International Journal of Computer Vision

DOI: 10.1007/s11263-006-8815-7

Photometric Stereo with General, Unknown Lighting

RONEN BASRI
Department of Computer Science and Applied Math, The Weizmann Institute of Science, Rehovot, 76100, Israel

DAVID JACOBS
Department of Computer Science, University of Maryland, College Park, MD 20742

IRA KEMELMACHER
Department of Computer Science and Applied Math, The Weizmann Institute of Science, Rehovot, 76100, Israel

Received June 19, 2005; Revised April 14, 2006; Accepted April 17, 2006

First online version published in June, 2006

Abstract. Work on photometric stereo has shown how to recover the shape and reflectance properties of an object using multiple images taken with a fixed viewpoint and variable lighting conditions. This work has primarily relied on known lighting conditions or the presence of a single point source of light in each image. In this paper we show how to perform photometric stereo assuming that all lights in a scene are distant from the object but otherwise unconstrained. Lighting in each image may be an unknown and may include arbitrary combination of diffuse, point and extended sources. Our work is based on recent results showing that for Lambertian objects, general lighting conditions can be represented using low order spherical harmonics. Using this representation we can recover shape by performing a simple optimization in a low-dimensional space. We also analyze the shape ambiguities that arise in such a representation. We demonstrate our method by reconstructing the shape of objects from images obtained under a variety of lightings. We further compare the reconstructed shapes against shapes obtained with a laser scanner.

Keywords: photometric stereo, Lambertian reflectance, Lorentz transformation, diffuse lighting

1. Introduction

Photometric stereo methods recover the shape and albedo of an object using multiple images in which viewpoint is fixed, and only the lighting conditions vary. Many solutions to this problem exist for laboratory conditions, in which lighting is either known or can be kept simple. In this paper[1] we show how to perform photometric stereo under quite general lighting conditions that need not be known ahead of time. We consider convex objects that are approximately Lambertian, and assume that lights are relatively distant and isotropic (no cast shadows or slide projectors). But otherwise, we allow for arbitrary lighting, including any combination of point sources, extended sources, and diffuse lighting.

Much work on photometric stereo has assumed that lighting comes from a single source, generally a point source or a controlled, diffused source of light (see Section 2). Some recent approaches allow for images containing a single point source and a diffuse component of lighting, provided that the diffuse component is the same for all images (Yuille et al., 1999). These assumptions are natural for many applications,

such as inspection, in which one controls viewing conditions.

In this work, we consider images produced by more general lighting conditions that are not known ahead of time. We do this, first, because we may wish to use shading to construct shapes under everyday lighting. This lighting is often quite complex, consisting of multiple sources of varying kinds, and large surfaces such as walls that reflect light. Second, for some applications, such as modeling large outdoor structures, it may not be practical to completely control the lighting. Third, for other applications, one may wish to reconstruct shape using previously taken photographs, without having access to the object itself. For example, one might wish to use photos of a person taken many years ago to build a model of them. Finally, theories for reconstruction under general, unknown lighting conditions can shed light on how humans perceive shape under similar conditions. This paper develops basic tools to handle complex lighting conditions, and provides a preliminary assessment of them using controlled images. It is a subject of future work to explore these more general applications.

The starting point for our work is a series of results that show that the set of images produced by a convex, Lambertian object under arbitrary lighting can be well approximated by a low-dimensional linear subspace of images. Shashua (1997) and Moses (1993) have shown analytically that, in the absence of all shadows, a Lambertian object produces a 3D subspace of images. Basri and Jacobs (2003) and Ramamoorthi and Hanrahan (2001) show that higher dimensional subspaces can be used to account for the effects of attached shadows. This subspace is 4D when the reflectance function (the amount of light reflected by Lambertian materials with unit albedo as a function of surface normal (Horn, 1986)) is approximated with the zero and first order spherical harmonics. The set of images is 9D when the second order spherical harmonics are included. These results imply that given a number of images under different lighting conditions, principal component analysis can provide a good approximation to an object's complete set of images. In this paper, we consider the problem of translating this linear description of an object's images into a description of its surface normals and albedos. Existing techniques can then be used to translate these normals into an integrable surface, if desired.

To do this, we must fit the low-dimensional space that represents the actual images with a similar space that could be produced by a 3D scene. With a first order approximation, the 4D space of an object's images corresponds to its albedo and its surface normals scaled by albedo (we call these *scaled surface normals*). Therefore, we must approximate the observed images with a 4D space in which one dimension equals the norm of the other three. This can be done by solving an overconstrained linear system, in a manner similar to that used by Tomasi and Kanade (1992) to determine motion for scaled orthographic projection. With a second order approximation, we must find a scene structure whose images under harmonic lighting match the observed images. This can be done with an efficient iterative process, because the scaled surface normals can be described as three linear combinations of nine basis images produced by SVD, requiring us to optimize over only 27 variables. We confirm experimentally that these optimization procedures produce good results.

We also determine the extent to which a linear description of an object's images uniquely determine its surface normals. With the 4D approximation, we show that the normals are determined up to a subgroup of the 4×4 linear transformations, called *Lorentz* transformations. That is, a (scaled) Lorentz transformation of surface normals and albedo generates new surface normals and albedo with the same 4D linear approximation to its images. We also show that after a small linear transformation of an object's scaled normals the 4D approximation to the new object's images lies in the old, 9D approximating subspace. This suggests that small linear transformations of a model have an especially small effect on the images it produces, making it difficult to resolve an unknown linear transformation of the object shape. Additional constraints, such as integrability, can be used to further reduce the ambiguity.

Finally, we describe experiments that illustrate the potential of these methods. We present simulations that show that in spite of the approximations made, in ideal cases we can reconstruct an object's normals up to a few degrees of angle. We also compare reconstructions of real objects to 3D models of the same objects produced with a laser scanner. In our experiments with the 4D approximation our reconstructed shapes achieve an accuracy of 95–98% with respect to the laser scanned shapes, while with the 9D approximation our reconstructed shapes achieve an accuracy of 97–99%.

2. Background

This section briefly reviews past work on photometric stereo (Section 2.1). Since our photometric stereo method relies on recent results on harmonic modeling of Lambertian reflectance we review these results in Section 2.2.

2.1. Previous Approaches

Classical work on photometric stereo has assumed that the illumination conditions are known, e.g., through the specification of a reflectance function (Woodham, 1980) (see (Horn, 1986) for a review). Much work has followed, including some that focuses on dealing with non-Lambertian objects (e.g., Coleman and Jain (1982), Barsky and Petrou (2003), Georghiades (2003a,b); Okatani and Deguchi (2001) and Tagare and Figueiredo (1991)), in some cases capitalizing on distributed light sources (Ikeuchi, 1981; Nayar et al., 1990), reflectance examples (Hertzmann and Seitz, 2005), and situations in which the light source is near the object and near the camera (Kim and Burger (1991) and Clark (1992)). Further work has attempted to improve photometric stereo reconstruction by applying local shape from shading (Sakarya and Erkmen, 2003) and by using known control points (Horovitz and Kiryati, 2004). A continuous form of photometric stereo has been suggested by Zhang et al. (1996). Applications to 3D texture anaylsis were suggested by Smith (1999). Finally, class constraints have been employed by Zhou et al. (2004), and photometric stereo has been combined with motion by Simakov et al. (2003) and Zhang et al. (2003).

Our work focuses on reconstruction from images of Lambertian objects under unknown lighting conditions. We now introduce some notation to allow us to discuss this situation in more detail. We first consider the image intensity, i, produced by a point on a 3D Lambertian object illuminated by a point light source. Let $\vec{l}_3$ denote a 3D column vector containing parameters describing the light. The direction of $\vec{l}_3$ provides the direction from the surface point to the light, while the magnitude of $\vec{l}_3$ encodes the magnitude of the light. Let ρ denote the albedo of the surface point, and $\hat{n} = (n_x, n_y, n_z)^T$ its surface normal ($\|\hat{n}\| = 1$). Then when the surface is facing towards the light source, Lambertian reflection is described by the equation:

$$i = \rho \vec{l}_3^{\,T} \hat{n}. \tag{1}$$

When there are many surface points, we can write this in matrix form as:

$$\vec{i}^{\,T} = \vec{l}_3^{\,T} \mathbf{S_3}. \tag{2}$$

Here $\vec{i}$ is an n-dimensional vector containing all image intensities. If we describe the albedo of all surface points with the vector $\vec{\rho}$, and the components of the surface normals in the vectors $\vec{n}_x, \vec{n}_y, \vec{n}_z$, then $\mathbf{S_3}$ is a $3 \times n$ matrix whose three rows, $\vec{\rho} \circ \vec{n}_x$, $\vec{\rho} \circ \vec{n}_y$, $\vec{\rho} \circ \vec{n}_z$ are each a vector describing the x, y or z component of the object's surface normal at each point, scaled by its albedo. We use the operator '$\circ$' to denote the Hadamard (i.e., *componentwise*) product of two vectors.

Shashua (1997) and Moses (1993) considered the case in which a set of images is produced each with a single point source with *unknown* intensity and direction.Using (2), they showed that three or more images provide enough information to determine the scaled surface normals of an object up to an unknown linear transformation; in fact each image is a linear combination of the x, y and z components of the object's normals scaled by albedo. This result ignores the effect of attached shadow, assuming that all the surface normals face and receive light from all the light sources. Woodham et al. (1991) also consider the problem of unknown light sources under the same assumptions. Hayakawa (1994) uses Shashua (1997) and Moses (1993)'s result in a factorization framework to handle many images. Belhumeur et al. (1999), Yuille et al. (1999), and Fan and Wolff (1997) (see also Onn and Bruckstein (1990)) have shown that integrability reduces the ambiguity to a "generalized bas-relief transformation," which allows the recovery of a surface up to a stretch and shear in the z-direction.

Koenderink and van Doorn (1997) extend this approach by allowing for a single point source plus a perfect ambient component to the lighting. This adds a fourth dimension to a linear description of an object's images, corresponding to albedo. Luong et al. (2002) used the same space to recover the albedo when the surface normals are known. Yuille et al. (1999) describe a reconstruction method for the case in which each image is lit by a single point source, and all images share a common background lighting, which can be arbitrary. All this work shows a progression towards lighting conditions that are less constrained, but still the emphasis is on inferring structure based on the assumption of a single point source in each image.

On the other hand, Jacobs et al. (1998) discusses the case of arbitrary, unknown lighting (assuming lighting is distant and isotropic). They show that, under these conditions, it is impossible to perfectly reconstruct the shape and albedo of an object using any number of images. In fact, they show that even if the shape of the object is known, perfect reconstruction of its albedo is impossible using shading information alone. We skirt this result because our reconstructions are approximate, although quite accurate.

2.2. *Harmonic Modeling of Lambertian Reflectance*

Recently, Basri and Jacobs (2003) and Ramamoorthi and Hanrahan (2001) have provided a new way to describe the effect of general lighting on a Lambertian object. They consider light that may strike an object from all directions. In this case we can express light intensity as a function of direction, $\ell(\theta_l, \phi_l)$ (denoted alternatively as $\ell(\hat{u}_l)$ where $\hat{u}_l$ denotes a point on the unit sphere). When light comes from all directions, we cannot assume, as we did above, that all surfaces in the scene face a common set of light sources. Even if we ignore the effect of cast shadows, all surfaces see light from a hemisphere of directions that varies with the direction of surface normals.

If we fix the lighting, and ignore the albedo 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 $r(\hat{v}_r)$. 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 $k(\hat{u}, \hat{v}) = \max(\hat{u}^T \hat{v}, 0)$, then, we can write:

$$r(\hat{v}_r) = \int_{S^2} k(\hat{u}_l, \hat{v}_r)\ell(\hat{u}_l)du_l. \tag{3}$$

where $\int_{S^2}$ denotes integration over the surface of the sphere. The max operation ensures that when a surface faces away from a light, the amount of reflected light is zero, not a negative value.

Using an analog to the convolution theorem for functions of direction (called the Funk-Hecke theorem), it can be shown that this equation describes a convolution on the surface of the sphere. It states that reflected light is obtained by convolving the incident light with, k, a *half-cosine* kernel, in which negative values are clamped to zero. This kernel can be shown to act as a low-pass filter. This means that only the low frequency components of the lighting function, ℓ, have a significant effect on a Lambertian object's reflectance function. It also means that light of a single low frequency is reflected as a somewhat attenuated function of the same low frequency.

These components are represented as low-order spherical harmonics. Analogous to the Fourier series, spherical harmonics form an orthonormal basis for describing functions on the surface of a sphere. Basri and Jacobs (2003) proves that for *any configuration* of distant lighting, at least 98% of the resulting reflectance function can be captured when light is represented by second order spherical harmonics. A first order approximation captures at least 75% of the reflectance. These bounds are not tight, and in fact many common lighting conditions yield significantly better approximations. For example, under a point source illumination the first and second order harmonics approximate the reflectance function to 87.5% and 99.22% respectively.

This means that to first order, we can approximate any image of an object, under complex lighting conditions, as a linear combination of four images of the object, under four low-frequency lighting conditions. These four *harmonic* images have the form: $(1/\sqrt{4\pi})\vec{\rho}$, $\sqrt{3/4\pi}\vec{\rho} \circ \vec{n}_x$, $\sqrt{3/4\pi}\vec{\rho} \circ \vec{n}_y$, $\sqrt{3/4\pi}\vec{\rho} \circ \vec{n}_z$. Because we are only interested in the linear subspace spanned by these images we can omit constant factors, obtaining $\vec{h}_1 = \vec{\rho}, \vec{h}_2 = \vec{\rho} \circ \vec{n}_x, \vec{h}_3 = \vec{\rho} \circ \vec{n}_y, \vec{h}_4 = \vec{\rho} \circ \vec{n}_z$. Note that the last three images are the images that Shashua (1997) and Moses (1993) show are produced by a point source that faces all surface normals. In this case, we can write: $\vec{i}^T = \vec{l}_4^{\,T}\mathbf{S_4}$, where $\vec{l}_4$ is a 4D vector describing the low frequency components of lighting, and $\mathbf{S_4}$ is a $4 \times n$ matrix whose rows contain these first four harmonic images. The four harmonic images obtained in this case are identical to the representation used in Koenderink and van Doorn (1997), although they interpreted this as relevant to the case of an object illuminated in each image by a perfectly diffuse light plus a single point source that was visible to all surface normals.

If we take a second order approximation to the lighting function, ℓ, an image is described by $\vec{i}^T = \vec{l}_9^T\mathbf{S_9}$, where $\vec{l}_9$ is a 9D vector and $\mathbf{S_9}$ contains the object's images under lighting from nine harmonics that provide a second order approximation. The first four rows of $\mathbf{S_9}$ are the same as $\mathbf{S_4}$. The other five rows are: $\vec{h}_5 = \vec{\rho} \circ (3\vec{n}_z \circ \vec{n}_z - \vec{1}), \vec{h}_6 = \vec{\rho} \circ \vec{n}_x \circ \vec{n}_y, \vec{h}_7 = \vec{\rho} \circ \vec{n}_x \circ \vec{n}_z$, $\vec{h}_8 = \vec{\rho} \circ \vec{n}_y \circ \vec{n}_z, \vec{h}_9 = \vec{\rho} \circ (\vec{n}_x \circ \vec{n}_x - \vec{n}_y \circ \vec{n}_y)$.

In principle, an accurate approximation of the reflectance function under some lighting conditions does

not guarantee an equally accurate approximation of an image of an object illuminated by the same lighting condition, since this accuracy depends on the distribution of the surface normals and albedos of the object. However, Frolova et al. (2004) showed that on average, assuming all lighting directions are equally likely, the accuracy of first and second order spherical harmonic approximations is *at least* 87.5% and 99.22% respectively. Simulations demonstrate that the actual accuracy is even higher, reaching 95% under a first order approximation and 99.5% under a second order approximation.

These results are key to our shape reconstruction method. Below we show how we can use a set of images of a Lambertian object to estimate the harmonic images of the object and consequently to recover its surface normals and albedos. The methods that we present will be appropriate under the same assumptions as in Basri and Jacobs (2003) and Ramamoorthi and Hanrahan (2001), namely, for a convex object illuminated by distant and isotropic lights that are otherwise unconstrained.

3. Shape Recovery

We assume that a number of images of an object are taken from the same viewpoint, but with different illumination. Denote the matrix of measurements by M. M is $f \times n$ where f denotes the number of images and n denotes the number of pixels in each image (so every image is a row in M). Then, M can be approximated by linear combinations of the harmonic images, that is,

$$M \approx LS, \tag{4}$$

where L ($f \times r$) contains the low order coefficients of the lighting and S ($r \times n$) contains the harmonic images. That is, r is either four or nine, and S is either $\mathbf{S_4}$ or $\mathbf{S_9}$. Our goal is to recover the harmonic images, S, since it is straightforward then to infer the surface normals and the albedos of the object.

The first step is to factor M using Singular Value Decomposition (SVD). Assuming $f, n \geq r$, L and S can be recovered up to a $r \times r$ linear ambiguity. Such a method was proposed by Hayakawa (1994) for the 3D linear space characterized in Shashua (1997) and Moses (1993) and by Koenderink and van Doorn (1997) for the 4D space that contains the zero and first order harmonics. Using SVD we obtain

$$M = U \Delta V^T \tag{5}$$

where U ($f \times f$) and V ($n \times n$) are orthonormal and Δ ($f \times n$) is diagonal and contains the singular values of M. The bulk of the energy in the images is contained in the first r components. Consequently:

$$M \approx \tilde{L}\tilde{S}, \tag{6}$$

where $\tilde{L} = U\sqrt{\Delta^{(fr)}}$ and $\tilde{S} = \sqrt{\Delta^{(rn)}}V^T$, where $\Delta^{(fr)}$ (and $\Delta^{(rn)}$) denote the first r columns (and respectively the first r rows) of Δ. The notation $\sqrt{\Delta}$ denotes the non-negative square roots of the components of Δ.

Both the first and second order methods are based on the assumption that the low order harmonics will lie very close to the subspace constructed by SVD. This is reasonable, since we know that they account for most of the energy in the images. The 4D method assumes that the first four harmonics span approximately the same space as the first four principal components of the images. This makes sense because we know that the 4D space produced by SVD is the best 4D approximation to the observed images, while the first four harmonics are known analytically to be the best possible approximation to the set of all images an object produces. Based on this assumption, we then determine which sets of scaled surface normals lie in this space with an albedo that also lies in this space. The 9D method makes the weaker assumption that the first order harmonics lie somewhere in the 9D space spanned by nine principal components. This merely assumes that the most significant components of an object's entire set of images will show up in the set of images the object has actually produced. In this case we look for scaled normals that lie in the 9D space, and generate harmonic images that span a similar space.

Ramamoorthi (2002) has analyzed the relationship between the subspaces produced by PCA and by images generated with harmonic lighting. It is a subject of future work to use these results to analyze the accuracy of each of our methods.

3.1. *The Case of Four Harmonics*

We first deal with the case of four harmonics ($r = 4$). We have factored M to $M \approx \tilde{L}\tilde{S}$, where M is $f \times n$, $\tilde{L}$ is $f \times 4$ and $\tilde{S}$ is $4 \times n$. This factorization is non-unique up to a 4×4 nonsingular linear transformation,

A, since $\tilde{L}\tilde{S} = \tilde{L}A^{-1}A\tilde{S}$. So $S \approx A\tilde{S}$ for an unknown matrix A. In a second step we now show how to use a constraint on S to reduce this ambiguity to a seven degree of freedom scaled Lorentz transformation. This ambiguity can be removed with additional constraints such as integrability.

Notice that every column $\vec{p} = (p_1, p_2, p_3, p_4)^T$ in S satisfies: $p_1^2 = p_2^2 + p_3^2 + p_4^2$ (since p_1 is the albedo at a point, and p_2, p_3 and p_4 are the components of the surface normal scaled by the albedo). This can be written in matrix notation as

$$\vec{p}^{\,T} J \vec{p} = 0, \tag{7}$$

where $J = \text{diag}\{-1, 1, 1, 1\}$. Note the geometric interpretation of this constraint; every column of S is a point on the surface of the canonical unit sphere in projective space $\mathcal{P}^3$. This constraint is generally not true for $\tilde{S}$, which may be a linear transformation of S. We therefore reduce the ambiguity by finding a linear transformation that forces the points to lie on the unit sphere.

The unknown ambiguity matrix A maps each column $\vec{q}$ of $\tilde{S}$ to its corresponding column p in S, i.e., $\vec{p} = A\vec{q}$. Substituting for $\vec{p}$ in (7) we obtain

$$\vec{q}^{\,T} A^T J A \vec{q} = 0. \tag{8}$$

Denoting $B = A^T J A$ (B is 4×4), this constraint becomes

$$\vec{q}^{\,T} B \vec{q} = 0. \tag{9}$$

This equation is linear and homogeneous in the components of B. Note that B is symmetric, so the equation has 10 unknowns, and so at least 9 points are required to determine B up to an unknown scale factor. To solve for B we construct a system of equations

$$Q\vec{b} = 0. \tag{10}$$

Q is $n \times 10$, and every row of Q is based on information contained in one column of $\tilde{S}$. So for a column $\vec{q}$ in $\tilde{S}$ the corresponding row in Q is $(q_1^2, \ldots, q_4^2, 2q_1q_2, \ldots, 2q_3q_4)$. $\vec{b}$ is a 10-dimensional vector $\vec{b} = (b_{11}, \ldots, b_{44}, b_{12}, \ldots, b_{34})^T$, where b_{ij} are the elements of B. So we can find B up to a scale factor by looking for the null space (or the best approximation to the null space) of Q. As a result of this step we find a 4×4 matrix

$$\tilde{B} = \beta B = \beta A^T J A \tag{11}$$

for some unknown scalar $\beta \neq 0$.

From $\tilde{B}$ we can obtain an estimate of the ambiguity matrix A up to scaled Lorenz transformation. Denote our estimate $\tilde{A}$ ($\tilde{A}$ too is 4×4), we obtain this estimate by factoring $\tilde{B}$ to a product of the form

$$\tilde{B} = \pm \tilde{A}^T J \tilde{A}. \tag{12}$$

Since $\tilde{B}$ is symmetric all its eigenvalues are real. Moreover, if all the previous assumptions hold one of the eigenvalues of $\tilde{B}$ must differ in sign from the remaining three eigenvalues. In other words, one eigenvalue may be negative while the other three are positive (or vice versa). To see this observe that

$$\begin{aligned} \det \tilde{B} &= \det(\beta A^T J A) \\ &= \beta^4 \det(J) \det^2(A) = -\beta^4 \det^2 A. \end{aligned} \tag{13}$$

Consequently, for nonsingular A and nonzero β, $\det \tilde{B} < 0$. Since $\det \tilde{B} < 0$ is a product of the eigenvalues of $\tilde{B}$ then clearly one of the eigenvalues must differ in sign from the remaining three.

We can use this property to factor $\tilde{B}$ as follows. We apply an eigenvalue decomposition to $\tilde{B}$,

$$\tilde{B} = WJ\Lambda W^T, \tag{14}$$

where the columns of the 4×4 matrix W contain the eigenvectors of $\tilde{B}$ and Λ is a 4×4 diagonal matrix that includes the absolute values of the eigenvalues of $\tilde{B}$. WLOG we order Λ and W so that the negative eigenvalue is first. (If there is only one positive eigenvalue we reverse the sign of $\tilde{B}$.) Next we define $\tilde{A} = \sqrt{\Lambda} W^T$, and so $\tilde{B} = \tilde{A}^T J \tilde{A}$. (Again, the notation $\sqrt{\Lambda}$ denotes the non-negative square roots of the components of Λ). When the assumptions do not strictly hold, or when there is significant noise, the eigenvalues of $\tilde{B}$ may not have the proper signs. In that case we resort to an iterative computation to find $\tilde{A}$ that minimizes the Frobenius norm

$$\| \pm \tilde{B} - \tilde{A}^T J \tilde{A} \|, \tag{15}$$

e.g., using gradient descent optimization. (Note that we must minimize this expression for both $\tilde{B}$ and $-\tilde{B}$ since the sign of β is unknown).

At this point we have recovered a valid ambiguity matrix $\tilde{A}$, so that the columns of our estimated shape matrix, given by $\tilde{A}\tilde{S}$, lie on the surface of the canonical unit sphere in projective space. However, there is still an ambiguity remaining because some linear transformations of $\tilde{A}\tilde{S}$ maintain the columns on the sphere. Specifically, the factorization of $\tilde{B}$ is not unique since we can obtain equally valid factorizations by multiplying $\tilde{A}$ by any 4×4 matrix $\tilde{C}$ that satisfies

$$\tilde{C}^T J \tilde{C} = J. \tag{16}$$

In other words, the matrix $\tilde{C}\tilde{A}$ is a valid ambiguity matrix for every matrix $\tilde{C}$ that satisfies (16), since, using (12),

$$\tilde{B} = \pm \tilde{A}^T J \tilde{A} = \pm \tilde{A}^T C^T J C \tilde{A}. \tag{17}$$

Matrices that satisfy this condition (16) represent all the projective transformations that map the unit sphere onto itself. This set of transformations forms the Lorentz group that arises in Einstein's relativity theory and many other disciplines (for its use in vision, see Kanatani (1993)).

A Lorentz transformation has six degrees of freedom. This is because the quadratic form $\tilde{C}^T J \tilde{C} = J$ provides ten quadratic equations (the form is symmetric) in 16 unknowns, the components of $\tilde{C}$. These degrees of freedom include three rotations of the (scaled) surface normals and three imaginary rotations that blend the albedo with the scaled surface normals.

Together with the unknown scale factor β we obtain a seven parameter ambiguity. Let C satisfy $A = C\tilde{A}$ (recall that A is the matrix that separates $\tilde{S}$ from the true harmonic space S, and that $\tilde{A}$ is the matrix obtained by enforcing the quadratic constraint), then using $\tilde{B} = \beta A^T J A$ and $\tilde{B} = \tilde{A}^T J \tilde{A}$ we see that C must satisfy $\beta C^T J C = J$.

We can resolve the ambiguity, for example, if we know the surface normals and albedos in two points. Or, we can remove the ambiguity by enforcing integrability as in Yuille et al. (1999).

In performing the procedure above we need to keep an eye on some numerical issues. In particular, the eigen-decomposition of $\tilde{B}$ (Eqs. (14) and (15)) is sensitive to the particular selection of $\tilde{L}$ and $\tilde{S}$, and its results can be made numerically stable by an initial step of normalization. This is analogous to the problem of motion estimation with a perspective camera, where a preliminary step of normalization has been proved essential (e.g., Hartley's normalization (Hartley, 1997)). In our experiments we noticed that in many cases the first row of $\tilde{S}$ was 3–4 orders of magnitude higher than the other three rows. Such discrepancies are further amplified by the eigen-decomposition process and can thus lead to large numerical errors. Since we are free to modify $\tilde{S}$ and $\tilde{L}$ by applying any non-singular 4×4 matrix to $\tilde{S}$ and its inverse to $\tilde{L}$ we first scaled the first row of $\tilde{S}$ to bring it to the same order of magnitude as the other rows. In our experiments we observed that such normalization significantly improved the accuracy of the reconstruction. Moreover, when a normalization was not applied, we often obtained a matrix $\tilde{B}$ with incorrect signs of eigenvalues, the numerical factorization (15) did not converge to a low value due to numerical errors, and consequently the ambiguity that remained was a non-Lorenz transformation.

In summary, the initial equation $M \approx \tilde{L}\tilde{S}$ tells us already that the scaled surface normals lie in the row space of $\tilde{S}$, obtained by SVD. Since $\tilde{S}$ has four rows, this leaves sixteen degrees of freedom in the albedo and the three scaled normals. We have shown that the constraint that the first row of S, the albedos, must equal the norm of the other three rows, the scaled surface normals, reduces these degrees of freedom to seven. We have also shown an effective procedure for computing a valid harmonic space by constructing a matrix $\tilde{A}$ that can be applied to $\tilde{S}$ to provide the albedos and scaled surface normals, up to this ambiguity.

Our complete algorithm is composed of the following steps:

1. Begin with a set of images, each composing a row of the matrix M.
2. Using SVD $M = U\Delta V^T$, factor $M = \tilde{L}\tilde{S}$, where $\tilde{L} = U\sqrt{\Delta^{(f4)}}$ and $\tilde{S} = \sqrt{\Delta^{(4n)}}V^T$.
3. Normalize $\tilde{S}$ by scaling its rows so as to have equal norms.
4. Construct Q. Each row of Q is constructed with quadratic terms computed from a column of $\tilde{S}$.
5. Using SVD, construct $\tilde{B}$ to approximate the null space of Q (i.e., solve $Q\vec{b} = 0$ and compose $\tilde{B}$ from the elements of $\vec{b}$).
6. Construct $\tilde{A}$.
   (a) If $\tilde{B}$ has exactly one positive or one negative eigenvalue, construct $\tilde{A}$ with an eigenvalue decomposition as follows:
      (i) If $\tilde{B}$ has exactly one positive eigenvalue, reverse its sign ($\tilde{B} \leftarrow -\tilde{B}$).

(ii) Then, $\tilde{B}$ has exactly one negative eigenvalue. Apply an eigenvalue decomposition $\tilde{B} = WJ\Lambda W^T$, and define $\tilde{A} = \sqrt{\Lambda}W^T$.

(b) Otherwise, perform an iterative computation to find $\tilde{A}$ that minimizes: $\| \pm \tilde{B} - \tilde{A}^T J \tilde{A}\|$.

7. Compute the structure $\tilde{A}\tilde{S}$, which provides the scene structure up to a scaled Lorentz transformation.

3.2. The Case of Nine Harmonics

We now present a method based on the weaker assumption that the scaled surface normals lie in the space spanned by the first nine principal components. Since at least 98% of the energy in the reflectance function lie in a 9D space (Basri and Jacobs (2003)), higher order components have little effect on the images, and therefore on the 9D space found by SVD.

3.2.1. Recovery. As before, let M denote a matrix containing the images. Our objective is to find an $f \times 9$ lighting matrix L and a $9 \times n$ shape matrix S that minimize

$$E = \min_{L,S} \|M - LS\|, \tag{18}$$

where $\|.\|$ denotes the Frobenius norm. The rows of S are further restricted to contain the second order harmonics of some shape, i.e., they include $\vec{h}_1, \ldots, \vec{h}_9$ in the form defined in Section 2.2.

We approach this problem as follows. We first use SVD to construct a 9D approximation such that $M \approx \tilde{L}\tilde{S}$. So $\tilde{L}$ is an $f \times 9$ matrix, and $\tilde{S}$ is $9 \times n$. If we assume that the scaled surface normals lie in the row space of $\tilde{S}$, then we can structure our search by seeking a 3×9 matrix $\mathbf{A}$, such that:

$$\mathbf{A}\tilde{S} = (\vec{h}_2, \vec{h}_3, \vec{h}_4)^T. \tag{19}$$

Given $\mathbf{A}$, we have an estimate of an object's structure. We can evaluate how well this structure matches the observed images by comparing it to the 9D linear subspace generated by the harmonic images. This linear method of comparing a model to an image is described in Basri and Jacobs (2003), but we review it briefly here. From $\mathbf{A}\tilde{S}$, we construct a $9 \times n$ matrix, S_A, containing the harmonic images of $\mathbf{A}\tilde{S}$. Then we determine how well this matches all images by computing the error (18), $E(\mathbf{A})$, as:

$$E(\mathbf{A}) = \min_{L} \|M - LS_A\| \tag{20}$$

where L is chosen so as to minimize the error. Finding this error is a linear problem. We can solve it by projecting M onto the space spanned by the harmonic images in S_A, and then measuring the distance from M to this projection. Our goal, then, is to find $\mathbf{A}$ that minimizes E.

We do this using an iterative optimization. First, as a starting point, we guess that the scaled surface normals will be the second, third and fourth rows of $\tilde{S}$, which are associated with the second through fourth largest singular values of M. We know that in theory, the scaled surface normals, which are the first order harmonics, are most important after the DC component, which contains the albedos of the object. We then can use any general purpose optimization method to try to find the $\mathbf{A}$ that minimizes E, from this starting point.

3.2.2. A Linear Ambiguity. When we represent an object's images using a 9D linear subspace, the question of ambiguity becomes: when do two different sets of scaled surface normals lead to the same 9D space? It is straightforward to show that if we rotate the scaled surface normals or scale their components uniformly, we will not change the 9D space of their harmonic images. Therefore, we know that our method can only recover the scaled normals up to a scaled rotation, at best. However, we also know that applying an arbitrary linear transformation to the scaled surface normals of an object will not change the entire set of images that it produces, called its *illumination cone* (Belhumeur et al., 1999). On the other hand, applying a 3×3 linear transformation to the scaled normals of an object *does* change its 9D space of harmonic images. This is easily verified numerically. So a linear transformation alters our 9D approximation to the illumination cone without altering the cone itself. We know the 9D approximation is quite accurate, and this leads us to suspect that our approach can only accurately recover the surface normals up to a linear transformation.

In the appendix we provide some evidence that this is indeed the case. We show analytically that small linear transformations applied to the surface normals will have an especially small effect on their first four harmonic images. We also show some simple simulations that quantify this. As a consequence, we expect that any solution produced by our 9D algorithm will be unstable with respect to linear transformations of the surface

normals. So our solutions should be considered valid only up to an unknown linear transformation.

In sum, the method of this section finds a set of scaled surface normals whose harmonic images match the observed images well. Two things make it possible to solve this problem efficiently. One is that because the true scaled surface normals lie near the 9D space produced by performing PCA on the images, we have reduced the number of unknowns needed to specify the scaled surface normals in a scene from $3n$ to 27. The second is that we can use simple linear methods to evaluate how well a set of normals fit the observed images. We can then solve this problem with straightforward optimization techniques, whose effectiveness is demonstrated in Section 5. However, we also show that this method can only stably recover the normals up to a linear transformation.

4. From Normals to Surfaces

Our work focuses on finding the scaled surface normals of an object. Often, we would like to turn these into an integrable surface, with height, $z = f(x, y)$ so that the partial derivatives $(\partial f/\partial x, \partial f/\partial y, -1)$ match the normals. We face two problems here, the linear ambiguity in the normals, and turning a set of normals into a surface.

The linear ambiguity can be resolved by finding a linear transformation of the normals that make them consistent with an integrable surface. This can be done up to a subset of transformations called the *generalized bas-relief* transformations (see Belhumeur et al. (1999) and Yuille et al. (1999)). Since our work focuses on finding the scaled normals, we remove remaining ambiguities by hand in our experiments, but here we make a few remarks about the interaction between the integrability constraint and the ambiguities of our approach.

With the addition of an integrability constraint, the 9D method can determine a surface up to a bas-relief transformation. However the bas-relief transformations are different from the Lorentz transformations. For example, if we scale the z component of the scaled surface normals, this is a simple bas-relief transformation. It preserves the first order harmonics, but transforms the albedos outside the space spanned by the scaled surface normals (this follows from the derivation in Section 3.2.2), and therefore changes the 4D harmonic space of the normals. So this is not a Lorentz transformation.

In fact, because the Lorentz ambiguity is different from the bas-relief ambiguity, adding integrability to the 4D method should in theory lead to a unique solution for the surface. This amounts to removing the bas-relief ambiguity using the statistical assumption that the first four harmonic images can be identified as the dominant components of the image. Our analysis of the ambiguity in Section 3.2.2, however, indicates that removing the ambiguity with this approach may be sensitive to measurement errors.

Once we have resolved any linear ambiguity, we may also wish to turn the normals into a surface. This is straightforward; we can use standard techniques to fit the surface, which has one degree of freedom per pixel, to unit surface normals, which have two degrees of freedom, by solving a quadratic minimization with linear constraints.

Specifically, denote the surface by $z(x, y)$. The directions of the normals are given by $n(x, y) = (p, q, -1)^T$, with $p = z_x$ and $q = z_y$ where z_x and z_y denote the partial derivatives of z with respect to x and y respectively. The (recovered) scaled surface normal (n_x, n_y, n_z) roughly satisfies

$$(n_x, n_y, n_z) = \frac{\rho}{\sqrt{p^2 + q^2 + 1}}(p, q, -1), \quad (21)$$

from which we obtain

$$p = -\frac{n_x}{n_z}$$
$$q = -\frac{n_y}{n_z}, \quad (22)$$

On a discrete grid, we may approximate p and q using forward differences by

$$p \approx z(x + 1, y) - z(x, y)$$
$$q \approx z(x, y + 1) - z(x, y). \quad (23)$$

Combining (22) with (23) we obtain the following constraints:

$$n_z z(x + 1, y) - n_z z(x, y) = n_x$$
$$n_z z(x, y + 1) - n_z z(x, y) = n_y. \quad (24)$$

These are merely linear constraints on $z(x, y)$, and can be solved as an overconstrained linear system. Note that the constraints are invalid near the rim of the object since there $n_z \approx 0$. In this case a different constraint can be used. This constraint depends on the two

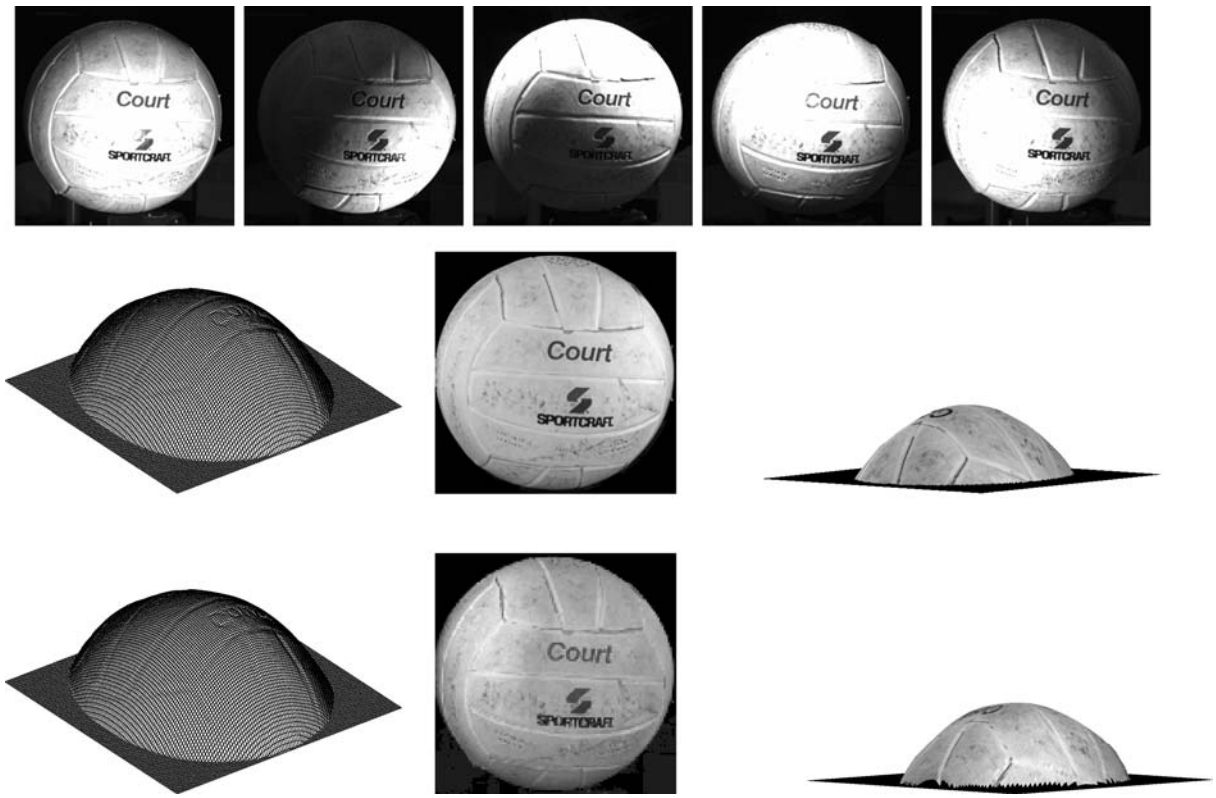

Figure 1. Top row: Five of the 64 volleyball images used for reconstruction. Middle row: reconstruction using the 4D method, including the recovered 3D shape (left), the recovered albedo (middle) and the recovered shape painted with the albedo. Bottom row: reconstruction using the 9D method.

Figure 2. Blowups of the reconstructed volleyball (using the 4D method) reveals the stripes (middle) and the logo and writing (right).

above constraints, but does not involve n_z. The constraint comes from the fact that $p/q = n_x/n_y$, and is given by:

$$n_y(z(x, y)-z(x+1, y)) = n_x(z(x, y) - z(x, y+1)). \tag{25}$$

We can solve the obtained linear set of equations using a least squares fit or we can apply in addition boundary constraints due to the rim points of the object (where the values of p and q can be estimated directly from the image).

5. Experiments

We now present experiments to evaluate these methods. Because they ignore higher order harmonics, even in the absence of any sensing error, our methods will have

Figure 3. A pair of images of a face statue obtained with a point source (left and middle) and their average produced in order to obtain an image with two point sources (right).

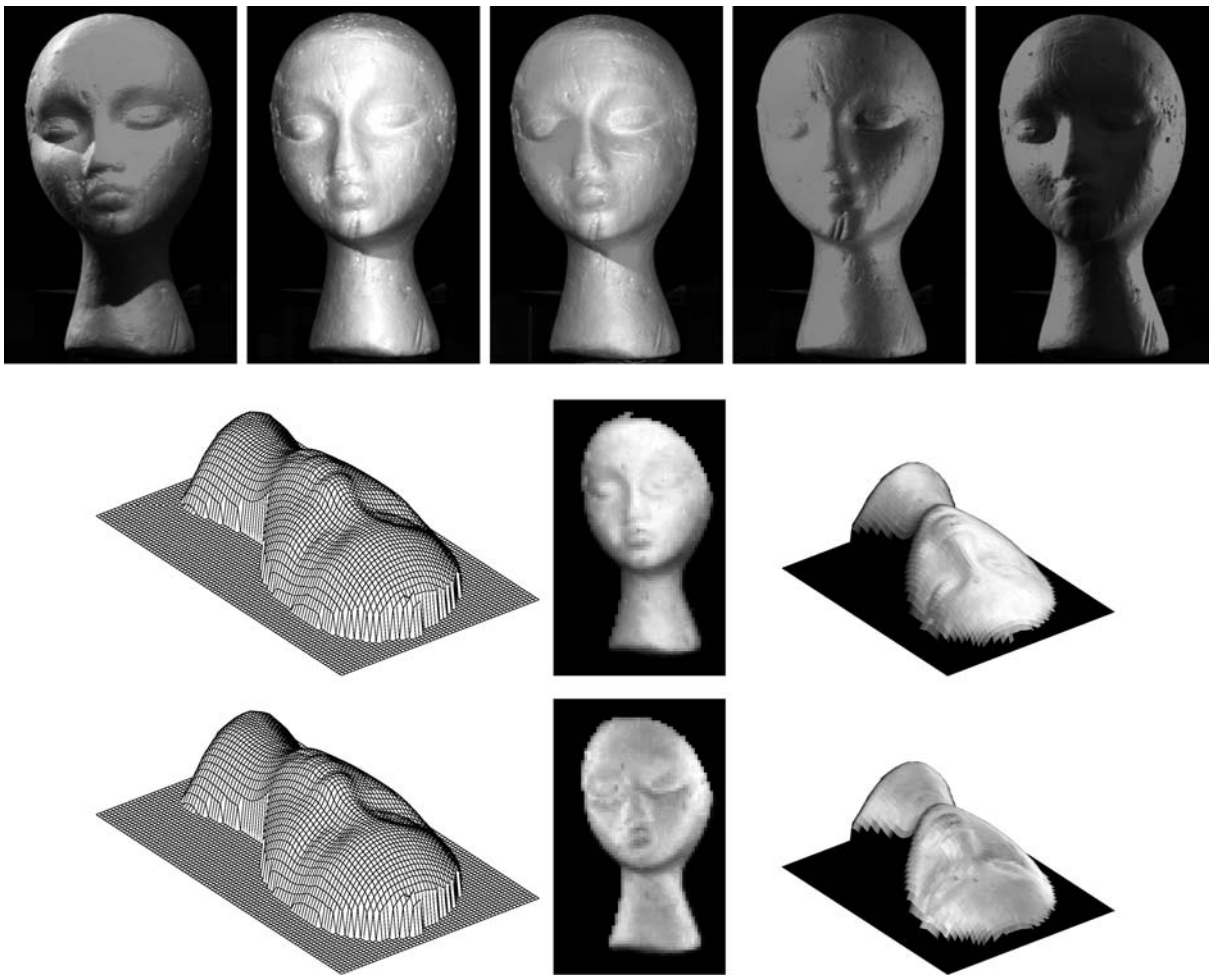

Figure 4. Top row: Five (of 32) averaged images used for reconstruction. Saturated pixels shown in white. Middle row: reconstruction using the 4D method (shape, albedo, and albedo painted shape). Bottom row: reconstruction using the 9D method.

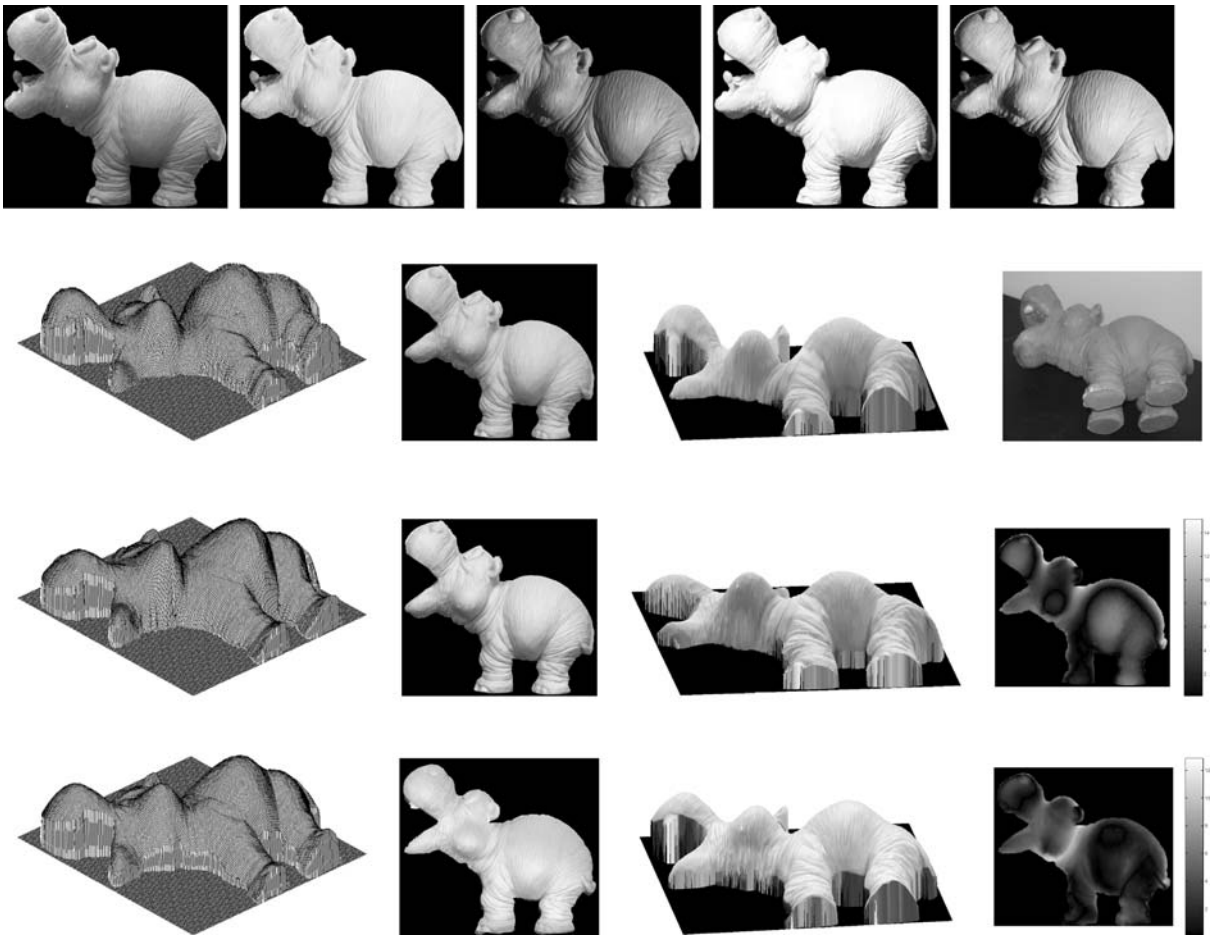

Figure 5. Top row: Five (of 11) images used for reconstruction. Second row: ground truth obtained with a laser scanner (surface, albedo, and albedo painted surface) and an image taken from roughly the same view (right). Third row: reconstruction using the 4D method (including shape, albedo, albedo painted shape and difference from ground truth surface). Bottom row: reconstruction using the 9D method.

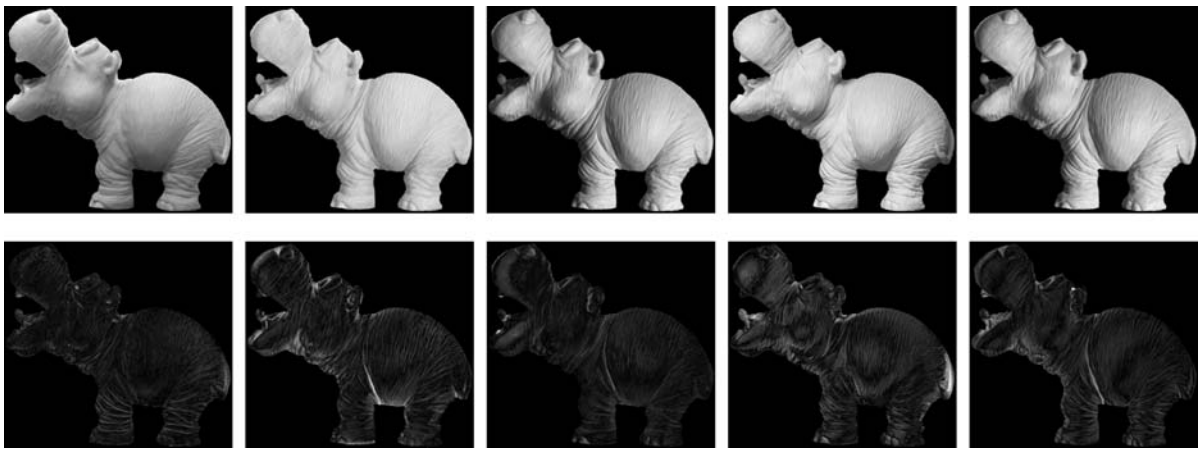

Figure 6. Top row: Rendering the reconstructed shape so as to best fit the original images (top row of Fig. 5). Bottom row: difference between rendered images and the original images.

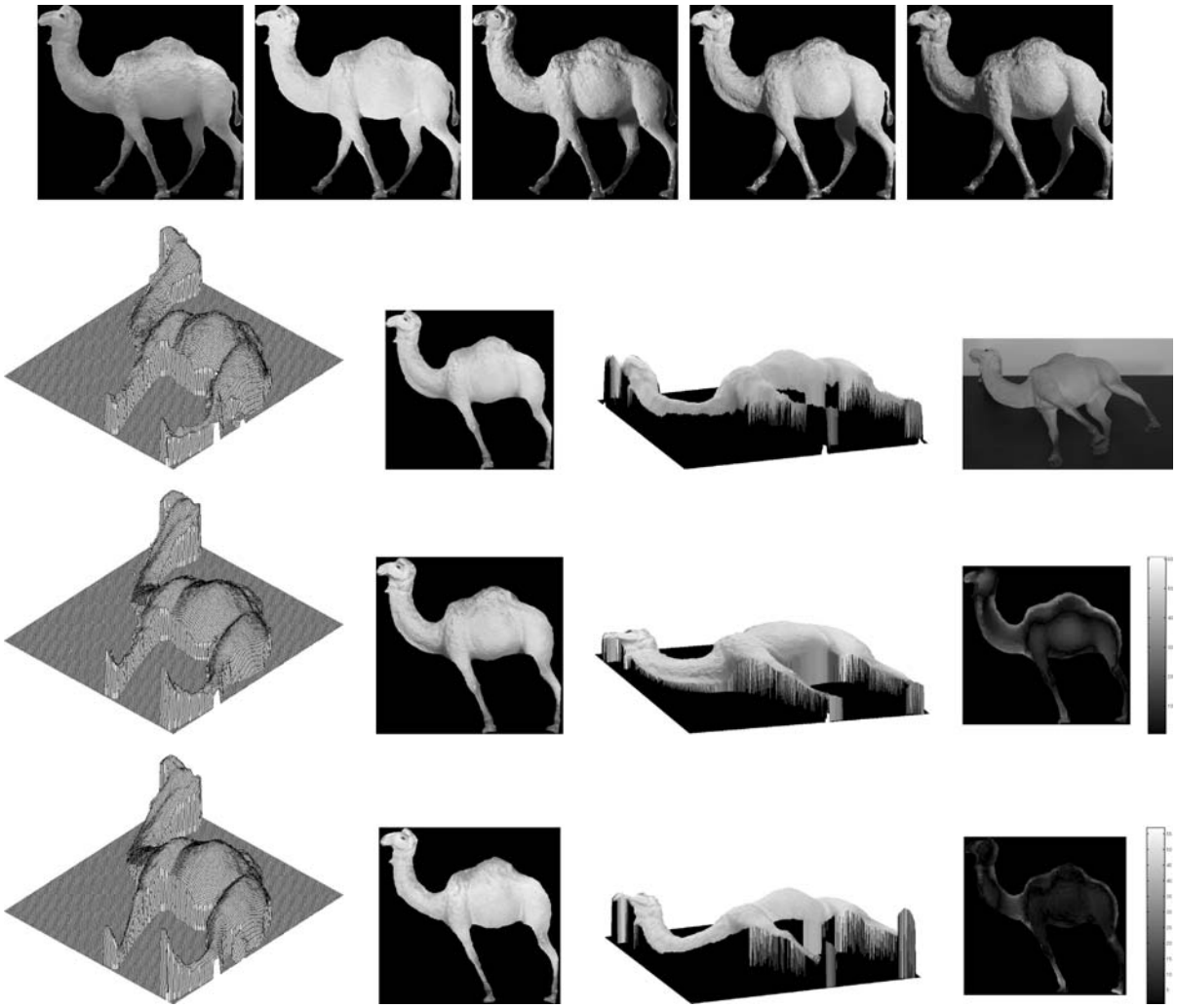

Figure 7. Top row: Five (of 13) images used for reconstruction. Second row: ground truth obtained with a laser scanner (surface, albedo, and albedo painted surface) and an image taken from roughly the same view (right). Third row: reconstruction using the 4D method (including shape, albedo, albedo painted shape and difference from ground truth surface). Bottom row: reconstruction using the 9D method.

some built in error. So we first describe experiments on synthetic data to establish some basic properties of the methods.

We generate square surfaces with random heights, and extract 81 surface normals from them. We then generate 20 images of each surface, using lighting conditions that are a combination of several random point sources and a diffuse component for each image. We then solve for the normals using the above two methods. For the 9D method, we optimize using the MATLAB minimization routine "fminunc," which performs a line search in the gradient direction. This has the virtue of being the easiest possible method to program. This method requires only a few seconds for small simulations and about an hour for real images, sufficient for our experiments.

Since we have ground truth available, once we compute surface normals, we can find the linear or Lorentz transformation that best fits our solution to the correct one. Then we measure the average angle between the true surface normals and the recovered ones. Repeating this 400 times, for the 9D method we find a mean error of 2.8 degrees, and error of 3.6 degrees for the 4D method. (The standard deviation of the means are 0.04 and 0.12 degrees, respectively.) This tells us that our method will produce small errors even in the absence

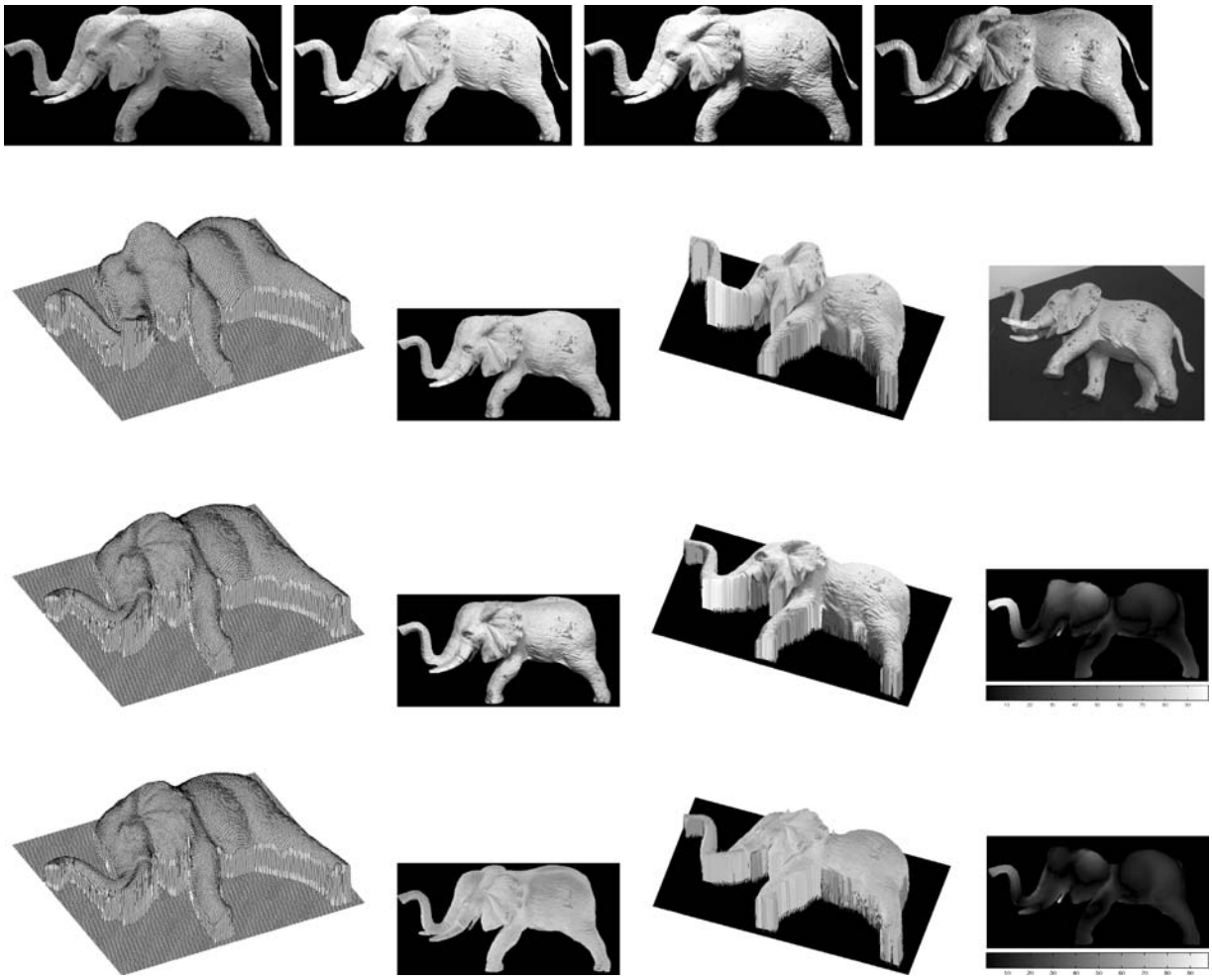

Figure 8. Top row: Four (of 11) images used for reconstruction. Second row: ground truth obtained with a laser scanner (surface, albedo, and albedo painted surface) and an image taken from roughly the same view (right). Third row: reconstruction using the 4D method (including shape, albedo, albedo painted shape and difference from ground truth surface). Bottom row: reconstruction using the 9D method.

of sensing error.

We can also use synthetic data to estimate how often optimization finds an optimal solution in the 9D case. To do this, we can project the scaled surface normals onto the 9D SVD space of the images. This is the solution closest to ground truth in the space the algorithm searches. We use this as a starting point for optimization to produce an estimate of the minimal error solution. We find that 97% of the time, our algorithm finds a solution in which the error function is no more than 1% greater than this solution.

In a second set of experiments we have run our algorithms on six sets of real images. In some cases, these images contain unreliable pixels that have been saturated. We remove these, and use Wiberg's algorithm (Wiberg, 1976) to fill in the missing data (see also Jacobs (2001) and Shum et al. (1995)). In each case, to resolve any ambiguity of the method we generate a surface after applying a transformation obtained by matching some points in the scene with hand chosen surface normals.

In the first two experiments we use images taken by researchers at Yale under controlled lighting. In the first experiment (Fig. 1) we use 64 images containing a volleyball lit by a single point source. These controlled images could be used by other algorithms (e.g., Yuille et al. (1999)), but our algorithms do not make any assumptions that take advantage of the presence of

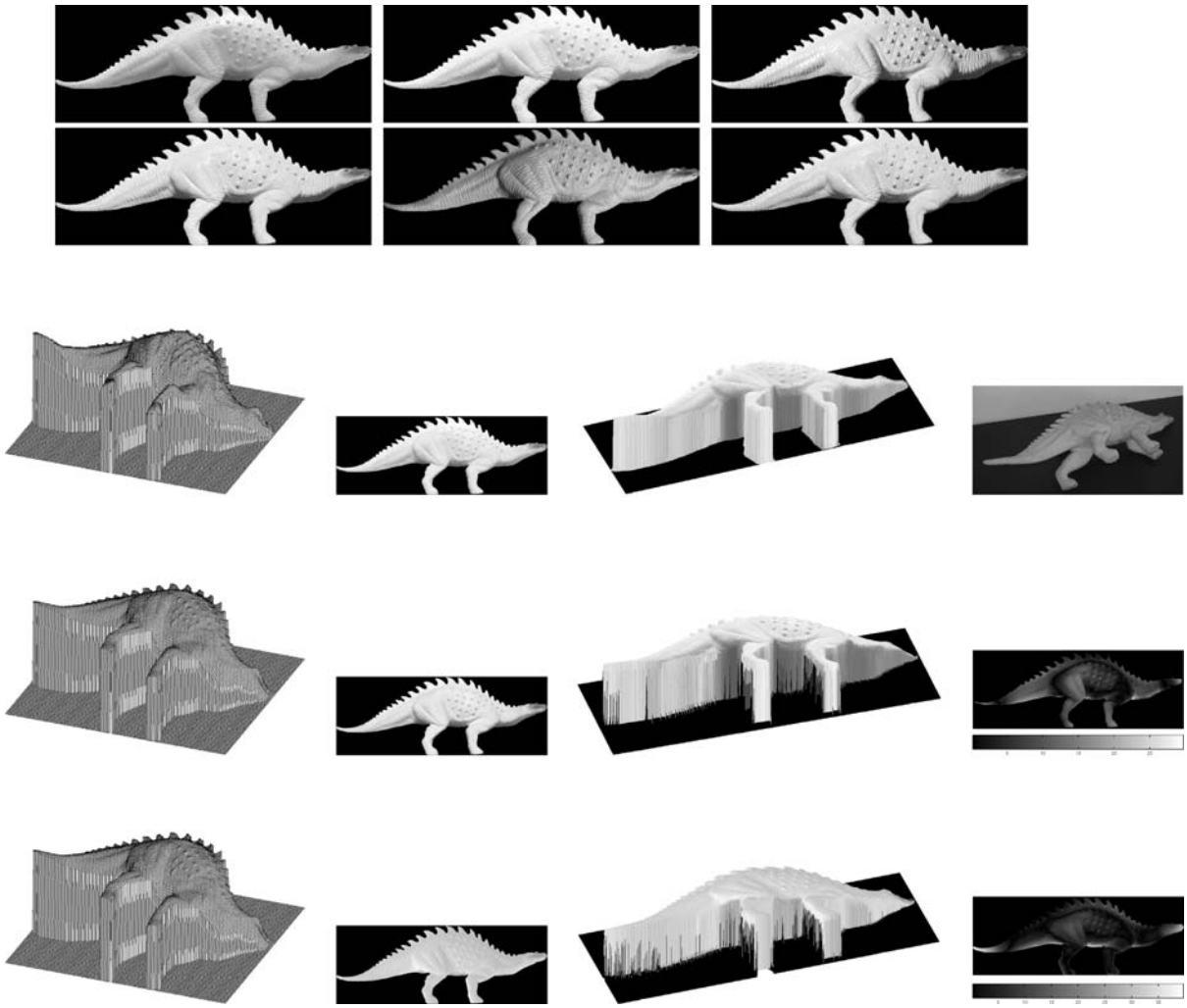

Figure 9. Top two rows: Six (of 10) images used for reconstruction. Third row: ground truth obtained with a laser scanner (surface, albedo, and albedo painted surface) and an image taken from roughly the same view (right). Fourth row: reconstruction using the 4D method (including shape, albedo, albedo painted shape and difference from ground truth surface). Bottom row: reconstruction using the 9D method.

a single source in each image. Figure 2 shows a blowup of the reconstruction to demonstrate its accuracy. Note the stripes and writing carved on the volleyball surface. Next, we use a similar set of controlled images of a statue of a face, but in this case we average pairs of images, to simulate having 32 images with two point sources in each image (Fig. 3). Pixels are marked saturated in the average image if saturated in either one of the original images. Results are shown in Fig. 4.

In the final four experiments we use images obtained under uncontrolled lighting that include different, unrestricted combinations of fluorescent lights, a light projector, and a desk lamp. In each case we provide 10–15 images to the algorithms. The images include toy objects with dimension along the longest axis is about 15 cm. The camera was placed at a distance of 1.5 m from the objects and the lights were placed at different locations in the room at distances between 1–2.5 m from the object. The images included 1024×768 pixels with the reconstructed objects taking the better part of these images. We further cropped the images before applying our algorithms. Reconstruction results are shown in Fig. 5–9. For comparison, we also used a handheld laser scanner to obtain ground truth surfaces. In addition in each case we display an image of the object taken from roughly the same

Table 1. Relative accuracy of reconstruction: $1 - \|\hat{z} - z\|^2/\|z\|^2$, where $\hat{z}$ denotes the reconstructed depth and z denotes the laser scanned depth.

	4D	9D
Hippo	0.96	0.97
Elephant	0.95	0.98
Camel	0.97	0.99
Dino	0.98	0.99

view. Using a Matlab implementation on a PC with a Pentium 4 processor our non-optimized implementation of the 4D method took 20–30 seconds to reconstruct each set and about an hour with the 9D method.

In general, the results of both our 4D and 9D algorithms appear to be very similar to those produced by the laser scanner, with the 9D algorithm producing slightly more accurate results than those produced by the 4D method. Difference maps are included in all figures showing the differences between our reconstructions and the laser scanned shapes. Table 1 further quantifies this, showing the relative accuracy of our reconstructions with respect to the laser scanned shapes. As can be seen, the 9D method reaches an accuracy of 97–99%, whereas the 4D method reaches an accuracy of 95–98%. We are unaware of any other algorithm designed to handle unknown lighting for such complex lighting conditions.

6. Summary and Conclusions

This paper describes new methods for recovering the surface normals in a scene using images produced under very general unknown lighting conditions. The first insight that allows us to do this is that results due to Basri and Jacobs (2003) and Ramamoorthi and Hanrahan (2001) show that even under very general lighting conditions, the scaled surface normals of a Lambertian object will lie in the low-dimensional space spanned by the principal components of the image. This reduces the search for surface normals to a problem with relatively few variables.

We have presented two reconstruction methods, one that uses a first order harmonic approximation operating in a 4-dimensional space, and a second method that uses a second order harmonic approximation operating in a 9-dimensional space. The 9D method produces slightly more favorable results, but involves an iterative optimization and requires at least nine images. In contrast, the 4D method is based on a sequence of closed form formulae, and requires fewer images (at least four images). Both methods lead to fairly accurate reconstructions. The small errors obtained can be attributed to several factors, including the accuracy of the first and second order harmonic approximation, the extent to which the low order harmonics coincide with the linear spaces derived with SVD, the adequacy of the Lambertian model, camera noise, etc.

Appendix

In this appendix we show analytically that small linear transformations applied to the surface normals will have an especially small effect on their first four harmonic images. We later demonstrate this with simulations. We suppose that we apply a linear transformation to the scaled surface normals of an object, $(\vec{h}_2, \vec{h}_3, \vec{h}_4)^T$. We denote this 3×3 transformation by T and denote

$$(\vec{h}'_2, \vec{h}'_3, \vec{h}'_4)^T = T(\vec{h}_2, \vec{h}_3, \vec{h}_4)^T, \tag{26}$$

where $\vec{h}'_i$ represents the harmonic images that result from the transformation.

We first note that, according to the above equation, $\vec{h}_2, \vec{h}_3$ and $\vec{h}_4$ all lie in the linear subspace spanned by $\vec{h}'_2, \vec{h}'_3$ and $\vec{h}'_4$. So any image effects produced by first order harmonic lighting can be accounted for by any surface normals that are a linear transformation of the true surface normals.

We next consider the effect of a small linear transformation on the first harmonic image, $\vec{h}_1$. We consider the derivative of the angle between $\vec{h}_1$ and the space spanned by $\vec{h}'_1, \ldots \vec{h}'_9$, with respect to a small linear transformation of the surface normals. We show that this derivative is zero. Consequently, a small linear transformation of surface normals has essentially no effect on their ability to account for zero and first harmonic components of a set of images.

Using SVD, we can write T as $R_1 D R_2$, where R_1 and R_2 are 3×3 rotations, and D is a 3×3 diagonal matrix. We can understand the effect of T on the 9D subspace generated by the scaled surface normals by considering the effect of each of these transformations. Rotating the scaled surface normals causes a phase shift; each harmonic image becomes a linear combination of the harmonics of that order, but no energy is shifted across frequencies. Therefore, the rotations leave the

9D space of harmonic images unchanged, and we need only consider changes caused by the diagonal matrix.

We write $D = \text{diag}\{1+\delta_2, 1+\delta_3, 1+\delta_4\}$, and take the limit as $\delta_i \to 0$. Without loss of generality, it will suffice to consider the case in which $\delta_2 = \delta_3 = 0$. By symmetry, a change in a different direction would have a similar effect on the rotated surface normals. Then our problem reduces to determining how $\vec{h}_1$, the albedo, changes as we scale $\vec{h}_4$ (which is the Hadamard product of the albedo and the z component of the surface normals). For notational simplicity, we consider the harmonic images at just one point.

Our approach will be to determine how a small change to h_4 affects the zero'th order harmonic, h_1, and also the second order harmonic h_5. We will show that the old harmonic h_1 lies in the space spanned by h_1' and h_5', for small δ_4. This will show that the new 9D subspace contains the first four harmonics of the original subspace, and can provide a good approximation to images well approximated by the original subspace.

We want to compute the following derivatives: $\frac{dh_1}{dh_4}$ and $\frac{dh_5}{dh_4}$. For this we use: $h_1 = \sqrt{h_2^2 + h_3^2 + h_4^2}$, and, using the substitutions $\rho = h_1$ and $z = \frac{h_4}{h_1}$:

$$h_5 = \frac{1}{2}h_1\left(\frac{3h_4^2}{h_1^2} - 1\right) = \frac{1}{2}\left(\frac{3h_4^2}{h_1} - h_1\right). \quad (27)$$

These imply:

$$\frac{dh_1}{dh_4} = \frac{h_4}{h_1} = z, \quad (28)$$

and

$$\frac{dh_5}{dh_4} = \frac{\partial h_5}{\partial h_1}\frac{dh_1}{dh_4} + \frac{\partial h_5}{\partial h_4} = \frac{1}{2}\left(-\frac{3h_4^2}{h_1^2} - 1\right)\frac{h_4}{h_1} + \frac{3h_4}{h_1} = -\frac{3}{2}z^3 + \frac{5}{2}z \quad (29)$$

If we now change h_4:

$$h_4' = h_4(1+\delta_4) = \rho z(1+\delta_4) \quad (30)$$

(so $h_4' = h_4 + \triangle h_4$ with $\triangle h_4 = \rho z \delta_4$). Then,

$$h_1' = h_1 + \frac{dh_1}{dh_4}\triangle h_4 + O\big(\delta_4^2\big) = \rho(1+z^2\delta_4) + O\big(\delta_2^2\big), \quad (31)$$

$$\begin{aligned} h_5' &= h_5 + \frac{dh_5}{dh_4}\triangle h_4 + O\big(\delta_4^2\big) \\ &= \rho\frac{1}{2}\big(3z^2-1\big) + \left(-\frac{3}{2}z^3 + \frac{5}{2}z\right)(\rho z\delta_4) + O\big(\delta_4^2\big) \\ &= \frac{\rho}{2}(3z^2 - 1 + (5-3z^2)z^2\delta_4) + O\big(\delta_4^2\big). \end{aligned} \quad (32)$$

Finally,

$$\begin{aligned} h_1' - \frac{2}{3}\delta_4 h_5' &= \rho\big(1+z^2\delta_4\big) - \frac{1}{3}\delta_4\rho \\ &\quad (3z^2 - 1 + (5-3z^2)z^2\delta_4) + O(\delta_4^2) \\ &= \rho\left(1 + \frac{1}{3}\delta_4 + O\big(\delta_4^2\big)\right). \end{aligned} \quad (33)$$

$$h_1\left(1 + \frac{1}{3}\delta_4\right) = h_1' - \frac{2}{3}\delta_4 h_5' + O\big(\delta_4^2\big) \quad (34)$$

When we take the limit as δ_4 goes to zero, we can ignore the $O(\delta_4^2)$ terms. This tells us that the derivative of the angle between $\vec{h}_1$ and the space spanned by $\vec{h}_1', \ldots, \vec{h}_9'$, taken with respect to a linear change in the surface normals, is zero. Therefore, all components of the images due to $\vec{h}_1, \vec{h}_2, \vec{h}_3, \vec{h}_4$ can also be produced by the new 9D linear subspace we get after any linear transformation near the identity.

We show a simple simulation to demonstrate this result. We generate a random object with 100 scaled surface normals. We then apply a linear transformation of the form $D = \text{diag}\{1+\delta_2, 1+\delta_3, 1+\delta_4\}$ to the object. We do this by randomly selecting $(\delta_2, \delta_3, \delta_4)$. We then scale these three values so that their total magnitude varies from zero to one. Near zero the transformation is almost an identity transformation; near one it is significantly different. In Fig. 10 we plot the angle between the new 9D linear subspace of the harmonic images and the old harmonic images. The angles for h_2, h_3 and h_4 are always zero, and not shown. The angle between h_1 and the subspace is shown as a solid curve. We can see that for small changes in the model, the slope of this curve is zero, as our analysis predicts. Moreover, we can see that this curve always remains extremely small, even outside the range in which a first order analysis might be expected to hold. The angles of the five second order harmonics, shown as dashed lines, increase much more rapidly, but still remain less than one degree. Repeating these simulations with different models and other linear transformations yielded similar results.

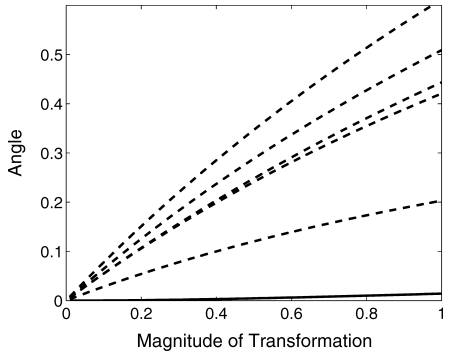

Figure 10. The solid line shows the angle between $\vec{h}_1$ and the 9D harmonic subspace formed after a linear transformation of surface normals, for a random object and a random change. The vertical axis shows the angle, in degrees, while the horizontal axis shows the magnitude of the change. The dashed lines show the same angles for the second order harmonics.

The zero and first order harmonics have a much greater effect on an object's images than the second order harmonics. This means that when the correct scaled surface normals account well for the images, this will be primarily because of the first four harmonic images they produce. Applying a linear transformation to the normals will, to first order, generate a linear space that contains these four harmonics, so that the new harmonics also account well for the images. Moreover, results from simulations, and known results about the illumination cone, suggest that even large linear transformations to a model will have only a small effect on the 9D harmonic subspace that it produces. Therefore, our reconstruction based on the nine, second order harmonics will be unstable to a linear transformation of the normals. We observe this experimentally, as well.

Acknowledgment

Major portions of this research were conducted while the authors were at NEC Research Institute, Princeton, NJ. At the Weizmann Institute Ronen Basri is supported in part by the European Commission Project IST-2000-26001 VIBES, the European Commission Project IST-2002-506766 Aim Shape, and by the Israel Science Foundation grant number 266/02. Additional support was obtained by the US-Israel Binational Science Foundation grant number 2002/254. The vision group at the Weizmann Inst. is supported in part by the Moross Laboratory for Vision Research and Robotics.

Note

1. A preliminary version of this work appeared in Basri and Jacobs (2001).

References

Basri, R. and Jacobs, D. 2001. Photometric stereo with general, unknown lighting. *IEEE International Conference on Computer Vision*, Kauai, II:374–381.

Basri, R. and Jacobs, D. 2003. Lambertian reflectances and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(2):218–233.

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

Belhumeur, P.N., Kriegman, D.J., and Yuille, A.L. 1999. The Bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44.

Barsky, S. and Petrou, M. 2003. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252.

Clark, J.J. 1992. Active photometric stereo. *IEEE International Conference on Computer Vision*, 29–34.

Coleman, Jr., E.N. and Jain, R. 1982. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry, *Computer Graphics and Image Processing*, 18(4):309–328.

Fan, J. and Wolff, L.B. 1997. Surface curvature and shape reconstruction from unknown multiple illumination and integrability. *Computer Vision and Image Understanding*, 65(2):347–359.

Frolova, D., Simakov, D., and Basri, R. 2004. Accuracy of spherical harmonic approximations for images of Lambertian objects under far and near lighting. *European Conf. on Computer Vision*, Prague, I:574–587.

Georghiades, A. 2003. Recovering 3-D shape and reflectance from a small number of photographs. In Christensen, P. and Cohen-Or D. (Eds.), *Eurographics Symposium on Rendering*, pp. 230–240.

Georghiades, A. 2003. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. *IEEE International Conference on Computer Vision*. 816–823.

Hartley, R.I. 1997. In Defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580–593.

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

Hertzmann, A. and Seitz, S.M. 2005. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(8):1254–1264.

Horn, B.K.P. 1986. *Robot Vision*. MIT Press.

Horovitz, I. and Kiryati, N. 2004. Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 22(9):681–694.

Ikeuchi, K. 1981. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 3(6):661–669.

Jacobs, D. 2001. Linear fitting with missing data for structure-from-motion. *Computer Vision and Image Understanding*, 82(1):57–81.

Jacobs, D., Belhumeur, P., and Basri, R. 1998. Comparing images under variable illumination. *IEEE International Conference on Computer Vision* Santa Barbara, 610–617.

Kanatani, K. 1993. *Geometric Computation for Machine Vision*. Oxford University Press.

Kim, B. and Burger, P. 1991. Depth and shape form shading using the photometric stereo method. *Computer Vision, Graphics and Image Processing*, 54(3):416–427.

Koenderink, J. and van Doorn, A. 1997. The generic bilinear calibration-estimation problem. *International Journal of Computer Vision*, 23(3):217–234.

Luong, Q.T., Fua, P., and Leclerc, Y.G. 2002. The radiometry of multiple images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(1):19–33.

Moses, Y. 1993. *Face Recognition: Generalization to Novel Images*. Ph.D. Thesis, Weizmann Inst. of Science.

Nayar, S.K., Ikeuchi, K., and Kanade, T. 1990. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431.

Okatani, T. and Deguchi, K. 2001. On uniqueness of solutions of the three-light-source photometric stereo: Conditions on illumination configuration and surface reflectance. *Computer Vision and Image Understanding*, 81(2):211-226.

Onn, R. and Bruckstein, A.M. 1990. Integrability disambiguates surface recovery in two-image photometric stereo. *International Journal of Computer Vision*, 5(1):105–113.

Ramamoorthi, R. 2002. Analytic PCA construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1322–1333.

Ramamoorthi, R. and Hanrahan, P. 2001. On the relationship between radiance and irradiance: Determining the illumination from images of convex Lambertian object. *Journal of the Optical Society in America A*, 2448–2459.

Sakarya, U. and Erkmen, I. 2003. An improved method of photometric stereo using local shape from shading. *Image and Vision Computing*, 21(11):941–954.

Shashua, A. 1997. On photometric issues in 3D visual recognition from a single 2D image. *International Journal of Computer Vision*, 21(1–2):99–122.

Shum, H.Y., Ikeuchi, K., and Reddy, R. 1995. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(9):854–867.

Simakov, D., Frolova, D., Basri, R. 2003. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. *IEEE Int. Conf. on Computer Vision*, 1202–1209.

Smith, M.L. 1999. The analysis of surface texture using photometric stereo acquisition and gradient space domain mapping. *Image and Vision Computing*, 17(14):1009–1019.

Tagare, H.D. and de Figueiredo, R.J.P. 1991. A theory of photometric stereo for a class of diffuse Non-Lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):133–152.

Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9 (2):137–154.

Woodham, R.J. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144.

Woodham, R.J., Iwahori, Y., and Barman, R.A. 1991. Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength. *University of British Columbia, TR-91-18*.

Wiberg, T. 1976. Computation of principal components when data are missing. In *Proc. Second Symposium of Computational Statistics*, pp. 229–236.

Yuille, A., Snow, D., Epstein, R., and Belhumeur, P. 1999. 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.

Zhang, L., Curless, B., Hertzmann, A., and Seitz, S.M. 2003. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. *IEEE International Conference on Computer Vision* pp. 618–625.

Zhang, R., Tsai, P.S., and Shah, M. 1996. Photomotion. *Computer Vision and Image Understanding*, 63(2):221–231.

Zhou, S.K., Chellappa, R., and Jacobs, D.W. 2004. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. *European Conference on Computer Vision*, Prague, I: 588–601.