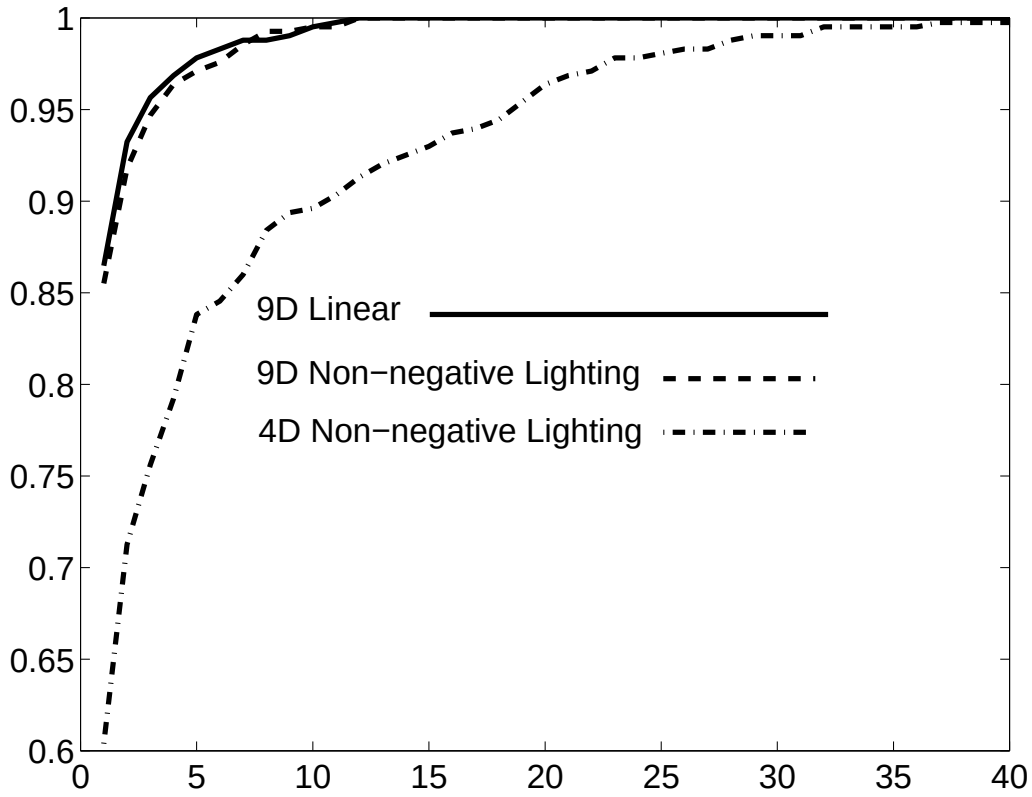

Figure 5: ROC curves for our recognition methods.

of Blicher and Roy [2]. In brief, features on the faces were identified by hand, and then a 3D rigid transformation was found to align the 3D features with the corresponding 2D image features.

In all methods, we only pay attention to image pixels that have been matched to some point in the 3D model of the face. We also ignore image pixels that are of maximum intensity, since these may be saturated, and provide misleading values. Finally, we subsample both the model and the image, replacing each $m \times m$ square with its average values. Preliminary experiments indicate that we can subsample quite a bit without significantly reducing accuracy. In the experiments below, we ran all algorithms subsampling with 16×16 squares, while the original images were 640×480.

Our methods produce coefficients that tell us how to linearly combine the harmonic images to produce the rendered image. These coefficients were computed on the sampled image, but then applied to harmonic images of the full, unsampled image. This process was repeated separately for each color channel. Then, a model was compared to the image by taking the root mean squared error, derived from the distance between the rendered face model and all corresponding pixels in the image.

Figure 5 shows ROC curves for three recognition methods: the 9D linear method, and the methods that enforce positive lighting in 9D and 4D. The curves show the percentage of query images for which the correct model is classified among the top k, as k varies from 1 to 40. The 4D positive lighting

method performs significantly less well than the others, getting the correct answer about 60% of the time. However, it is much faster, and seems to be quite effective under the simpler pose and lighting conditions. The 9D linear method and 9D positive lighting method each pick the correct model first 86% of the time. With this data set, the difference between these two algorithms is quite small compared to other sources of error. These may include limitations in our model for handling cast shadows and specularities, but also includes errors in the model building and pose determination processes. In fact, on examining our results we found that one pose (for one person) was grossly wrong because a human operator selected feature points in the wrong order. We eliminated the six images (under six lighting conditions) that used this pose from our results.

4 Specularity

In general, it is a subject of future work to consider how this sort of analysis may be applied to more complex imaging situations that include specularities and cast shadows. However, in this section we will make one basic remark about these situations.

We note that a low-dimensional set of images can also result when the lighting itself is low-dimensional. This can occur when the lights are all diffuse, as when the sun is behind clouds or lighting is due to inter-reflections. In this case, the lighting itself may be well approximated by only low order harmonics. If the lighting is a linear combination of a small number of harmonics, then images will be a linear combination of those produced when the scene is rendered separately by each of these harmonics. This low-dimensionality is due simply to the linearity of lighting, the fact that the sum of two images produced by any two lighting conditions will be the image produced by the sum of these lighting conditions. Therefore, this will be true under the most general imaging assumptions, including cast shadows and specularities.

We also note that with specular objects, the bidirectional reflection distribution function (BRDF) is generally much more sharply peaked than it is with the cosine function. This provides the intuition that specular objects will be more affected by high-order harmonic components of the lighting. In the extreme case of a mirror, the entire lighting function passes into the reflectance function, preserving all components of the lighting. Therefore, we expect that for specular objects, a low order approximation to the image set will be less accurate. A representation in terms of harmonic images may still provide a useful approximation, however. This is consistent with the experiments of Epstein et al. [7].

5 Conclusions

Lighting can be arbitrarily complex. But in many cases its effect is not. When objects are Lambertian, we show that a simple, nine-dimensional linear subspace can capture the set of images they produce. This explains prior empirical results. It also gives us a new and effective way of understanding the effects of Lambertian reflectance as that of a low-pass filter on lighting.

Moreover, we show that this 9D space can be directly computed from a model, as low-degree polynomial functions of its scaled surface normals. This description allows us to produce efficient recognition algorithms in which we know we are using an accurate approximation to the model's images. We can compare models to images in a 9D space that captures at least 98% of the energy of all the model's images. We can enforce the constraint that lighting be positive by performing a non-negative least squares optimization in this 9D space. Or, if we are willing to settle for a less accurate approximation, we can compute the positive lighting that best matches a model to an image by just solving a six-degree polynomial in one variable. We evaluate the effectiveness of all these algorithms using a data base of models and images of real faces.

Appendix

A The Harmonic Transform of the Lambertian Kernel

The Lambertian kernel is given by $k(\theta) = \max(\cos\theta, 0)$, where θ denotes the solid angle between the light direction and the surface normal. The harmonic transform of k is defined as

$$k = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} k_{nm} h_{nm},$$

where the coefficients k_{nm} are given by

$$k_{nm} = \int_0^{2\pi} \int_0^{\pi} k(\theta) h_{nm}(\theta, \phi) \sin\theta d\theta d\phi.$$

Without loss of generality, we set the coordinate system on the sphere as follows. We position one of the poles at the center of k, θ then represents the angle along a longitude and varies from 0 to π, and ϕ represents an angle along a latitude and varies from 0 to 2π. In this coordinate system k is independent of ϕ and is rotationally symmetric about the pole. Consequently, all its energy is split between the zonal harmonics (the harmonics with $m = 0$), and the coefficients for every $m \neq 0$ vanish. Below we denote $k_n = k_{n0}$.

We next determine an explicit form for the coefficients k_n. First, we can limit the integration to the positive portion of the cosine function by integrating over θ only to $\pi/2$, that is,

$$k_n = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \cos\theta h_{n0}(\theta) \sin\theta d\theta d\phi = 2\pi \int_0^{\frac{\pi}{2}} \cos\theta h_{n0}(\theta) \sin\theta d\theta.$$

Now,

$$h_{n0} = \sqrt{\frac{2n+1}{4\pi}} P_n(\cos\theta),$$

where $P_n(z)$ is the associated Legendre function of order n defined by

$$P_n(z) = \frac{1}{2^n n!} \frac{d^n}{dz^n} (z^2 - 1)^n.$$

Substituting $z = \cos\theta$ we obtain

$$k_n = \sqrt{(2n+1)\pi}\int_0^1 zP_n(z)dz.$$

We now turn to computing the integral

$$\int_0^1 zP_n(z)dz.$$

This integral is equal to

$$\frac{1}{2^n n!}\int_0^1 z\frac{d^n}{dz^n}(z^2-1)^n dz.$$

Integrating by parts yields

$$\frac{1}{2^n n!}\left[z\frac{d^{n-1}}{dz^{n-1}}(z^2-1)^n\Big|_0^1 - \int_0^1 \frac{d^{n-1}}{dz^{n-1}}(z^2-1)^n dz\right].$$

The first term vanishes and we are left with

$$-\frac{1}{2^n n!}\int_0^1 \frac{d^{n-1}}{dz^{n-1}}(z^2-1)^n dz = -\frac{1}{2^n n!}\frac{d^{n-2}}{dz^{n-2}}(z^2-1)^n\Big|_0^1 .$$

This formula vanishes for $z=1$ and so we obtain

$$\frac{1}{2^n n!}\frac{d^{n-2}}{dz^{n-2}}(z^2-1)^n\Big|_{z=0} .$$

Now,

$$(z^2-1)^n = \sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}z^{2k}.$$

When we take the $n-2$ derivative all terms whose exponent is less than $n-2$ disappear. Moreover, since we are evaluating the derivative at $z=0$ all the terms whose exponent is larger than $n-2$ vanish. Thus, only the term whose exponent is $2k = n-2$ survives. Denote the $n-2$ coefficient by b_{n-2}, then, when n is odd $b_{n-2}=0$, and when n is even

$$b_{n-2} = \binom{n}{\frac{n}{2}-1}(-1)^{\frac{n}{2}+1}.$$

In this case

$$\frac{d^{n-2}}{dz^{n-2}}(z^2-1)^n\Big|_{z=0} = (n-2)!b_{n-2} = (n-2)!\binom{n}{\frac{n}{2}-1}(-1)^{\frac{n}{2}+1},$$

and we obtain

$$\int_0^1 zP_n(z)dz = \frac{(-1)^{\frac{n}{2}+1}(n-2)!}{2^n n!}\binom{n}{\frac{n}{2}-1} = \frac{(-1)^{\frac{n}{2}+1}(n-2)!}{2^n(\frac{n}{2}-1)!(\frac{n}{2}+1)!}.$$

The above derivation holds for $n \geq 2$. The special cases that $n = 0$ and $n = 1$ should be handled separately. In the first case $P_0(z) = 1$ and in the second case $P_1(z) = z$. For $n = 0$ the integral becomes

$$\int_0^1 z dz = \frac{1}{2},$$

and for $n = 1$ it becomes

$$\int_0^1 z^2 dz = \frac{1}{3}.$$

Consequently,

$$k_n = \sqrt{(2n+1)\pi} \int_0^1 z P_n(z) dz = \begin{cases} \frac{\sqrt{\pi}}{2} & n = 0 \\ \sqrt{\frac{\pi}{3}} & n = 1 \\ (-1)^{\frac{n}{2}+1} \frac{(n-2)!\sqrt{(2n+1)\pi}}{2^n (\frac{n}{2}-1)!(\frac{n}{2}+1)!} & n \geq 2, \text{ even} \\ 0 & n \geq 2, \text{ odd} \end{cases}$$

B Recognition with Four Harmonics

Finding the nearest image in the 4D harmonic space subject to the constraint that the light is non-negative has the general form

$$\min_x \|Ax - b\| \quad \text{s.t.} \quad x^T B x = 0,$$

with A $(n \times 4)$, b $(n \times 1)$, and B (4×4). In this representation the columns of A contain the unscaled harmonic images, b is the image to be recognized, and $B = \text{diag}(4, -1, -1, -1)$. (The method we present below, however, can be used with an arbitrary nonsingular matrix B.)

First, we can solve the linear system

$$\min_x \|Ax - b\|$$

and check if this solution satisfies the constraint. If it does, we are done. If not, we must seek a minimum that occurs when the constraint is satisfied at equality. We will divide the solution into two parts. In the first part we will convert the problem to the form:

$$\min_z \|z - c\| \quad \text{s.t.} \quad z^T D z \geq 0,$$

Later, we will show how to turn the new problem into a sixth degree polynomial.

Step 1:

First, we can assume WLOG that b resides in the column space of A, since the component of b orthogonal to this space does not affect the solution to the problem. Furthermore, since b lies in the column space of A we can assume that A is 4×4 full rank and b is 4×1. This can be achieved, for

example, using a QR decomposition. Now, define b' such that $Ab' = b$ (this is possible because A is full rank). Then, $Ax - b = A(x - b')$, implying that our problem is equivalent to:

$$\min_x \|A(x - b')\| \quad \text{s.t.} \quad x^T B x = 0.$$

Using the method presented in Golub and van Loan [11] (see the second edition, pages 466–471, especially algorithm 8.7.1) we simultaneously diagonalize $A^T A$ and B. This will produce a non-singular matrix X such that $X^T A^T A X = I$ and $X^T B X = D$, I denotes the identity matrix, and D is a 4×4 diagonal matrix. Thus, we obtain

$$\min_x \|X^{-1}(x - b')\| \quad \text{s.t.} \quad x^T X^{-T} D X^{-1} x = 0.$$

where X^{-1} denotes the inverse of X, and X^{-T} denotes its transpose. Denote $z = X^{-1}x$ and $c = X^{-1}b'$, then we obtain

$$\min_z \|z - c\| \quad \text{s.t.} \quad z^T D z = 0.$$

This has the desired form.

Step 2:

At this point we attempt to solve a problem of the form

$$\min_z \|z - c\| \quad \text{s.t.} \quad z^T D z = 0.$$

We solve this minimization problem using Lagrange multipliers. That is,

$$\min_z \|z - c\| + \lambda z^T D z.$$

Taking the derivatives with respect to x and λ we get

$$z - c + \lambda D z = 0,$$

and

$$z^T D z = 0.$$

From the first equation we get

$$z = (I + \lambda D)^{-1} c.$$

Since D is diagonal the components of z are given by

$$z_i = \frac{c_i}{1 + \lambda d_i},$$

where $z = (z_1, ..., z_4)$, $c = (c_1, ..., c_4)$, and $D = \text{diag}(d_1, ..., d_4)$. The constraint $z^T D z = 0$ thus becomes

$$\sum_{i=1}^{4} \frac{c_i^2 d_i}{(1 + \lambda d_i)^2} = 0,$$

which, after multiplying out the denominator, becomes a sixth degree polynomial in λ. This polynomial can be efficiently and accurately solved using standard techniques (we use the MATLAB function *roots*). We plug in all solutions to determine x, as indicated above, and choose the real solution that minimizes our optimization criteria.

Acknowledgements

We are grateful to Bill Bialek, Peter Blicher, Mike Langer and Warren Smith. Their insightful comments and suggestions have been of great assistance to us. We are also grateful to Rui Ishiyama, Shizuo Sakamoto, and Johji Tajima for their helpful comments and for providing us with data for our experiments.

References

[1] P. Belhumeur, D. Kriegman. "What is the Set of Images of an Object Under All Possible Lighting Conditions?", *IEEE Conf. on Computer Vision and Pattern Recognition*: 270–277, 1996.

[2] A.P. Blicher and S. Roy. "Fast Lighting/Rendering Solution for Matching a 2D Image to a Database of 3D Models." (forthcoming).

[3] B. Cabral, N. Max, R. Springmeyer. "Bidirectional Reflection Functions from Surface Bump Maps", *Computer Graphics*, **21**(4): 273–281, 1987.

[4] T.F. Cootes, C,J. Taylor, D.H. Cooper, J. Graham, "Training Models of Shape from Sets of Examples," *Proceedings of the British Machine Vision Conference*, Springer-Verlag, 1992: 9–18.

[5] Y. Dobashi, K. Kaneda, H. Nakatani, H. Yamashita, "A quick rendering method using basis functions for interactive lighting design," *Eurographics*: 229–240, 1995.

[6] M. D'Zmura, 1991. "Shading Ambiguity: Reflectance and Illumination," in *Computational Models of Visual Processing*, edited by M. Landy, and J. Movshon.

[7] R. Epstein, P. Hallinan, A. Yuille. "5 $\pm$ 2 Eigenimages Suffice: An Empirical Investigation of Low-Dimensional Lighting Models," *IEEE Workshop on Physics-Based Vision*: 108–116, 1995.

[8] A. Georghiades, D. Kriegman, P. Belhumeur. "Illumination Cones for Recognition Under Variable Lighting: Faces", *IEEE Conf. on Computer Vision and Pattern Recognition*: 52–59, 1998.

[9] A. Georghiades, P. Belhumeur ,D. Kriegman. "From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination", *Int. Conf. on Automatic Face and Gesture Recognition*, 2000.

[10] H. Groemer, *Geometric applications of Fourier series and spherical harmonics*, Cambridge University Press.

[11] G. Golub, C. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1989.

[12] P. Hallinan. "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions", *IEEE Conf. on Computer Vision and Pattern Recognition*: 995–999, 1994.

[13] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *Journal of the Optical Society of America*, **11**(11): 3079–3089, 1994.

[14] B.K.P. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1986.

[15] D. Jacobs, P. Belhumeur, and R. Basri. "Comparing Images Under Variable Illumination," NECI TR:97-183, 1997.

[16] M. Kirby, and L. Sirovich, "The application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE transactions on Pattern Analysis and Machine Intelligence*, **12**(1): 103-108, 1990.

[17] J. Koenderink and A. Van Doorn, "Bidirectional reflection distribution function expressed in terms of surface scattering modes," *European Conference on Computer Vision*, **2**: 28–39, 1996.

[18] J. Koenderink and A. Van Doorn, 1997. "The Generic Bilinear Calibration-Estimation Problem," *International Journal of Computer Vision*, **23**(3):217–234.

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

[20] Y. Moses. *Face recognition: generalization to novel images*, Ph.D. Thesis, Weizmann Institute of Science, 1993.

[21] H. Murase, S. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, **14**(1): 5–25, 1995.

[22] S. Nayar and H. Murase, "Dimensionality of illumination manifolds in appearance matching," *NSF/ARPA Workshop on 3D Object Representation for Computer Vision*, 1996: 165.

[23] J. Nimeroff, E. Simoncelli, J. Dorsey, "Efficient re-rendering of naturally illuminated environments," *5th Eurographics Workshop on Rendering*, 1994.

[24] R. Ramamoorthi, P. Hanrahan. Personnal communication.

[25] A. Shashua. "On Photometric Issues in 3D Visual Recognition from a Single 2D Image", *International Journal of Computer Vision*, **21**(1-2): 99–122, 1997.

[26] P.C. Teo, E.P. Simoncelli, D.J. Heeger, "Efficient linear re-rendering for interactive lighting design," *Tech. Rep. STAN-CS-TN-97-60*, Stanford Univ., 1997.

[27] M. Turk, A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, **3**(1): 71–96, 1991.

[28] S. Ullman, R. Basri, "Recognition by Linear Combinations of Models," *IEEE transactions on Pattern Analysis and Machine Intelligence*, **13**(10): 992–1007, 1991.

[29] S.H. Westin, J.R. Arvo, K.E. Torrance, "Predicting reflectance functions from complex surfaces," *Computer Graphics*, **26**(2): 255–264, 1992.

[30] A. Yuille, D. Snow, R. Epstein, P. Belhumeur, "Determining Generative Models of Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability", *International Journal of Computer Vision*, **35**(3): 203–222, 1999.

[31] L. Zhao and Y.H. Yang, "heoretical analysis of illumination in PCA-based vision systems," *Pattern Recognition* **32**, 1999: 547–564.