
Contents

Chapter 5. Illumination Modeling for Face Recognition	
<i>Ronen Basri, David Jacobs</i>	1
Index	27

Chapter 5. Illumination Modeling for Face Recognition

Ronen Basri¹ and David Jacobs²

¹ The Weizmann Institute of Science, Rehovot 76100, Israel
`ronen.basri@weizmann.ac.il`

² University of Maryland, College Park, MD 20742, `djacobs@umiacs.umd.edu`

1 Introduction

Changes in lighting can produce large variability in the appearance of faces, as illustrated in Figure 1. Characterizing this variability is fundamental to understanding how to account for the effects of lighting in face recognition. In this chapter³, we will discuss solutions to the problem: given a 3D description of a face, its pose, and its reflectance properties, and a 2D query image, how can we efficiently determine whether lighting conditions exist that can cause this model to produce the query image? We describe methods that solve this problem by producing simple, linear representations of the set of all images that a face can produce under all lighting condition. These results can be directly used in face recognition systems that capture 3D models of all individuals to be recognized. They also have the potential to be used in recognition systems that compare strictly 2D images, but that do so using generic knowledge of 3D face shape.

One way to measure the difficulties presented by lighting, or any variability, is the number of degrees of freedom needed to describe it. For example, the pose of a face relative to the camera has six degrees of freedom, three rotations and three translations. Facial expression has a few tens of degrees of freedom if one considers the number of muscles that may contract to change expression. To describe the

³ Portions reprinted, with permission, from [5], © 2004 IEEE.



Fig. 1. The same face, under two different lighting conditions.

light that strikes a face, we must describe the intensity of light hitting each point on the face from each direction. That is, light is a function of position and direction, meaning that light has an infinite number of degrees of freedom. In this chapter, however, we will show that effective systems can account for the effects of lighting using fewer than ten degrees of freedom. This can have considerable impact on the speed and accuracy of recognition systems.

Support for low-dimensional models is both empirical and theoretical. Principal component analysis on images of a face taken under different lighting conditions shows that this image set is well approximated by a low-dimensional, linear subspace of the space of all images (eg., [18]). And experimentation shows that algorithms that take advantage of this observation can achieve high performance (eg., [17, 21]). In addition, we will describe theoretical results that, with some simplifying assumptions, prove the validity of low-dimensional, linear approximations to the set of images produced by a face. For these results we assume that light sources are distant from the face, but we do allow arbitrary combinations of point sources (such as the sun) and diffuse sources (such as the sky). We also consider only diffuse components of reflectance, modelled as Lambertian reflectance, and we ignore the effects of cast shadows, such as those produced by the nose. We do, however, model the effects of attached shadows, as when one side of a head faces away from a light. Theoretical predictions from these models provide a good fit to empirical observations, and produce useful recognition systems. This suggests that the approximations made capture the most significant effects of lighting on facial appearance. Theoretical models are valuable because they provide insight into the role of lighting in face recognition, but also because they lead to analytically derived, low-dimensional, linear representations of the effects of lighting on facial appearance, which in turn can lead to more efficient algorithms.

An alternate stream of work attempts to compensate for lighting effects without the use of 3D face models. This work performs matching directly between 2D images, using representations of images that are found to be insensitive to lighting variations. These include image gradients (Brunelli and Poggio[11]), Gabor jets (Lades et al.[26]), the direction of image gradients (Jacobs et al.[23], Chen et al.[12]), and projections to subspaces derived from linear discriminants (Belhumeur et al.[7]). These methods are certainly of interest, especially for applications in which 3D face models are not available. However, methods based on 3D models may be more powerful, since they have the potential to completely compensate for lighting changes, while 2D methods cannot achieve such invariance (Chen et al.[12], Moses and Ullman[32]; see also Adini et al.[1]). Another approach of interest, which is discussed in Chapter 15 in this volume, is to use general 3D knowledge of faces to improve methods of image comparison.

2 Background on Reflectance and Lighting

Throughout this chapter, we will consider only distant light sources. By a *distant* light source we mean that it is valid to make the approximation that a light shines

on each point in the scene from the same angle, and with the same intensity (this also rules out, for example, slide projectors).

We consider two types of lighting conditions. A *point* source is described by a single direction, represented by the unit vector u_l , and intensity, l . These can be combined into a vector with three components, $\bar{l} = lu_l$. Or, lighting may come from multiple sources, including diffuse sources such as the sky. In that case we can describe the intensity of the light as a function of its direction, $\ell(u_l)$, which does not depend on the position in the scene. Light, then, can be thought of as a non-negative function on the surface of a sphere. This allows us to represent scenes in which light comes from multiple sources, such as a room with a few lamps, and also to represent light that comes from extended sources, such as light from the sky, or light reflected off a wall.

Some of the analysis in this chapter will account for *attached shadows*, which occur when a point in the scene faces away from a light source. That is, if a scene point has a surface normal v_r , and light comes from the direction u_l , when $u_l \cdot v_r < 0$, none of the light strikes the surface. We will also discuss methods of handling *cast shadows*, which occur when one part of a face blocks the light from reaching another part of the face. Cast shadows have been treated by methods based on rendering a model to simulate shadows [16], while attached shadows can be accounted for with analytically derived linear subspaces.

Building truly accurate models of the way the face reflects light is a complex task. This is in part because skin is not homogeneous; light striking the face may be reflected by oils or water on the skin, by melanin in the epidermis, or by hemoglobin in the dermis, below the epidermis (see, for example, Angelopoulou et al.[3], Angelopoulou[2], Meglinski and Matcher[30], which discuss these effects and build models of skin reflectance. See also the Chapter 8 in this volume). Based on empirical measurements of skin, Marschner et al.[29] state: “The BRDF itself is quite unusual; at small incidence angles it is almost Lambertian, but at higher angles strong forward scattering emerges.” Furthermore, light entering the skin at one point may scatter below the surface of the skin, and exit from another point. This phenomena, known as subsurface scattering, cannot be modelled by a Bidirectional Reflectance Function (BRDF), which assumes that light leaves a surface from the point that it strikes it. Jensen et al.[24] present one model of subsurface scattering.

For purposes of realistic computer graphics, this complexity must be confronted in some way. For example, Borshukov and Lewis[10] report that in *The Matrix Reloaded*, they began by modelling face reflectance using a Lambertian diffuse component, and a modified Phong model to account for a Fresnel-like effect. “As production progressed it became increasingly clear that realistic skin rendering couldn’t be achieved without subsurface scattering simulations.”

However, simpler models may be adequate for face recognition. They also lead to much simpler and more efficient algorithms. This suggests that even if one wishes to more accurately model face reflectance, simple models may provide useful, approximate algorithms that can initialize more complex ones. In this chapter we will discuss analytically derived representation of the images produced by a con-

vex, Lambertian object illuminated by distant light sources. We restrict ourselves to convex objects, so we can ignore the effect of shadows cast by one part of the object on another part of it. We assume that the surface of the object reflects light according to Lambert’s law [27], which states that materials absorb light and reflect it uniformly in all directions. The only parameter of this model is the *albedo* at each point on the object, which describes the fraction of the light reflected at that point.

Specifically, according to Lambert’s law, if a light ray of intensity l and coming from the direction u_l reaches a surface point with albedo ρ and normal direction v_r , then the intensity, i , reflected by the point due to this light is given by

$$i = l(u_l)\rho \max(u_l \cdot v_r, 0). \quad (1)$$

If we fix the lighting, and ignore ρ for now, then the reflected light is a function of the surface normal alone. We write this function as $r(\theta_r, \phi_r)$, or $r(v_r)$. If light reaches a point from a multitude of directions then the light reflected by the point would be the integral over the contribution for each direction. If we denote $k(u \cdot v) = \max(u \cdot v, 0)$, then, we can write:

$$r(v_r) = \int_{S^2} k(u_l \cdot v_r) \ell(u_l) du_l. \quad (2)$$

where \int_{S^2} denotes integration over the surface of the sphere.

3 Using PCA to Generate Linear Lighting Models

We can consider a face image as a point in a high-dimensional space by treating each pixel as a dimension. Then one can use principal component analysis to determine how well one can approximate a set of face images using a low-dimensional, linear subspace. PCA was first applied to images of faces by Sirovitch and Kirby[40], and used for face recognition by Turk and Pentland[41]. Hallinan [18] used PCA to study the set of images that a single face in a fixed pose produces when illuminated by a floodlight placed in different positions. He found that a five or six dimensional subspace accurately models this set of images. Epstein et al. [13] and Yuille et al. [43] 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. This work suggests that lighting variation has a low-dimensional effect on face images, although it does not make clear the exact reasons for this.

Because of this low-dimensionality, linear representations based on PCA can be used to compensate for lighting variation. In Georghiades et al.[16] a 3D model of a face is used to render images with attached or also with cast shadows. PCA is used to compress these images to a low-dimensional subspace, in which they are compared to new images (also using non-negative lighting constraints which we

discuss in Section 5). One issue raised by this approach is that the linear subspace produced depends on the face’s pose. Computing this on-line, when pose is determined, is potentially very expensive. Georghiades et al.[17] attack this problem by sampling pose space and generating a linear subspace for each pose. Ishiyama and Sakamoto[21] instead generate a linear subspace in a model-based coordinate system, so that this subspace can be transformed in 3D, as pose varies.

4 Linear lighting models: without shadows

This empirical work was to some degree motivated by a previous result that showed that Lambertian objects, in the absence of *all* shadows, produce a set of images that form a three-dimensional linear subspace (Shashua[37], Moses[31]). To see this, consider a Lambertian object illuminated by a point source described by the vector, \bar{l} . Let p_i denote a point on the object, let n_i be a unit vector describing the surface normal at p_i , and let ρ_i denote the albedo at p_i , and define $\bar{n}_i = \rho_i n_i$. Then in the absence of attached shadows, Lambertian reflectance is described by $\bar{l}^T \bar{n}_i$. If we combine all of an object’s surface normals into a single matrix N , so that the i ’th column of N is \bar{n}_i , then the entire image is described by $I = \bar{l}^T N$. This implies that any image is a linear combination of the three rows of N . These are three vectors consisting of the x , y , and z components of the object’s surface normals, scaled by albedo. Consequently, all images of an object lie in a three-dimensional space spanned by these three vectors. Note that if we have multiple light sources, $\bar{l}_1 \dots \bar{l}_d$, we have:

$$I = \sum_i (\bar{l}_i N) = \left(\sum_i \bar{l}_i \right) N$$

so that this image, too, lies in this three-dimensional subspace. Belhumeur et al.[7] report face recognition experiments using this 3D linear subspace. They find that this approach partially compensates for lighting variation, but not as well as methods that account for shadows.

Hayakawa[19] uses factorization to build 3D models using this linear representation. Koenderink and van Doorn[25] augmented this space in order to account for an additional, perfect diffuse component. When in addition to a point source there is also an ambient light, $\ell(u_l)$, which is constant as a function of direction, and we ignore cast shadows, this has the effect of adding the albedo at each point, scaled by a constant, to the image. This leads to a set of images that occupy a four-dimensional linear subspace.

5 Attached shadows: non-linear models

Belhumeur and Kriegman[8] began the analytic study of the images that an object produces when shadows are present. First, they point out that for arbitrary illumination, scene geometry, and reflectance properties, the set of images produced by

an object forms a convex cone in image space. It is a cone, because the intensity of lighting can be scaled by any positive value, creating an image scaled by the same positive value. It is convex because two lighting conditions that create two images can always be added together to produce a new lighting condition that creates an image that is the sum of the original two images. They call this set of images the *Illumination Cone*.

Then they show that for a convex, Lambertian object, in which there are attached shadows but no cast shadows, the dimensionality of the illumination cone is $O(n^2)$ where n is the number of distinct surface normals visible on the object. For an object such as a sphere, in which every pixel is produced by a different surface normal, the illumination cone has volume in image space. This proves that the images of even a simple object do not lie in a low-dimensional linear subspace. They do note, however, that simulations indicate that the illumination cone is “thin”, that is, it lies near a low-dimensional image space, which is consistent with the experiments described in Section 3. They further show how to construct the cone using the representation of Shashua[37]. Given three images taken with lighting that produces no attached or cast shadows, they construct a 3D linear representation, clip all negative intensities at zero, and take convex combinations of the resulting images.

Georghiades, Belhumeur and Kriegman[16, 17] present several algorithms that use the illumination cone for face recognition. The cone can be represented by sampling its extremal rays; this corresponds to rendering the face under a large number of point light sources. An image may be compared to a known face by measuring its distance to the illumination cone, which they show can be computed using non-negative least squares algorithms. This is a convex optimization guaranteed to find a global minimum, but it is slow when applied to a high-dimensional image space. So they suggest running the algorithm after projecting the query image and the extremal rays to a lower-dimensional subspace, using PCA.

Also of interest is the approach of Blicher and Roy[9], which buckets nearby surface normals, and renders a model based on the average intensity of image pixels that have been matched to normals within a bucket. This method assumes that similar normals produce similar intensities (after the intensity is divided by albedo), so it is suitable for handling attached shadows. It is also extremely fast.

6 Linear Lighting Models: Spherical Harmonic Representations

The empirical evidence showing that for many common objects the illumination cone is “thin” even in the presence of attached shadows has remained unexplained until recently, when Basri and Jacobs[4, 5], and in parallel Ramamoorthi and Hanrahan[35], analyzed the illumination cone in terms of spherical harmonics. This analysis shows that, when we account for attached shadows, the images of a convex Lambertian object can be approximated to high accuracy using nine (or even fewer) basis images. In addition, this analysis provides explicit expressions for the basis

images. These expressions can be used to construct efficient recognition algorithms that handle faces under arbitrary lighting. At the same time these expressions can be used to construct new shape reconstruction algorithms that work under unknown combinations of point and extended light sources. Below we review this analysis. Our discussion is based primarily on Basri and Jacobs[5].

6.1 Spherical Harmonics and the Funk-Hecke Theorem

The key to producing linear lighting models that account for attached shadows lies in noticing that Eq. (2), which describes how lighting is transformed to reflectance, is analogous to a convolution on the surface of a sphere. For every surface normal v_r , reflectance is determined by integrating the light coming from all directions weighted by the kernel $k(u_l \cdot v_r) = \max(u_l \cdot v_r, 0)$. For every v_r this kernel is just a rotated version of the same function, which contains the positive portion of a cosine function. We shall denote the (unrotated) function $k(u_l)$ (defined by fixing v_r at the north pole) and refer to this as the *half-cosine* function. Note that on the sphere convolution is well defined only when the kernel is rotationally symmetric about the north pole, which indeed is the case for this 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. We now introduce these tools, and use them to show that in producing reflectance, k acts as a low-pass filter.

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 Y_{nm} , with $n = 0, 1, 2, \dots$ and $-n \leq m \leq n$:

$$Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_{n|m|}(\cos\theta) e^{im\phi}, \quad (3)$$

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

$$P_{nm}(z) = \frac{(1-z^2)^{m/2}}{2^n n!} \frac{d^{n+m}}{dz^{n+m}} (z^2-1)^n. \quad (4)$$

We say that Y_{nm} is an n 'th *order* harmonic.

Below it will sometimes be convenient to parameterize Y_{nm} as a function of space coordinates (x, y, z) rather than angles. The spherical harmonics, written $Y_{nm}(x, y, z)$, then become polynomials of degree n in (x, y, z) . The first nine harmonics then become

$$\begin{aligned} Y_{00} &= \frac{1}{\sqrt{4\pi}} & Y_{10} &= \sqrt{\frac{3}{4\pi}} z \\ Y_{11}^e &= \sqrt{\frac{3}{4\pi}} x & Y_{11}^o &= \sqrt{\frac{3}{4\pi}} y \\ Y_{20} &= \frac{1}{2} \sqrt{\frac{5}{4\pi}} (3z^2 - 1) & Y_{21}^e &= 3 \sqrt{\frac{5}{12\pi}} xz \\ Y_{21}^o &= 3 \sqrt{\frac{5}{12\pi}} yz & Y_{22}^e &= \frac{3}{2} \sqrt{\frac{5}{12\pi}} (x^2 - y^2) \\ Y_{22}^o &= 3 \sqrt{\frac{5}{12\pi}} xy, \end{aligned} \quad (5)$$

where the superscripts e and o denote the even and the odd components of the harmonics respectively (so $Y_{nm} = Y_{n|m|}^e \pm iY_{n|m|}^o$, 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).

Because the spherical harmonics form an orthonormal basis, this means that any piecewise continuous function, f , on the surface of the sphere can be written as a linear combination of an infinite series of harmonics. Specifically, for any f ,

$$f(u) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} Y_{nm}(u), \quad (6)$$

where f_{nm} is a scalar value, computed as:

$$f_{nm} = \int_{S^2} f(u) Y_{nm}^*(u) du, \quad (7)$$

and $Y_{nm}^*(u)$ denotes the complex conjugate of $Y_{nm}(u)$.

If we rotate a function f , this acts as a phase shift. Define for every n the n 'th order amplitude of f as

$$A_n \stackrel{\text{def}}{=} \sqrt{\frac{1}{2n+1} \sum_{m=-n}^n f_{nm}^2}. \quad (8)$$

Then, rotating f does not change the amplitude of a particular order. It may shuffle values of the coefficients, f_{nm} , for a particular order, but it does not shift energy between harmonics of different orders.

Both the lighting function, ℓ , and the Lambertian kernel, k , can be written as sums of spherical harmonics. Denote by

$$\ell = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} Y_{nm}, \quad (9)$$

the harmonic expansion of ℓ , and by

$$k(u) = \sum_{n=0}^{\infty} k_n Y_{n0}. \quad (10)$$

Note that, because $k(u)$ is circularly symmetric about the north pole, only the zonal harmonics participate in this expansion, and

$$\int_{S^2} k(u) Y_{nm}^*(u) du = 0, \quad m \neq 0. \quad (11)$$

Spherical harmonics are useful in understanding the effect of convolution by k because of the Funk-Hecke theorem, which is analogous to the convolution theorem. Loosely speaking, the theorem states that we can expand ℓ and k in terms of

spherical harmonics, and then convolving them is equivalent to multiplication of the coefficients of this expansion (see [5] for details).

Following the Funk-Hecke theorem, the harmonic expansion of the reflectance function, r , can be written as:

$$r = k * \ell = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left(\sqrt{\frac{4\pi}{2n+1}} k_n l_{nm} \right) Y_{nm}. \quad (12)$$

6.2 Properties of the Convolution Kernel

The Funk-Hecke theorem implies that in producing the reflectance function, r , the amplitude of the light, ℓ , at every order n is scaled by a factor that depends only on the convolution kernel, k . We can use this to infer analytically what frequencies will dominate r . To achieve this we treat ℓ as a signal and k as a filter, and ask how the amplitudes of ℓ change as it passes through the filter.

The harmonic expansion of the Lambertian kernel (10) can be derived (see [5]) yielding

$$k_n = \begin{cases} \frac{\sqrt{\pi}}{2} & n = 0 \\ \sqrt{\frac{\pi}{3}} & n = 1 \\ (-1)^{\frac{n}{2}+1} \frac{\sqrt{(2n+1)\pi}}{2^n(n-1)(n+2)} \binom{n}{\frac{n}{2}} & n \geq 2, \text{ even} \\ 0 & n \geq 2, \text{ odd} \end{cases} \quad (13)$$

The first few coefficients, for example, are

$$\begin{aligned} k_0 &= \frac{\sqrt{\pi}}{2} \approx 0.8862 & k_1 &= \sqrt{\frac{\pi}{3}} \approx 1.0233 \\ k_2 &= \frac{\sqrt{5\pi}}{8} \approx 0.4954 & k_4 &= -\frac{\sqrt{\pi}}{16} \approx -0.1108 \\ k_6 &= \frac{\sqrt{13\pi}}{128} \approx 0.0499 & k_8 &= \frac{\sqrt{17\pi}}{256} \approx -0.0285. \end{aligned} \quad (14)$$

($k_3 = k_5 = k_7 = 0$), $|k_n|$ approaches zero as $O(n^{-2})$. A graph representation of the coefficients is shown in Figure 2.

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

$$\int_0^{2\pi} \int_0^{\pi} k^2(\theta) \sin \theta d\theta d\phi = 2\pi \int_0^{\frac{\pi}{2}} \cos^2 \theta \sin \theta d\theta = \frac{2\pi}{3}. \quad (15)$$

(Here we simplify our computation by integrating over θ and ϕ rather than u . The $\sin \theta$ factor is needed to account for the varying length of the latitude over the sphere.) Figure 2 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 $(\frac{\pi}{4} + \frac{\pi}{3} + \frac{5\pi}{64}) / \frac{2\pi}{3} \approx 99.22\%$ of the energy. With this approximation the half cosine function can be written as:

$$k(\theta) \approx \frac{3}{32} + \frac{1}{2} \cos \theta + \frac{15}{32} \cos^2 \theta. \quad (16)$$

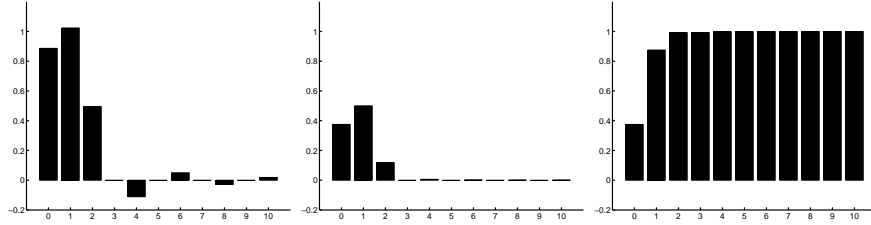


Fig. 2. From left to right: a graph representation of the first 11 coefficients of the Lambertian kernel, the relative energy captured by each of the coefficients, and the cumulative energy.

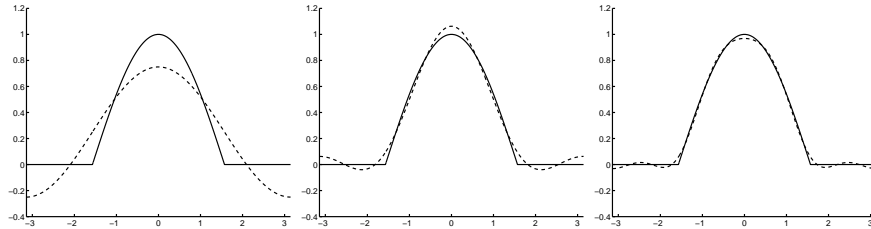


Fig. 3. A slice of the Lambertian kernel (solid) and its approximations (dashed) of first (left), second (middle), and fourth order (right).

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 3 shows a 1D slice of the Lambertian kernel and its various approximations.

6.3 Approximating the Reflectance Function

Because the Lambertian kernel, k , acts as a low-pass filter, the high frequency components of the lighting have little effect on the reflectance function. This implies that we can approximate the reflectance function that occurs under any lighting conditions using only low-order spherical harmonics. In this section, we show that this leads to an approximation that is always quite accurate.

We achieve a low-dimensional approximation to the reflectance function by truncating the sum in Equation (12). That is, we have:

$$r = k * \ell \approx \sum_{n=0}^N \sum_{m=-n}^n \left(\sqrt{\frac{4\pi}{2n+1}} k_n l_{nm} \right) Y_{nm} \quad (17)$$

for some choice of order N . This means considering only the effects of the low order components of the lighting on the reflectance function. Intuitively, we know that since k_n is small for large n , this approximation should be good. However, the accuracy of the approximation also depends on l_{nm} , the harmonic expansion of the lighting.

To evaluate the quality of the approximation consider first, as an example, lighting, $\ell = \delta$, generated by a unit directional (distant point) source at the z direction ($\theta = \phi = 0$). In this case the lighting is simply a delta function whose peak is at the north pole ($\theta = \phi = 0$). It can be readily shown that

$$r(v) = k * \delta = k(v). \quad (18)$$

If the sphere is illuminated by a single directional 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 consists of 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 is given by

$$\frac{k_0^2}{\frac{2\pi}{3} - \sum_{n=1}^N k_n^2}. \quad (19)$$

(A proof appears in [5].)

It can be shown 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 (19) 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.

6.4 Generating Harmonic Reflectances

Constructing a basis to the space that approximates the reflectance functions is straightforward - we can simply use the low order harmonics as a basis (see Equation (17)). However, in many cases we will want a basis vector for the nm component of the reflectances to indicate the reflectance produced by a corresponding basis vector describing the lighting, Y_{nm} . This makes it easy for us to relate reflectances

and lighting, which is important when we want to enforce the constraint that the reflectances arise from non-negative lighting (see Section 7.1 below). We call these reflectances *harmonic reflectances* and denote them by r_{nm} . Using the Funk-Hecke theorem, r_{nm} is given by

$$r_{nm} = k * Y_{nm} = \left(\sqrt{\frac{4\pi}{2n+1}} k_n \right) Y_{nm}. \quad (20)$$

Then, following (17),

$$r = k * \ell \approx \sum_{n=0}^N \sum_{m=-n}^n l_{nm} r_{nm}. \quad (21)$$

The first few harmonic reflectances are given by

$$\begin{aligned} r_{00} &= \pi Y_{00} & r_{1m} &= \frac{2\pi}{3} Y_{1m} & r_{2m} &= \frac{\pi}{4} Y_{2m} \\ r_{4m} &= \frac{\pi}{24} Y_{4m} & r_{6m} &= \frac{\pi}{64} Y_{6m} & r_{8m} &= \frac{\pi}{128} Y_{8m} \end{aligned} \quad (22)$$

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

6.5 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.

We can write this explicitly, as follows. Let p_i denote the i 'th object point. Let n_i denote the surface normal at p_i , and let ρ_i denote the albedo of p_i . Let the illumination be expanded with the coefficients l_{nm} (Equation (9)). Then the image, I_i of p_i is:

$$I_i = \rho_i r(n_i), \quad (23)$$

where

$$r(n_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} r_{nm}(n_i). \quad (24)$$

Then any image is a linear combination of *harmonic images*, b_{nm} , of the form:

$$b_{nm}(p_i) = \rho_i r_{nm}(n_i) \quad (25)$$

with

$$I_i = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} b_{nm}(p_i). \quad (26)$$

Figure 4 shows the first nine harmonic images derived from a 3D model of a face.



Fig. 4. 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.

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. The accuracy of our low dimensional linear approximation can vary, according to the shape and albedos of the object. Every shape is characterized by a different distribution of surface normals, and this distribution may significantly differ from the distribution of normals on the sphere. Viewing direction also affects this distribution, since all normals facing away from the viewer are not visible in the image. Albedo further affects the accuracy of our low dimensional approximation since it may scale every pixel by a different amount. In the worst case, this can make our approximation arbitrarily bad. For many objects it is possible to illuminate the object by lighting configurations that will produce images for which low order harmonic representations provide a poor approximation.

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.

The following argument shows that the lower bound on the accuracy of a harmonic approximation to the reflectance function also provides a lower bound on the average accuracy of the harmonic approximation for *any* convex object. (This result is derived in [14].) We assume that lighting is equally likely from all directions. Given an object, we can construct a matrix M whose columns contain the images obtained by illuminating the object by a single point source, for all possible source directions. (Of course there are infinitely many such directions, but we can sample them to any desired accuracy.) The average accuracy of a low rank representation of the images of the object then is determined by

$$\min_{M^*} \frac{\|M^* - M\|^2}{\|M\|^2}, \quad (27)$$

where M^* is low rank. Now consider the rows of M . Every row represents the reflectance of a single surface point under all point sources. Such reflectances are identical to the reflectances of a sphere with uniform albedo under a single point source. (To see this simply let the surface normal and the lighting directions change roles.) We know that under a point source the reflectance function can be approximated by a combination of the first nine harmonics to 99.22%. Since by this argument every row of M can be approximated to the same accuracy, there exists a rank nine matrix M^* that approximates M to 99.22%. This argument can be applied to convex objects of any shape. Thus, on average, nine harmonic images approximate the images of an object by at least 99.22%, and likewise four harmonic images approximate the images of an object by at least 87.5%. Note that this approximation can even be improved somewhat by selecting optimal coefficients to better fit the images of the object. Indeed, simulations indicate that optimal selection of the coefficients often increases the accuracy of the second order approximation up to 99.5% and that of the first order approximation to about 95%.

Ramamoorthi[34] further derived expressions to calculate the accuracies obtained with spherical harmonics for orders less than nine. His analysis in fact demonstrates that generically the spherical harmonics of the same order are not equally significant. The reason is 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 $b_{00} = \rho$ vs. $b_{10} = \rho z$. Ramamoorthi's calculations show a good fit (with a slight overshoot) to the empirical results. With his derivations the accuracy obtained for a 3D representation of a human face is 92% (as opposed to 90.2% in empirical studies) and for 7D 99% (as opposed to 95.3%). The somewhat lower accuracies obtained in empirical studies may be attributed to the presence of specularities, cast shadows, and noisy measurements.

Finally, it is interesting to compare the basis images determined by our spherical harmonic representation with the basis images derived for the case of no shadows. As we have mentioned earlier in Section 4, Shashua[37] and Moses[31] point 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. They therefore propose using these three components to produce a 3D linear subspace to represent a model's images. Interestingly, these three vectors are identical, up to a scale factor, to the basis images produced by the first order harmonics in our method.

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. We noted in Section 4 that Koenderink and van Doorn[25] have suggested augment-

ing 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.

7 Applications

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 in various tasks, including object recognition and shape reconstruction. We begin by describing methods for recognizing faces under different illumination and pose. Later we briefly describe reconstruction algorithms for stationary (“photometric stereo”) and moving objects.

7.1 Recognition

In a typical recognition problem, the 3D shape and reflectance properties (including surface normals and albedos) of faces may be available. The task then is, given an image of a face seen under unknown pose and illumination, to recognize the individual. Our spherical harmonic representation enables us to perform this task while accounting for complicated, unknown lightings that include combinations of point and extended sources. Below we 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 query image to each model in turn. To compare to a model we compute the distance between the query 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 possible 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.

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, as derived by our harmonic representation. To do this, we produce the harmonic basis images of each model, as described in Section 6.5. Given an image I we seek the distance from I to the space spanned by the basis images. Let B denote the basis images. Then we seek a vector a that minimizes $\|Ba - I\|$. 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. Then Q is an orthonormal basis for B , and $Q^T Q I$ is the projection of I into the space spanned by B . 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 $p \gg r$.

The use of an analytically derived basis can have a substantial effect on the speed of the recognition process. In a previous work Georghiadis et al.[17] performed recognition by rendering the images of an object under many possible lightings and finding an 11D subspace that approximates these images. With our method this expensive rendering step is unnecessary. When s sampled images are used (typically $s \gg r$), with $s \ll p$ 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 Georghiadis 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 affect 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.

Enforcing Non-Negative 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.

When we use a 9D approximation to an object's images, we can efficiently enforce the non-negative lighting constraint in a manner similar to that proposed by Belhumeur and Kriegman[8], after projecting everything into the appropriate 9D linear subspace. Specifically, we approximate any arbitrary lighting function as a non-negative combination of a fixed set of directional light sources. We solve for the best such approximation by fitting to the query image a non-negative combination of images each produced by a single, directional source.

We can do this efficiently using the 9D subspace that represents an object's images. We project into this subspace a large number of images of the object, in which each image is produced by a single directional light source. Such a light source is represented as a delta function; we can derive the representation of the resulting image in the harmonic basis simply by taking the harmonic transform of the delta function that represents the lighting. Then, we can also project a query image into this 9D subspace, and find the non-negative linear combination of directionally lit images that best approximate the query image. Finding the non-negative combination of vectors that best fit a new vector is a standard, convex optimization problem. We can solve it efficiently because we have projected all the images into a space that is only nine-dimensional.

Note that this method is similar to that presented in Georghiades et al.[16]. The primary difference is that we work in a low dimensional space constructed for each model using its harmonic basis images. Georghiades 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 given in closed form by the spherical 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. In this case, we can reduce the resulting optimization problem to one of finding the roots of a sixth degree polynomial, which is extremely efficient. Further details of both methods can be found in [5].

Specularity

Recent work has built on this spherical harmonic representation to also account for non-Lambertian reflectance (Osadchy et al.[33]). The method first computes Lambertian reflectance. This constrains the possible location of a dominant compact source of light. Then, it extracts highlight candidates as pixels that are brighter than we can predict from Lambertian reflectance. Next, we determine which of these candidates are consistent with a known 3D object. A general model of specular reflectance is used that implies that if one thresholds specularities based on intensity, the surface normals that produce specular points will form a disk on the



Fig. 5. Test images used in the experiments.

Gaussian sphere. Therefore, the method proceeds by selecting candidate specularities consistent with such a disk. It maps each candidate specularity to the point on the sphere having the same surface normal. Next, a plane is found that separates the specular pixels from the other pixels with a minimal number of misclassifications. The presence of specular reflections that are consistent with the object's known 3D structure then serves as a cue that the model and image match.

This method has succeeded in recognizing very shiny objects, such as pottery. However, informal face recognition experiments with this method, using the data set described in the next Section, have not shown significant improvements. Our sense is that most of our recognition errors are due to misalignments in pose, and that when a good alignment is found between a 3D model and image, a Lambertian model is sufficient to produce good performance on a data set of 42 individuals.

In other recent work, Georghiadis[15] has augmented the recognition approach of Georghiadis et al.[17] to include specular reflectance. After initialization using a Lambertian model, the position of a single light source and parameters of the Torrance-Sparrow model of specular reflectance are optimized to fit a 3D model of an individual. Face recognition experiments with a data set of ten individuals show that this produces a reduction in overall errors from 2.96% to 2.47%. It seems probable that experiments with data sets containing larger numbers of individuals will be needed to truly gauge the value of methods that account for specular reflectance.

Experiments

We have experimented with these recognition methods using a database of faces collected at NEC, Japan. The database contains models of 42 faces, each includes the 3D shape of the face (acquired using a structured light system) and estimates of the 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 5). In our experiment, each of the query images is compared to each of the 42 models, and then the best matching model is selected.

In all methods, we first obtain a 3D alignment between the model and the image, using the algorithm of Blicher and Roy [9]. In brief, a dozen or fewer 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 6 shows Receiver Operating Characteristic (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 fraction 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.

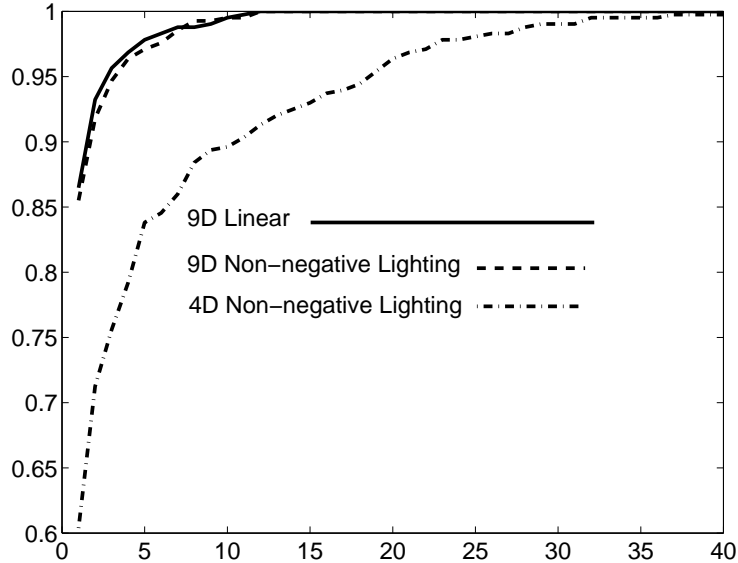


Fig. 6. ROC curves for our recognition methods. The vertical axis shows the percentage of times that the correct model was found among the k best matching models, while the horizontal axis shows k .

7.2 Modeling

The recognition methods described in the previous section require detailed 3D models of faces, as well as their albedos. Such models can be acquired in various ways. For example, in the experiments above we used a laser scanner to recover the 3D shape of a face, and we estimated the albedos from an image taken under ambient lighting (which was approximated by averaging several images of a face). As an alternative it is possible to recover the shape of a face from images illuminated by structured light, or by using stereo reconstruction, although stereo algorithms may give somewhat inaccurate reconstructions for non-textured surfaces. Finally, recent studies developed reconstruction methods that use the harmonic formulation to simultaneously recover both the shape and the albedo of an object. In the remainder of this section we briefly describe two such methods. We first describe how to recover the shape of an object when the input images are obtained with a stationary object illuminated by variable lighting, a problem commonly referred to as “photometric stereo.” Later, we discuss an approach for shape recovery of a moving object.

Photometric Stereo

In photometric stereo, we are given a collection of images of a stationary object, under varying illumination. Our objective is to recover the 3D shape of the object and

its reflectance properties, which for a Lambertian object include the albedo at every surface point. Previous approaches to photometric stereo under unknown lighting generally assume that in every image the object is illuminated by a dominant point source (e.g., [19, 25, 43]). However, by using spherical harmonic representations it is possible to reconstruct the shape and albedo of an object under unknown lighting configurations that include arbitrary collections of point and extended sources. In this section we summarize this work, which is described in more detail in [6].

We begin by stacking the input images into a matrix M of size $f \times p$, in which every input image of p pixels occupies a single row, and f denotes the number of images in our collection. The low dimensional harmonic approximation then implies that there exist two matrices, L and S , of sizes $f \times r$ and $r \times p$ respectively, that satisfy:

$$M \approx LS, \quad (28)$$

where L represents the lighting coefficients, S the harmonic basis, and r is the dimension used in the approximation (usually 4 or 9). If indeed we can recover L and S then obtaining the surface normals and albedos of the shape is straightforward using Eqs. 22 and 25.

We can attempt to recover L and S using SVD. This will produce a factorization of M into two matrices \tilde{L} and \tilde{S} , which are related to the correct lighting and shape matrices by an unknown, arbitrary $r \times r$ ambiguity matrix A . So we can try to reduce this ambiguity. Consider the case that we use a first order harmonic approximation ($r = 4$). Omitting unnecessary scale factors, the zero order harmonic contains the albedo at every point, and the three first order harmonics contain the surface normal scaled by the albedo. For a given point we can write these four components in a vector, i.e., $p = (\rho, \rho n_x, \rho n_y, \rho n_z)^T$. Then p should satisfy $p^T J p = 0$, where $J = \text{diag}\{-1, 1, 1, 1\}$. Enforcing this constraint reduces the ambiguity matrix from 16 degrees of freedom to just seven. Further resolution of the ambiguity matrix requires additional constraints, which can be obtained by specifying a few surface normals, or by enforcing integrability.

A similar technique can be applied in the case of a second order harmonic approximation ($r = 9$). In this case there exist many more constraints on the nine basis vectors, and those can be satisfied by applying an iterative procedure. Using the 9 harmonics the surface normals can be recovered up to a rotation, and further constraints are required to resolve the remaining ambiguity.

An application of these photometric stereo methods is demonstrated in Figure 7. A collection of 32 images of a statue of a face illuminated by two point sources in each image were used to reconstruct the 3D shape of the statue. (The images were simulated by averaging pairs of images obtained with single light sources taken by researchers at Yale.) Saturated pixels were removed from the images and filled in using Wiberg's algorithm [42] (see also [22, 38]). We resolved the remaining ambiguity by matching some points in the scene with hand chosen surface normals.

Photometric stereo is one way to produce a 3D model for face recognition. An alternative approach is to determine a discrete set of lighting directions that will produce a set of images that span the 9D set of harmonic images of an object.

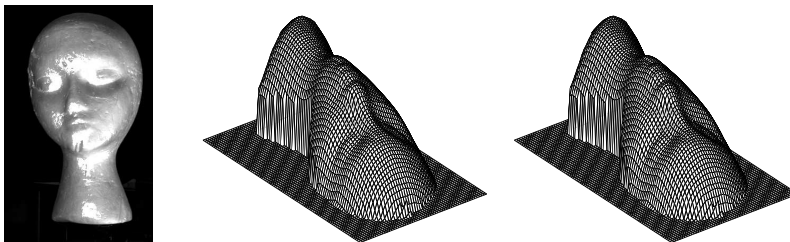


Fig. 7. On the left, two face images averaged together to produce an image with two point sources. Saturated pixels shown in white. In the center, the surface produced by the 4D method. On the right, the surface from the 9D method. Reprinted, with permission, from [6], © 2004 IEEE.

In this way, the harmonic basis can be constructed directly from images, without building a 3D model. This problem is addressed by Lee et al.[28] and by Sato et al.[36]. Other approaches use harmonic representations to cluster the images of a face under varying illumination[20] or determine the harmonic images of a face from just one image using a statistical model derived from a set of 3D models of other faces[45].

Objects in Motion

Photometric stereo methods require a still object while lighting varies. For faces this requires a cooperative subject and controlled lighting. An alternative approach is to use video of a moving face. Such an approach, presented by Simakov et al.[39], is briefly described below.

We assume that the motion of a face is known, for example by tracking a few feature points such as the eyes and the tips of the mouth. Thus we know the epipolar constraints between the images and (in case the cameras are calibrated) also the mapping from 3D to each of the images. To obtain a dense shape reconstruction we need to find correspondences between points in all images. Unlike stereo, in which we can expect corresponding points to maintain approximately the same intensity, in the case of a moving object we expect points to change their intensity as they turn away or toward light sources.

We therefore adopt the following strategy. For every point in 3D we associate a “correspondence measure,” a measure that indicates if its projections in all the images could come from the same surface point. To this end we collect all the projections and compute the residual of the following set of equations:

$$I_j = \rho l^T R_j Y(n). \quad (29)$$

in this equation $1 \leq j \leq f$, f is the number of images, I_j denote the intensity of the projection of the 3D point in the j 'th image, ρ is the unknown albedo, l denotes the unknown lighting coefficients, R_j denotes the rotation of the object in the j 'th

image, and $Y(n)$ denotes the spherical harmonics evaluated for the unknown surface normal. Thus to compute the residual we need to find l and n that minimize the difference between the two sides of this equation. (Note that for a single 3D point ρ and l can be combined to produce a single vector.)

Once we have computed the correspondence measure for every 3D point we can incorporate the measure in any stereo algorithm to extract the surface that minimizes the measure, possibly subject to some smoothness constraints.

The algorithm of Simakov et al.[39] described above assumes that the motion between the images is known. Zhang et al. [44] proposed an iterative algorithm that simultaneously recovers the motion assuming infinitesimal motion between images and modeling reflectance using a first order harmonic approximation.

8 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. In addition, we can use the harmonic formulation to develop reconstructions algorithms to recover the 3D shape and albedos of an object. We evaluate the effectiveness of our recognition algorithms using a database of models and images of real faces.

Acknowledgements

Major portions of this research were conducted while Ronen Basri and David Jacobs were at the NEC Research Institute, Princeton, NJ. At the Weizmann Institute Ronen Basri is supported in part by the European Community grant number IST-2000-26001 and by the Israel Science Foundation grant number 266/02. The vision group at the Weizmann Inst. is supported in part by the Moross Foundation.

References

- [1] Y. Adini, Y. Moses, S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**,(7): 721–732, 1997.
- [2] E. Angelopoulou, "Understanding the color of human skin," *Proc. of the SPIE Conf. on Human Vision and Electronic Imaging VI* SPIE **4299**: 243–251, 2001.

- [3] E. Angelopoulou, R. Molana, and K. Daniilidis, "Multispectral Skin Color Modeling," *IEEE Conf. on Computer Vision and Patt. Rec.*: 635–642., 2001.
- [4] R. Basri, D.W. Jacobs, "Lambertian reflectances and linear subspaces," *IEEE Int. Conf. on Computer Vision*, **II**: 383–390, 2001.
- [5] R. Basri, D.W. Jacobs, "Lambertian reflectances and linear subspaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **25**(2):218–233, (2003).
- [6] R. Basri and D.W. Jacobs, "Photometric stereo with general, unknown lighting," *IEEE Conf. on Computer Vision and Pattern Recognition*, **II**: 374–381, 2001.
- [7] P. Belhumeur, J. Hespanha, and D. Kriegman. "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19**(7): 711–720, 1997.
- [8] P. Belhumeur, D. Kriegman. "What is the set of images of an object under all possible lighting conditions?," *International Journal of Computer Vision*, **28**(3): 245–260, 1998.
- [9] A.P. Blicher, S. Roy. "Fast Lighting/Rendering Solution for Matching a 2D Image to a Database of 3D Models: 'LightSphere'", *IEICE Transactions on Information and Systems*, **E84-D**(12) p.1722-27, 2001.
- [10] G. Borshukov and J.P. Lewis. "Realistic human face rendering for 'The Matrix Reloaded'," *SIGGRAPH-2003 Sketches and Applications Program*, 2003.
- [11] R. Brunelli, T. Poggio, T., "Face recognition: Features versus templates", *IEEE Trans. on pattern analysis and machine intelligence*, **15**(10):1042–1062, 1993.
- [12] H. Chen, P. Belhumeur, D. Jacobs, "In search of illumination invariants", *IEEE Proc. Computer Vision and Pattern Recognition*, **I**:254–261, 2000.
- [13] R. Epstein, P. Hallinan, A. Yuille. " 5 ± 2 eigenimages suffice: an empirical investigation of low-dimensional lighting models," *IEEE Workshop on Physics-Based Vision*: 108–116, 1995.
- [14] D. Frolova, D. Simakov, R. Basri, "Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting," forthcoming.
- [15] A. Georghiades. "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo", *International Conference on Computer Vision*, **II**:816–823, 2003.
- [16] 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.
- [17] A. Georghiades, P. Belhumeur, D. Kriegman. "From few to many: generative models for recognition under variable pose and illumination", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **23**(6):643-660, 2001.
- [18] P. Hallinan. "A low-dimensional representation of human faces for arbitrary lighting conditions", *IEEE Conf. on Computer Vision and Pattern Recognition*: 995–999, 1994.
- [19] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *Journal of the Optical Society of America*, **11**(11): 3079–3089, 1994.

- [20] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. “Clustering appearances of objects under varying illumination conditions”, *IEEE Conf. on Computer Vision and Pattern Recognition*, **1**:11–18, 2003.
- [21] R. Ishiyama and S. Sakamoto. “Geodesic illumination basis: compensating for illumination variations in any pose for face recognition,” *IEEE Int. Conf. on Pattern Recognition*, **4**:297-301, 2002.
- [22] D. Jacobs, “Linear fitting with missing data for structure-from-motion,” *Computer Vision and Image Understanding*, **82**(1):57–81, 2001.
- [23] D. Jacobs, P. Belhumeur, and R. Basri. “Comparing images under variable illumination”, *IEEE Proc. Computer Vision and Pattern Recognition*, 610-617, 1998.
- [24] H.W. Jensen, S.R. Marschner, M. Levoy, and P. Hanrahan. “A practical model for subsurface light transport”. In *Proc. SIGGRAPH*, 511–518, 2001.
- [25] J. Koenderink, A. Van Doorn, “The generic bilinear calibration-estimation problem,” *International Journal of Computer Vision*, **23**(3): 217–234, 1997.
- [26] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture”, *IEEE Trans. on Computers*, **42**(3):300-311, 1993.
- [27] J. Lambert, “Photometria sive de mensura et gradibus luminus, colorum et umbrae,” Eberhard Klett, 1760.
- [28] K.C. Lee, J. Ho, D. Kriegman, “Nine points of light: acquiring subspaces for face recognition under variable lighting,” *IEEE Conf. on Computer Vision and Pattern Recognition*: 519–526, 2001.
- [29] S. Marschner, S. Westin, E. Lafortune, K. Torrance, and D. Greenberg. “Image-based BRDF measurement including human skin,” *10th Eurographics Workshop on Rendering*, pps. 131–144, 1999.
- [30] I.V. Meglinski and S.J. Matcher, “Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions,” *Physiol. Meas.* **23**, 741–753, 2002.
- [31] Y. Moses, *Face recognition: generalization to novel images*, Ph.D. Thesis, Weizmann Institute of Science, 1993.
- [32] Y. Moses and S. Ullman, “Limitations of Non Model-Based Recognition Schemes,” *Second European Conference on Computer Vision*:820-828, 1992.
- [33] M. Osadchy, D. Jacobs, R. Ramamoorthi, “Using specularities for recognition”, *International Conference on Computer Vision*, II:1512–1519, 2003.
- [34] R. Ramamoorthi, “Analytic PCA construction for theoretical analysis of lighting variability in a single image of a Lambertian object,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **24**(10), 2002.
- [35] R. Ramamoorthi, P. Hanrahan, “On the relationship between radiance and irradiance: determining the illumination from images of convex Lambertian object.” *Journal of the Optical Society of America*, **18**(10): 2448–2459, 2001.
- [36] I. Sato, T. Okabe, Y. Sato, and K. Ikeuchi, “ Appearance sampling for obtaining a set of basis images for variable illumination”, *IEEE Int. Conf. on Computer Vision*, **II**:800–807, 2003.

- [37] 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.
- [38] H.Y. Shum, K. Ikeuchi, R. Reddy, “Principal component analysis with missing data and its application to polyhedral object modeling,” *PAMI*, **17**(9):854–867, 1995.
- [39] D. Simakov, D. Frolova, R. Basri, “Dense shape reconstruction of a moving object under arbitrary, unknown lighting,” *IEEE Int. Conf. on Computer Vision*:1202–1209, 2003.
- [40] L. Sirovitch and M. Kirby, “Low-dimensional procedure for the characterization of human faces,” *Journal of the Optical Society of America*, **2**:586–591, 1987.
- [41] M. Turk, A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, **3**(1): 71–96, 1991.
- [42] T. Wiberg, “Computation of principal components when data are missing”, *Proc. Second Symp. Computational Statistics*:229–236, 1976.
- [43] 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.
- [44] L. Zhang, B. Curless, A. Hertzmann, and S.M. Seitz, “Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multi-view stereo,” *IEEE Int. Conf. on Computer Vision*: 618–625, 2003.
- [45] L. Zhang and D. Samaras. “Face recognition under variable lighting using harmonic image exemplars”, *IEEE Conf. on Computer Vision and Pattern Recognition*, **I**:19–25, 2003.

Index

Face recognition with spherical harmonic
representations, 15
Funk-Hecke theorem, 7
Harmonic reflectances, 11
Illumination Cone, 6
Photometric stereo, 20
Principal component analysis, 4
Shape reconstruction, 20
Specular reflectance, 17
Spherical harmonic representations, 6
Subsurface scattering, 3