

Chapter 7

Illumination Modeling for Face Recognition

Ronen Basri and David Jacobs

7.1 Introduction

Changes in lighting can produce large variability in the appearance of faces, as illustrated in Fig. 7.1. Characterizing this variability is fundamental to understanding how to account for the effects of lighting on face recognition. In this chapter, we will discuss solutions to a problem: Given (1) a three-dimensional description of a face, its pose, and its reflectance properties, and (2) 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 a face can produce under all lighting conditions. 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 shapes.

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 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

Portions of this chapter are reprinted, with permission, from Basri and Jacobs [6], © 2003 IEEE.

R. Basri (✉)

The Weizmann Institute of Science, Rehovot 76100, Israel
e-mail: ronen.basri@weizmann.ac.il

D. Jacobs

University of Maryland, College Park, MD 20742, USA
e-mail: djacobs@umiacs.umd.edu

Fig. 7.1 Same face under different lighting conditions



using fewer than 10 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 (PCA) on images of a face obtained under various lighting conditions shows that this image set is well approximated by a low-dimensional, linear subspace of the space of all images (see, e.g., [19]). Experimentation shows that algorithms that take advantage of this observation can achieve high performance, for example, [17, 21].

In addition, we describe theoretical results that, with some simplified 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 (e.g., the Sun) and diffuse sources (e.g., the sky). We also consider only diffuse components of reflectance, modeled 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 not only 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 directly matches 2D images using representations of images that are found to be insensitive to lighting variations. These include image gradients [12], Gabor jets [29], the direction of image gradients [13, 24], and projections to subspaces derived from linear discriminants [8]. A large number of these methods are surveyed in [50]. 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, as they have the potential to compensate completely for lighting changes, whereas 2D methods cannot achieve such invariance [1, 13, 35]. Another approach of interest, the Morphable Model, is to use general 3D knowledge of faces to improve methods of image comparison.

7.2 Background on Reflectance and Lighting

Throughout this chapter, we 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 lighting conditions. A *point* source is described by a single direction, represented by the unit vector u_l , and intensity, l . These factors can be combined into a vector with three components, $\vec{l} = lu_l$. Lighting may also 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 nonnegative 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.

Most of the analysis in this chapter accounts 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 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 [18], whereas 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, [2, 3, 33], which discuss these effects and build models of skin reflectance; see also Chap. 6). Based on empirical measurements of skin, Marschner et al. [32] 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 phenomenon, known as subsurface scattering, cannot be modeled by a bidirectional reflectance function (BRDF), which assumes that light leaves a surface from the point that it strikes it. Jensen et al. [25] presented one model of subsurface scattering.

For purposes of realistic computer graphics, this complexity must be confronted in some way. For example, Borshukov and Lewis [11] reported that in *The Matrix Reloaded*, they began by modeling 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, more efficient algorithms. This suggests that even if one wishes to model face reflectance more accurately, simple models may provide useful, approximate algorithms that can initialize more complex ones. In this chapter, we discuss

analytically derived representation of the images produced by a convex, 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 [30], 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 coming from the direction u_l reaches a surface point with albedo ρ and normal direction v_r , the intensity i reflected by the point due to this light is given by

$$i = l(u_l)\rho \max(u_l \cdot v_r, 0). \quad (7.1)$$

If we fix the lighting and ignore ρ for now, 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, the light reflected by the point would be the integral over the contribution for each direction. If we denote $k(u \cdot v) = \max(u \cdot v, 0)$, we can write:

$$r(v_r) = \int_{S^2} k(u_l \cdot v_r) \ell(u_l) du_l \quad (7.2)$$

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

7.3 PCA Based 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 PCA 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 [44], and used for face recognition by Turk and Pentland [45]. Hallinan [19] used PCA to study the set of images that a single face in a fixed pose produces when illuminated by a floodlight placed in various positions. He found that a five- or six-dimensional subspace accurately models this set of images. Epstein et al. [14] and Yuille et al. [47] described experiments on a wide 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, whereas 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 it.

Because of this low-dimensionality, linear representations based on PCA can be used to compensate for lighting variation. Georgiades et al. [18] used a 3D model of a face to render images with attached or 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 nonnegative lighting constraints we discuss in Sect. 7.5). 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 expensive. Georghiades et al. [17] attacked this problem by sampling pose space and generating a linear subspace for each pose. Ishiyama and Sakamoto [21] instead generated a linear subspace in a model-based coordinate system, so this subspace can be transformed in 3D as the pose varies.

7.4 Linear Lighting Models without Shadows

The empirical study of the space occupied by the images of various real objects 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 [34, 40]. 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 , let ρ_i denote the albedo at p_i , and define $\bar{n}_i = \rho_i n_i$. 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 the i th column of N is \bar{n}_i , 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

$$I = \sum_i (\bar{l}_i N) = \left(\sum_i \bar{l}_i \right) N \quad (7.3)$$

so this image, too, lies in this three-dimensional subspace. Belhumeur et al. [8] reported face recognition experiments using this 3D linear subspace. They found that this approach partially compensates for lighting variation, but not as well as methods that account for shadows.

Hayakawa [20] used factorization to build 3D models using this linear representation. Koenderink and van Doorn [28] augmented this space 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, it 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.

7.5 Nonlinear Models with Attached Shadows

Belhumeur and Kriegman [9] conducted an analytic study of the images an object produces when shadows are present. First, they pointed 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 showed 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 noted, 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 Sect. 7.3. They further showed how to construct the cone using the representation of Shashua [40]. Given three images obtained with lighting that produces no attached or cast shadows, they constructed a 3D linear representation, clipped all negative intensities at zero, and took convex combinations of the resulting images.

Georgiades and colleagues [17, 18] presented 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 showed can be computed using nonnegative 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. Therefore, they suggested 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 [10], 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 the albedo), so it is suitable for handling attached shadows. It is also extremely fast.

7.6 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, 6], and in parallel Ramamoorthi and Hanrahan [38], analyzed the illumination cone in terms of spherical harmonics. This analysis showed 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. We next review this analysis. Our discussion is based primarily on the work of Basri and Jacobs [6].

7.6.1 Spherical Harmonics and the Funk–Hecke Theorem

The key to producing linear lighting models that account for attached shadows lies in noting that (7.2), 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 denote the (unrotated) function $k(u_l)$ (defined by fixing v_r at the north pole) and refer to it as the *half-cosine* function. Note that on the sphere convolution is well defined only when the kernel is rotationally symmetrical 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 when 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 $-n \leq m \leq n$:

$$Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_{n|m|}(\cos\theta) e^{im\phi} \quad (7.4)$$

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

$$P_{nm}(z) = \frac{(1-z^2)^{m/2}}{2^n n!} \frac{d^{n+m}}{dz^{n+m}} (z^2-1)^n. \quad (7.5)$$

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

It is sometimes 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 be-

come polynomials of degree n in (x, y, z) . The first nine harmonics then become

$$\begin{aligned}
 Y_{00} &= \frac{1}{\sqrt{4\pi}}, & Y_{10} &= \sqrt{\frac{3}{4\pi}}z, \\
 Y_{11}^e &= \sqrt{\frac{3}{4\pi}}x, & Y_{11}^o &= \sqrt{\frac{3}{4\pi}}y, \\
 Y_{20} &= \frac{1}{2}\sqrt{\frac{5}{4\pi}}(3z^2 - 1), & Y_{21}^e &= 3\sqrt{\frac{5}{12\pi}}xz, \\
 Y_{21}^o &= 3\sqrt{\frac{5}{12\pi}}yz, & Y_{22}^e &= \frac{3}{2}\sqrt{\frac{5}{12\pi}}(x^2 - y^2), \\
 Y_{22}^o &= 3\sqrt{\frac{5}{12\pi}}xy,
 \end{aligned} \tag{7.6}$$

where the superscripts e and o denote the even and 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 because the reflectance function is real).

Because the spherical harmonics form an orthonormal basis, 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 ,

$$f(u) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} Y_{nm}(u) \tag{7.7}$$

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

$$f_{nm} = \int_{S^2} f(u) Y_{nm}^*(u) du \tag{7.8}$$

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

Rotating a function f results in a phase shift. Define for every n the n 'th order amplitude of f as

$$A_n \stackrel{\text{def}}{=} \sqrt{\frac{1}{2n+1} \sum_{m=-n}^n f_{nm}^2}. \tag{7.9}$$

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

$$\ell = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} Y_{nm} \tag{7.10}$$

the harmonic expansion of ℓ , and by

$$k(u) = \sum_{n=0}^{\infty} k_n Y_{n0}. \quad (7.11)$$

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

$$\int_{S^2} k(u) Y_{nm}^*(u) du = 0, \quad m \neq 0. \quad (7.12)$$

Spherical harmonics are useful for 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 Basri and Jacobs [6] for details).

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

$$r = k * \ell = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left(\sqrt{\frac{4\pi}{2n+1}} k_n l_{nm} \right) Y_{nm}. \quad (7.13)$$

7.6.2 Properties of the Convolution Kernel

The Funk–Hecke theorem implies that when 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 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 (7.11) can be derived [6] yielding

$$k_n = \begin{cases} \frac{\sqrt{\pi}}{2} & n = 0, \\ \sqrt{\frac{\pi}{3}} & n = 1, \\ (-1)^{\frac{n}{2}+1} \frac{\sqrt{(2n+1)\pi}}{2^n (n-1)(n+2)} \binom{n}{\frac{n}{2}} & n \geq 2, \text{ even}, \\ 0 & n \geq 2, \text{ odd}. \end{cases} \quad (7.14)$$

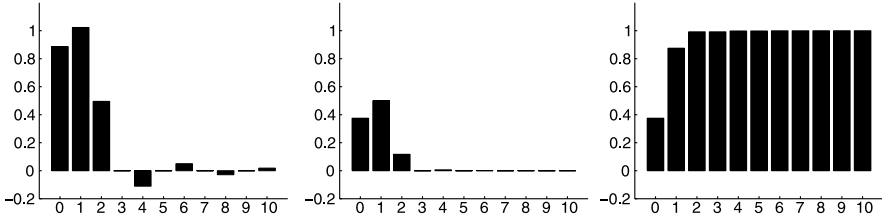


Fig. 7.2 From left to right: the first 11 coefficients of the Lambertian kernel; the relative energy captured by each of the coefficients; and the cumulative energy

The first few coefficients, for example, are

$$\begin{aligned}
 k_0 &= \frac{\sqrt{\pi}}{2} \approx 0.8862, & k_1 &= \sqrt{\frac{\pi}{3}} \approx 1.0233, \\
 k_2 &= \frac{\sqrt{5\pi}}{8} \approx 0.4954, & k_4 &= -\frac{\sqrt{\pi}}{16} \approx -0.1108, \\
 k_6 &= \frac{\sqrt{13\pi}}{128} \approx 0.0499, & k_8 &= \frac{\sqrt{17\pi}}{256} \approx -0.0285
 \end{aligned}
 \tag{7.15}$$

($k_3 = k_5 = k_7 = 0$), $|k_n|$ approaches zero as $O(n^{-2})$. A graphic representation of the coefficients may be seen in Fig. 7.2.

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

$$\int_0^{2\pi} \int_0^\pi k^2(\theta) \sin \theta \, d\theta \, d\phi = 2\pi \int_0^{\frac{\pi}{2}} \cos^2 \theta \sin \theta \, d\theta = \frac{2\pi}{3}.
 \tag{7.16}$$

(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 7.2 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:

$$k(\theta) \approx \frac{3}{32} + \frac{1}{2} \cos \theta + \frac{15}{32} \cos^2 \theta.
 \tag{7.17}$$

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 7.3 shows a one-dimensional slice of the Lambertian kernel and its various approximations.

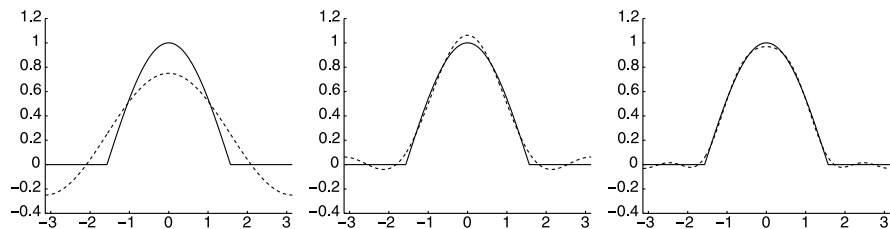


Fig. 7.3 A slice of the Lambertian kernel (*solid line*) and its approximations (*dashed line*) of first (*left*), second (*middle*), and fourth (*right*) order

7.6.3 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 (7.13). That is, we have:

$$r = k * \ell \approx \sum_{n=0}^N \sum_{m=-n}^n \left(\sqrt{\frac{4\pi}{2n+1}} k_n l_{nm} \right) Y_{nm} \quad (7.18)$$

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 because 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

$$r(v) = k * \delta = k(v). \quad (7.19)$$

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, 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 high. This is a consequence of the nonnegativity of light. A lower bound on the accuracy of the approximation for *any* light function is given by

$$\frac{k_0^2}{\frac{2\pi}{3} - \sum_{n=1}^N k_n^2}. \quad (7.20)$$

(Proof appears in Basri and Jacobs [6].)

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 (7.20) is not tight, as 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.

7.6.4 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 (7.18)). However, in many cases we 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 nonnegative lighting (see Sect. 7.7.1). We call these reflectances *harmonic reflectances* and denote them by r_{nm} . Using the Funk–Hecke theorem, r_{nm} is given by

$$r_{nm} = k * Y_{nm} = \left(\sqrt{\frac{4\pi}{2n+1}} k_n \right) Y_{nm}. \quad (7.21)$$

Then, following (7.18),

$$r = k * \ell \approx \sum_{n=0}^N \sum_{m=-n}^n l_{nm} r_{nm}. \quad (7.22)$$

The first few harmonic reflectances are given by

$$\begin{aligned} r_{00} &= \pi Y_{00}, & r_{1m} &= \frac{2\pi}{3} Y_{1m}, & r_{2m} &= \frac{\pi}{4} Y_{2m}, \\ r_{4m} &= \frac{\pi}{24} Y_{4m}, & r_{6m} &= \frac{\pi}{64} Y_{6m}, & r_{8m} &= \frac{\pi}{128} Y_{8m} \end{aligned} \quad (7.23)$$

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

7.6.5 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 represent efficiently 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} (7.10). Then the image, I_i of p_i is

$$I_i = \rho_i r(n_i) \quad (7.24)$$

where

$$r(n_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} r_{nm}(n_i). \quad (7.25)$$

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

$$b_{nm}(p_i) = \rho_i r_{nm}(n_i) \quad (7.26)$$

with

$$I_i = \sum_{n=0}^{\infty} \sum_{m=-n}^n l_{nm} b_{nm}(p_i). \quad (7.27)$$

Figure 7.4 shows the first nine harmonic images derived from a 3D model of a face.

We now 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. Each 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



Fig. 7.4 First nine harmonic images for a model of a face. The *top row* contains the zeroth 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*

this distribution, as all normals facing away from the viewer are not visible in the image. Albedo further affects the accuracy of our low dimensional approximation, as it may scale each pixel by a different amount. In the worst case, this can make our approximation arbitrarily poor. For many objects, it is possible to illuminate the object by lighting configurations that produce images for which low order harmonic representations provide a poor approximation.

However, generally, things are not so bad. In general, occlusion renders an arbitrary half of the normals on the unit sphere invisible. Albedo variations and curvature emphasize some normals and deemphasize others. In general, though, the normals whose reflectances are poorly approximated are not 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 was derived by Frolova et al. [15].) 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

$$\min_{M^*} \frac{\|M^* - M\|^2}{\|M\|^2} \quad (7.28)$$

where M^* is low rank, and $\|\cdot\|$ denotes the Frobenius Norm of a matrix. Now consider the rows of M . Each 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%. Because 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 [37] 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 are not generally orthogonal, and in some cases are quite similar. For example, if the z components of the surface normals of an object do not vary much, some of the harmonic images are quite similar, such as $b_{00} = \rho$ versus $b_{10} = \rho z$. 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% (in contrast to 90.2% in empirical studies) and for 7D 99% (in contrast 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 mentioned in Sect. 7.4, Shashua [40] and Moses [34] pointed 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 proposed 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% 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 zeroth harmonic (commonly referred to as the DC component). These are the images produced by a perfectly diffuse light source. Nonnegative lighting must always have a significant DC component. We noted in Sect. 7.4 that Koenderink and van Doorn [28] 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 proposed it as apparently a heuristic suggestion, without analysis or reference to a harmonic representation of lighting.

7.7 Applications

We have developed an analytic description of the linear subspace that lies near the set of images 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 illuminations and poses. Later, we briefly describe reconstruction algorithms for stationary and moving objects.

7.7.1 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 lighting that includes 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 has identified features, such as the eyes and the tip of the nose, that allow us to determine the pose for each face in the database, but that the database is too large 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 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 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 it 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 depends on the type of recognition problem 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.

7.7.1.1 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, as derived

by our harmonic representation. To do this, we produce the harmonic basis images of each model, as described in Sect. 7.6.5. 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 and $r = 18$ 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. Hence 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 $\|Q Q^T I - I\|$. 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 previous work Georghiades 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 Georghiades 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 10000, but this does not affect the relative costs of the methods.) This may not be significant if pose is known ahead of time and this computation takes place off line. When pose is computed at run time, however, the advantages of our method can become significant.

7.7.1.2 Enforcing Nonnegative 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 nonnegative light.

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

Note that this method is similar to that presented in Georghiades et al. [18]. The primary difference is that we work in a low dimensional space constructed for each model using its harmonic basis images. Georghiades et al. performed a similar computation after projecting all images into a 100-dimensional space constructed using PCA on images rendered from models in a 10-model database. 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 in a single variable, which is extremely efficient. Further details of both methods can be found elsewhere [6].

The approach of enforcing nonnegative lighting for 9 harmonics relies on representing lighting as the nonnegative sum of a large number of delta functions. In this way, the nonnegativity of the lighting follows from the nonnegativity of the coefficients of the delta functions. However, in recent work, Shirdhonkar and Jacobs [41] have shown that nonnegativity can be enforced when representing lighting using low frequency spherical harmonics. To do this, one must be able to determine whether a set of low frequency spherical harmonics are consistent with a nonnegative function; that is, could one add higher frequency harmonics to make the complete function nonnegative. By extending Szego's eigenvalue distribution theorem to spherical harmonics, Shirdhonkar and Jacobs show that a matrix constructed using the coefficients of low frequency lighting, represented as spherical harmonics, must be positive semi-definite in order for these harmonics to be consistent with non-negative lighting. This allows them to compute the low frequency lighting that best matches a 3D model to an image by solving a semi-definite programming problem. This leads to solutions that are more accurate and efficient than previous methods that represent lighting using delta functions.

7.7.1.3 Specularity

Other work has built on this spherical harmonic representation to account for non-Lambertian reflectance [36]. The method first computes Lambertian reflectance, which 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 is consistent with a known 3D object. A general model of specular reflectance is used that implies that the surface normals of specular points obtained by thresholding intensity form a disk on the Gaussian sphere. Therefore, the method proceeds by selecting candidate specularities consistent with such a disk. It maps each candidate specular-ity 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 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 work, Georghiades [16] augmented the recognition approach of Georghiades 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 10 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 large numbers of individuals are needed to truly gauge the value of methods that account for specular reflectance.

7.7.1.4 Experiments

We have experimented with these recognition methods using a database of faces collected at NEC in Japan. The database contains models of 42 faces, each including 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 10 individuals taken across seven poses and six lighting conditions (shown in Fig. 7.5). 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 [10]. 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



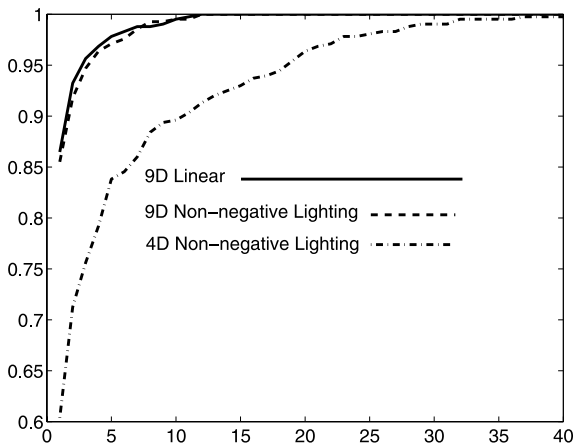
Fig. 7.5 Test images used in the experiments

of maximum intensity, as they 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 combine the harmonic images linearly 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 7.6 shows performance 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 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. Such errors may include limitations in our model for handling cast shadows and specularities, but they also include errors in the model building and pose determination processes. In fact, on examining our results, we found that one pose (for one person)

Fig. 7.6 Performance curves for our recognition methods. The vertical axis shows the percentage of times the correct model was found among the k best matching models; the horizontal axis shows k



was grossly wrong because a human operator selected feature points in the wrong order. We eliminated from our results the six images (under six lighting conditions) that used this pose.

7.7.2 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 described 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 nontextured surfaces. Finally, other studies have developed reconstruction methods that use the harmonic formulation to recover both the shape and the albedo of an object simultaneously. In the remainder of this section, we briefly describe three 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. We conclude with an approach that can recover the shape of faces from single images by exploiting prior knowledge of the generic shape of faces.

7.7.2.1 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 for example, [20, 28, 47]. 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 elsewhere [5, 7].

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

$$M \approx LS \tag{7.29}$$

where L represents the lighting coefficients, S is the harmonic basis, and r is the dimension used in the approximation (usually 4 or 9). If indeed we can recover L and S , obtaining the surface normals and albedos of the shape is straightforward using (7.23) and (7.26).

We can attempt to recover L and S using singular value decomposition (SVD). This produces 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 . 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: $p = (\rho, \rho n_x, \rho n_y, \rho n_z)^T$. 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 7. 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 are many more constraints on the nine basis vectors, and they can be satisfied by applying an iterative procedure. Using the nine 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 Fig. 7.7. 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 [46]; see also [23, 42]. 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 produce a set of images that span the 9D set of harmonic images of an object. In this

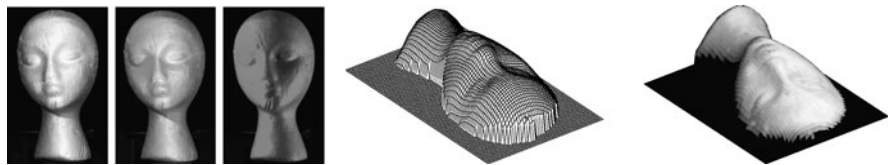


Fig. 7.7 *Left*: three images of a bust illuminated each by two point sources. *Right*: the surface produced by the 4D method (a mesh, and painted with albedo). From Basri, Jacobs, and Kemelmacher [7], © 2007 Springer, with permission

way, the harmonic basis can be constructed directly from images, without building a 3D model. This problem was addressed by Lee et al. [31] and by Sato et al. [39]. Other approaches use harmonic representations to cluster the images of a face under varying illumination [22] 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 [49].

7.7.2.2 Objects in Motion

Photometric stereo methods require a still object while the 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. [43], 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 from or toward light sources.

We therefore adopt the following strategy. For every point in 3D, we associate a “correspondence measure,” which 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.

$$I_j = \rho l^T R_j Y(n). \quad (7.30)$$

In this equation, $1 \leq j \leq f$, f is the number of images, I_j denotes 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 each 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. [43] described above assumes that the motion between the images is known. Zhang et al. [48] proposed an iterative algorithm that simultaneously recovers the motion assuming infinitesimal motion between images and modeling reflectance using a first order harmonic approximation.

7.7.2.3 Reconstruction with Shape Prior

While the previous methods utilize collections of images to achieve 3D reconstruction, it is of interest to explore methods that can recover the shape of faces from just a single image. Recently, Kemelmacher–Shlizerman and Basri [26, 27] proposed such an approach that exploits prior knowledge of the rough shape of faces to make the problem of single view reconstruction well-posed.

The algorithm obtains as input an image of a face to be reconstructed along with a 3D model (shape and albedo) of some different face. Such a model can depict an individual whose 3D shape is available, or an “averaged” model of a collection of faces. The algorithm then attempts to reconstruct the shape of the face in the input image essentially by solving a shape from shading (SFS) problem. However, while SFS is ill-posed and its solution requires knowledge of the lighting conditions, the reflectance properties (albedo) of the object to be reconstructed, and boundary conditions (i.e., depth values at extremal points), this algorithm estimates their values by exploiting the similarity of the input model to the desired shape.

Specifically, Kemelmacher–Shlizerman and Basri seek a solution to the following optimization problem:

$$\min_{l, \rho, z} \int_{\Omega} (I - \rho l^T Y(n))^2 + (\lambda_1 \Delta_z^2 + \lambda_2 \Delta_\rho^2) dx dy. \quad (7.31)$$

In this expression, $I(x, y)$ is the input image ($x, y \in \Omega$), l represents the unknown lighting conditions, $\rho(x, y)$ the unknown albedo, $z(x, y)$ the unknown depth, and $Y(n)$ the spherical harmonic basis derived from z . The first term therefore is a data term fitting the desired reconstruction to the image. For the second term, λ_1 and λ_2 are preset constants and we define $\Delta_z(x, y)$ and $\Delta_\rho(x, y)$ to represent respectively, the (smoothed) difference in shape and albedo between the desired shape and the input model. The role of this regularization term is to keep those differences small. Figure 7.8 shows a reconstruction obtained with this method.

7.8 Conclusions

Lighting can be arbitrarily complex, but in many cases its effect is not. When objects are Lambertian, we show that a simple, 9D linear subspace can capture the set of



Fig. 7.8 Single view reconstruction. The figure shows two triplets of images; each includes an input image, 3D reconstruction (output), and the input image overlaid on the reconstruction. The reference shape used in these runs is shown on the right. Notice that veridical shape is recovered despite change in expression relative to the reference shape. From Kemelmacher–Shlizerman and Basri [27], © 2010 IEEE, with permission

images they produce. This explains prior empirical results. It also gives us a new and effective way to understand 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 of the model’s images. In addition, we can use the harmonic formulation to develop reconstruction 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 European Community grants IST-2000-26001 VIBES and IST-2002-506766 Aim Shape and by the Israel Science Foundation grant 266/02. The vision group at the Weizmann Institute is supported in part by the Moross Foundation. David Jacobs was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government.

References

1. Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 721–732 (1997)
2. Angelopoulou, E.: Understanding the color of human skin. In: *SPIE Conf. on Human Vision and Electronic Imaging VI*, vol. 4299, pp. 243–251. SPIE, Bellingham (2001)
3. Angelopoulou, E., Molana, R., Daniilidis, K.: Multispectral skin color modeling. In: *IEEE Conf. on Computer Vision and Patt. Recognition*, pp. 635–642 (2001)
4. Basri, R., Jacobs, D.: Lambertian reflectances and linear subspaces. In: *IEEE Int. Conf. on Computer Vision*, vol. II, pp. 383–390 (2001)
5. Basri, R., Jacobs, D.: Photometric stereo with general, unknown lighting. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. II, pp. 374–381 (2001)
6. Basri, R., Jacobs, D.: Lambertian reflectances and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(2), 218–233 (2003)
7. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. *Int. J. Comput. Vis.* **72**(3), 239–257 (2007)

8. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
9. Belhumeur, P., Kriegman, D.: What is the set of images of an object under all possible lighting conditions? *Int. J. Comput. Vis.* **28**(3), 245–260 (1998)
10. Blicher, A., Roy, S.: Fast lighting/rendering solution for matching a 2d image to a database of 3d models: ‘lightsphere’. *IEICE Trans. Inf. Syst.* **E84-D**(12), 1722–1727 (2001)
11. Borshukov, G., Lewis, J.: Realistic human face rendering for ‘the matrix reloaded’. In: *SIGGRAPH-2003 Sketches and Applications Program* (2003)
12. Brunelli, R., Poggio, T.: Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(10), 1042–1062 (1993)
13. Chen, H., Belhumeur, P., Jacobs, D.: In search of illumination invariants. In: *IEEE Proc. Computer Vision and Pattern Recognition*, vol. I, pp. 254–261 (2000)
14. Epstein, R., Hallinan, P., Yuille, A.: *pm2* eigenimages suffice: an empirical investigation of low-dimensional lighting models. In: *IEEE Workshop on Physics-Based Vision*, pp. 108–116 (1995)
15. Frolova, D., Simakov, D., Basri, R.: Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. In: *ECCV*, pp. 574–587 (2004)
16. Georgiades, A.: Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In: *International Conference on Computer Vision*, vol. II, pp. 816–823 (2003)
17. Georgiades, A., Belhumeur, P., Kriegman, D.: From few to many: generative models for recognition under variable pose and illumination. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 643–660 (2001)
18. Georgiades, A., Kriegman, D., Belhumeur, P.: Illumination cones for recognition under variable lighting: faces. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 52–59 (1998)
19. Hallinan, P.: A low-dimensional representation of human faces for arbitrary lighting conditions. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 995–999 (1994)
20. Hayakawa, H.: Photometric stereo under a light source with arbitrary motion. *J. Opt. Soc. Am.* **11**(11), 3079–3089 (1994)
21. Ishiyama, R., Sakamoto, S.: Geodesic illumination basis: compensating for illumination variations in any pose for face recognition. In: *IEEE Int. Conf. on Pattern Recognition*, vol. 4, pp. 297–301 (2002)
22. Ho, J., Yang, M., Lim, J., Lee, K., Kriegman, D.: Clustering appearances of objects under varying illumination conditions. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 11–18 (2003)
23. Jacobs, D.: Linear fitting with missing data for structure-from-motion. *Comput. Vis. Image Underst.* **82**(1), 57–81 (2001)
24. Jacobs, D., Belhumeur, P., Basri, R.: Comparing images under variable illumination. In: *IEEE Proc. Computer Vision and Pattern Recognition*, pp. 610–617 (1998)
25. Jensen, H., Marschner, S., Levoy, M., Hanrahan, P.: A practical model for subsurface light transport. In: *Proc. SIGGRAPH*, pp. 511–518 (2001)
26. Kemelmacher, I., Basri, R.: Molding face shapes by example. In: *European Conf. on Computer Vision. LNCS*, vol. 3951, pp. 277–288 (2006)
27. Kemelmacher-Shlizerman, I., Basri, R.: 3d face reconstruction from a single image using a single reference face shape. *IEEE Trans. Pattern Anal. Mach. Intell.* (forthcoming)
28. Koenderink, J., Doorn, A.V.: The generic bilinear calibration-estimation problem. *Int. J. Comput. Vis.* **23**(3), 217–234 (1997)
29. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **42**(3), 300–311 (1993)
30. Lambert, J.: *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhard Klett (1760)

31. Lee, K., Ho, J., Kriegman, D.: Nine points of light: acquiring subspaces for face recognition under variable lighting. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 519–526 (2001)
32. Marschner, S., Westin, S., Lafortune, E., Torrance, K., Greenberg, D.: Image-based brdf measurement including human skin. In: 10th Eurographics Workshop on Rendering, pp. 131–144 (1999)
33. Meglinski, I., Matcher, S.: 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)
34. Moses, Y.: Face recognition: generalization to novel images. Ph.D. thesis, Weizmann Institute of Science (1993)
35. Moses, Y., Ullman, S.: Limitations of non model-based recognition schemes. In: Second European Conference on Computer Vision, pp. 820–828 (1992)
36. Osadchy, M., Jacobs, D., Ramamoorthi, R.: Using specularities for recognition. In: International Conference on Computer Vision, vol. II, pp. 1512–1519 (2003)
37. Ramamoorthi, R.: 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)
38. Ramamoorthi, R., Hanrahan, P.: On the relationship between radiance and irradiance: determining the illumination from images of convex Lambertian object. *J. Opt. Soc. Am.* **18**(10), 2448–2459 (2001)
39. Sato, I., Okabe, T., Sato, Y., Ikeuchi, K.: Appearance sampling for obtaining a set of basis images for variable illumination. In: IEEE Int. Conf. on Computer Vision, vol. II, pp. 800–807 (2003)
40. Shashua, A.: On photometric issues in 3d visual recognition from a single 2d image. *Int. J. Comput. Vis.* **21**(1–2), 99–122 (1997)
41. Shirdhonkar, S., Jacobs, D.: Non-negative lighting and specular object recognition. In: IEEE International Conference on Computer Vision, vol. II, pp. 1323–1330 (2005)
42. Shum, H., Ikeuchi, K., Reddy, R.: Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(9), 854–867 (1995)
43. Simakov, D., Frolova, D., Basri, R.: Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In: IEEE Int. Conf. on Computer Vision, pp. 1202–1209 (2003)
44. Sirovitch, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am.* **2**, 586–591 (1987)
45. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–96 (1991)
46. Wiberg, T.: Computation of principal components when data are missing. In: Proc. Second Symp. Computational Statistics, pp. 229–236 (1976)
47. Yuille, A., Snow, D., Epstein, R., Belhumeur, P.: Determining generative models of objects under varying illumination: shape and albedo from multiple images using svd and integrability. *Int. J. Comput. Vis.* **35**(3), 203–222 (1999)
48. Zhang, L., Curless, B., Hertzmann, A., Seitz, S.: Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multi-view stereo. In: IEEE Int. Conf. on Computer Vision, pp. 618–625 (2003)
49. Zhang, L., Samaras, D.: Face recognition under variable lighting using harmonic image exemplars. In: IEEE Conf. on Computer Vision and Pattern Recognition, vol. I, pp. 19–25 (2003)
50. Zou, X., Kittler, J., Messer, K.: Illumination invariant face recognition: A survey. In: *Biometrics: Theory, Applications, and Systems*, pp. 1–8 (2007)