

Photometric Stereo with General, Unknown Lighting

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

David Jacobs
NEC Research Institute
4 Independence Way
Princeton, NJ 08540

Abstract

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 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 isotropic and distant from the object but otherwise unconstrained. Lighting in each image may be an unknown and 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.

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 can be kept simple. In this paper 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. 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 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 set of images ([1, 20]). This set is 4D for a first order approximation, 9D for a second order approximation. This implies 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

*Research was supported in part by the European Commission Project IST-2000-26001 VIBES.

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 [23] 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 the images, 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 to first order, any linear transformation of an object's scaled normals will not change the 9D approximation to its images. Consequently, using the 9D approximation, an object's normals can only be stably reconstructed up to a linear transformation.

Finally, we present some preliminary experiments to 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 show experiments with real objects, in which we reconstruct shapes that appear to be quite veridical.

2 Background

Classical work on photometric stereo has assumed that the illumination conditions are known, e.g., through the specification of a reflectance function [24] (see [8] for a review). Much work has followed, including some that focuses on dealing with non-Lambertian objects (e.g., [5, 26]), in some cases capitalizing on distributed light sources ([9, 17]), and in situations in which the light source is near the object and near the camera ([13, 4]).

[21, 16] have pointed out that when each image is produced by a single point source with *unknown* intensity and direction one could recover 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. [25] also considers the problem of unknown light sources. [7] uses [21, 16]'s result in a factorization framework to handle many images. These results assume that the surface normals face the light source direction in all images (that is, no attached shadows). [3, 28, 6] (see also [18]) have shown that integrability re-

duces 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.

[14] extends 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. [15] used the same space to recover the albedo when the surface normals are known. [28] describes a reconstruction method when each image is lit by a single point source, and all images share a common background lighting, which can be arbitrary. 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.

Recently, [1, 20] have provided a new way to describe the effect of general lighting on a Lambertian object. They do this by focusing on the reflectance function, which describes the amount of light reflected by each surface normal. They point out that the reflectance function results from convolving a function describing the light with a kernel that describes Lambertian reflectance. This kernel is the max of the cosine function and zero, and can be shown to act as a low-pass filter. This means that only the low frequency components of lighting have a significant effect on a Lambertian object's reflectance function. 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. [1] proves that for *any* distant, isotropic lighting, at least 98% of the resulting 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.

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 [1, 20], namely, for a convex object illuminated by distant and isotropic lights that are otherwise unconstrained.

3 Shape Recovery

For our purposes, the key consequence of [1, 20] is that they provide an approximate, analytic description of an object's images as a 4 or 9D linear subspace of image space. To first order, any image of an object can be described as:

$\vec{i} = \vec{l}_4 \mathbf{S}_4$. \vec{i} is an n -dimensional vector containing the intensity of each pixel, and \vec{l}_4 is a 4D vector describing the low frequency components of lighting. \mathbf{S}_4 is a $4 \times n$ matrix whose rows each describe an image the object produces when lighting consists of a single, low-frequency, spherical harmonic. The first row, $\vec{\lambda}$, is a vector describing the albedo at each pixel. (We omit additional constant factors since they do not change the space spanned by the vectors.) The next three rows, $\vec{\lambda} \vec{n}_x$, $\vec{\lambda} \vec{n}_y$, $\vec{\lambda} \vec{n}_z$ are each a vector describing the x , y or z component of the object's surface normal at each point, scaled by its albedo. Abusing notation, we write a product of two vectors to denote *componentwise* multiplication of the elements of the two vectors. The four harmonic images obtained in this case are identical to the representation used in [14], 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 lighting, an image is described by $\vec{i} = \vec{l}_9 \mathbf{S}_9$, 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 rows are: $\vec{\lambda}(3\vec{n}_z\vec{n}_z - \vec{I})$, $\vec{\lambda}\vec{n}_x\vec{n}_y$, $\vec{\lambda}\vec{n}_x\vec{n}_z$, $\vec{\lambda}\vec{n}_y\vec{n}_z$, $\vec{\lambda}(\vec{n}_x\vec{n}_x - \vec{n}_y\vec{n}_y)$.

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 a linear combination of the harmonic images, that is,

$$M \approx LS,$$

where L ($f \times r$) contains the low order coefficients of the lighting and S ($r \times n$) contains the harmonic images. r is either four or nine. 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 [7] for the 3D linear space characterized in [21, 16] and by [14] for the 4D space that contains the zero and first order harmonics. Using SVD we obtain $M = U\Delta V^T$ 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:

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

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 .

Both the first and second order methods are based on the assumption that the low order harmonics will show up in the space produced 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 the same space as the first four principal components of the images. Then we 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 appear somewhere in the 9D space spanned by nine principal components. In this case we look for scaled normals that lie in the 9D space, and generate harmonic images that span a similar space.

Recently, [19] 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 this analysis to analyze the accuracy of each of our methods.

3.1 The Case of Four Harmonics

We have factored M to $M \approx \tilde{L}\tilde{S}$. This factorization is non-unique up to a linear transformation, A , since $\tilde{L}\tilde{S} = \tilde{L}A^{-1}A\tilde{S}$. So $S \approx A\tilde{S}$ for an unknown A . In the 4D case A is 4×4 . 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 $p = (p_1, p_2, p_3, p_4)^T$ 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 $p^T J p = 0$, 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 not true of \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. Denote $p = Aq$, where q is the corresponding column in \tilde{S} . Then $q^T A^T J A q = 0$. Denoting $B = A^T J A$, this constraint becomes $q^T B q = 0$. 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 $Qb = 0$. Q is $n \times 10$, and every row of Q corresponds to one column of \tilde{S} . So for a column q in \tilde{S} the corresponding row in Q is $(q_1^2, \dots, q_4^2, 2q_1q_2, \dots, 2q_3q_4)$. b is a 10-vector $b = (b_{11}, \dots, b_{44}, b_{12}, \dots, b_{34})^T$, 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 matrix $\tilde{B} = \beta A^T J A$ for some unknown scalar $\beta \neq 0$.

Next, we turn to factoring \tilde{B} to a product of the form

$\tilde{B} = \pm \tilde{A}^T J \tilde{A}$. Notice that if all the previous assumptions hold \tilde{B} should have one negative eigenvalue and three positive ones (or vice versa). To see this let $A = U_A \Delta_A V_A^T$ denote the singular value decomposition of A , then $\tilde{B} = \beta A^T J A = \beta V_A \Delta_A U_A^T J U_A \Delta_A V_A^T$. Since diagonal matrices commute, and $U_A^T U_A = I$ we obtain that $\tilde{B} = \beta V_A J \Delta_A^2 V_A^T$. Clearly, this is an eigenvalue decomposition of \tilde{B} with the eigenvalues given in the diagonal of $\beta J \Delta_A^2$, and so the first of them differs in sign from the remaining three. So we can factor \tilde{B} as follows. We apply an eigenvalue decomposition to \tilde{B} , $\tilde{B} = W \Lambda W^T$ where the columns of W contain the eigenvectors of \tilde{B} and Λ is a 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 $\tilde{B} = \tilde{A}^T J \tilde{A}$. 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 Frobenous norm $\|\pm \tilde{B} - \tilde{A}^T J \tilde{A}\|$. (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 $\tilde{A}\tilde{S}$. However, there is still an ambiguity because some linear transformations of $\tilde{A}\tilde{S}$ preserve the above constraints. Specifically, the factorization of \tilde{B} is not unique since we can obtain equally valid factorizations by multiplying \tilde{A} by any matrix \tilde{C} that satisfies $\tilde{C}^T J \tilde{C} = J$. Matrices that satisfy this condition represent all the projective transformations that keep the unit sphere fixed. 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 [12]).

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 [28].

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 twelve degrees of freedom in 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.

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 ([1]), 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. We 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:

$$\mathbf{A}\tilde{S} = \begin{pmatrix} \vec{\lambda}\vec{n}_x \\ \vec{\lambda}\vec{n}_y \\ \vec{\lambda}\vec{n}_z \end{pmatrix}.$$

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 [1], 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 an error, $E(\mathbf{A})$, as:

$$E(\mathbf{A}) = \min_L \|M - LS_A\|$$

where $\|\cdot\|$ denotes the Frobenous norm, and L is chosen so as to minimize the error. 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 apply any scaled, 3D rotation to the scaled surface normals, 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* ([3]). 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. This leads us to suspect that our approach can only accurately recover the surface normals up to a linear transformation. We now show this.

Suppose we alter the scaled surface normals. Since these are the second through fourth harmonics, we abbreviate: $h_2 = \vec{\lambda}\vec{n}_z$, $h_3 = \vec{\lambda}\vec{n}_x$, $h_4 = \vec{\lambda}\vec{n}_y$. Similarly, we let $h_1 = \vec{\lambda}$, $h_5 = \vec{\lambda}(3\vec{n}_z\vec{n}_z - \vec{I})$. Then, we consider applying a linear transformation:

$$\begin{pmatrix} h'_3 \\ h'_4 \\ h'_2 \end{pmatrix} = T \begin{pmatrix} h_3 \\ h_4 \\ h_2 \end{pmatrix},$$

where h'_i represents the harmonic images that results after the transformation. Using SVD, we can write T as R_1DR_2 , where R_1 and R_2 are rotations, and D is a diagonal matrix. Rotating the scaled surface normals h_2, h_3 and h_4 is just a coordinate change. This causes a phase shift; each harmonic 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 now show that to first order this has no effect on the first harmonic, the albedo.

We write $D = \text{diag}\{1 + d_3, 1 + d_4, 1 + d_2\}$. We analyze this to first order, taking the limit as $d_i \rightarrow 0$ and considering the effect on first order harmonics. To first order, the effect of scaling all of these components by different amounts is just the sum of the effects of scaling each one separately. So, without loss of generality, we can consider only the effect of scaling one, the z component. Then our problem reduces to determining how h_1 , the albedo, changes as we scale the second harmonic, h_2 , since scaling h_2 leaves the direction of h_2, h_3, h_4 unchanged. For notational simplicity, we consider the harmonic images at just one point.

We want to compute the following derivatives: $\frac{dh_1}{dh_2}$ and $\frac{dh_5}{dh_2}$. For this we use: $h_1 = \sqrt{h_2^2 + h_3^2 + h_4^2}$, and, using the

substitutions $\lambda = h_1$ and $z = \frac{h_2}{h_1}$:

$$h_5 = \frac{1}{2} \left(\frac{3h_2^2}{h_1} - h_1 \right)$$

These imply:

$$\frac{dh_1}{dh_2} = \frac{h_2}{h_1} = z,$$

and

$$\frac{dh_5}{dh_2} = \frac{\partial h_5}{\partial h_1} \frac{dh_1}{dh_2} + \frac{\partial h_5}{\partial h_2} = -\frac{3}{2}z^3 + \frac{5}{2}z$$

If we now change h_2 :

$$h'_2 = h_2(1 + \delta) = \lambda z(1 + \delta)$$

(so $h'_2 = h_2 + \Delta h_2$ with $\Delta h_2 = \lambda z\delta$). Then,

$$h'_1 \approx h_1 + \frac{dh_1}{dh_2} \Delta h_2 = \lambda(1 + z^2\delta),$$

$$h'_5 \approx h_5 + \frac{dh_5}{dh_2} \Delta h_2 = \frac{\lambda}{2}(3z^2 - 1 + (5 - 3z^2)z^2\delta).$$

Finally,

$$h'_1 - \frac{2}{3}\delta h'_5 = \lambda \left(1 + \frac{1}{3}\delta + O(\delta^2) \right).$$

This tells us that to first order, when we scale h_2 , the previous albedo, h_1 , lies in the linear space spanned by the new harmonic images. Therefore, all components of the images due to h_1, h_2, h_3, h_4 can also be produced by the new 9D linear subspace we get after any linear transformation near the identity.

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. 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.

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 , given as a function $f(x, y)$ 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 [3, 28]). 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 will 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.

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 approximately given by $n(x, y) = (p, q, -1)$, 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 respec-

tively. On a discrete grid, we may approximate p and q by

$$\begin{aligned} p &\approx z(x+1, y) - z(x, y) \\ q &\approx z(x, y+1) - z(x, y). \end{aligned}$$

Given a (recovered) scaled surface normal (n_x, n_y, n_z) , we know that roughly

$$\begin{aligned} p &= -\frac{n_x}{n_z} \\ q &= -\frac{n_y}{n_z}, \end{aligned}$$

so we obtain the following constraints:

$$\begin{aligned} n_z z(x, y) - n_x z(x+1, y) &= n_x \\ n_z z(x, y) - n_y z(x, y+1) &= n_y. \end{aligned}$$

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 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:

$$n_y(z(x, y) - z(x+1, y)) = n_x(z(x, y) - z(x, y+1)).$$

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 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. Other methods, such as gradient descent may lead to faster optimization, but our current 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

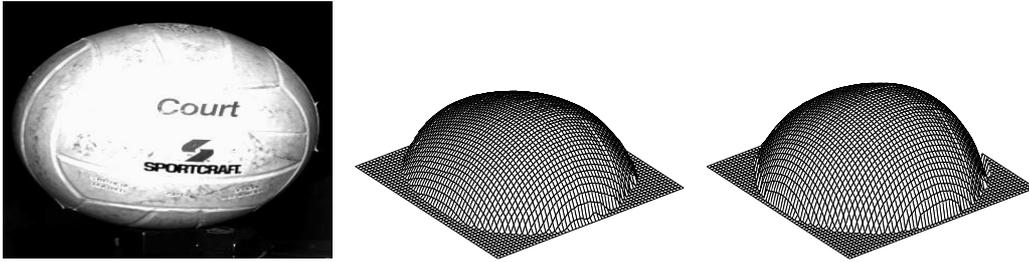


Figure 1: One of the volleyball images (left), the surface produced by the 4D method (center) and by the 9D method (right).

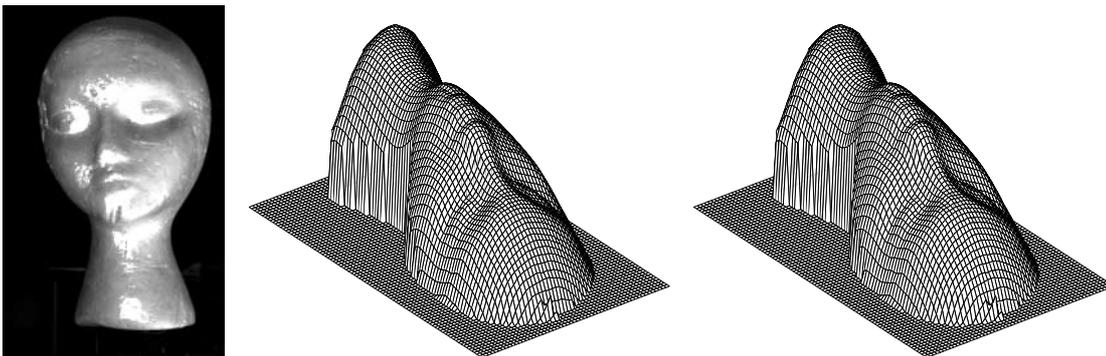


Figure 2: On the left, two face images averaged together to produce an image with two point sources. Saturated pixels shown in white. In the center, the surface produced by the 4D method. On the right, the surface from the 9D method.

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 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.

Finally, we have run our algorithms on two sets of real images. In all cases we subsampled the images considerably (resolution varied between 60×60 and 100×80). In some cases, these images contain unreliable pixels that have been saturated. We remove these, and use Wiberg's algorithm [27] to fill in the missing data (see also [10, 22]). In the volleyball images (Figure 1), we use 64 images taken by researchers at Yale. Each image is lit by a single point source. These controlled images could be used by other algorithms (eg., [28]), but our algorithms do not make any

assumptions that take advantage of the presence of a single source in each image. 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. Pixels are marked saturated if saturated in either image. Results are shown in Figure 2. In each case, we generate a surface by matching some points in the scene with hand chosen surface normals and using these to resolve any ambiguity of the method.

6. Summary and Conclusions

This paper describes new methods for recovering the surface normals in a scene using images produced under very general lighting conditions. The first insight that allows us to do this is that results due to [1, 20] 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 then show that with a 4D approximation, we can recover the normals by solving an overconstrained linear system. With a 9D approximation, we resort to a more general optimization method. We analyze the ambiguities that result from these linear methods. Finally, we show experimen-

tally, that these methods can produce accurate reconstructions under lighting conditions that are beyond the scope of previous algorithms.

References

- [1] R. Basri, D. Jacobs, "Lambertian reflectances and linear subspaces," *ICCV*, II:383–390, 2001.
- [2] P.N. Belhumeur, D.J. Kriegman. "What is the Set of Images of an Object Under All Possible Lighting Conditions?," *IJCV*, **28**(3):245-260, 1998.
- [3] P.N. Belhumeur, D.J. Kriegman, A.L. Yuille, "The Bas-Relief Ambiguity", *IJCV*, **35**(1):33–44,1999.
- [4] J.J. Clark, "Active Photometric Stereo," *CVPR*:29–34, 1992.
- [5] E.N. Coleman Jr., R. Jain, "Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," *CGIP*, **18**(4):309-328, 1982.
- [6] J. Fan, L.B. Wolff, "Surface Curvature and Shape Reconstruction from Unknown Multiple Illumination and Integrability," *CVIU* **65**(2):347–359, 1997.
- [7] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *JOSA*, **11**(11):3079–3089, 1994.
- [8] B.K.P. Horn, *Robot Vision*, MIT Press, 1986.
- [9] K. Ikeuchi, "Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method", *PAMI* **3**(6):661-669, 1981.
- [10] D. Jacobs, "Linear Fitting with Missing Data for Structure-from-Motion," *CVIU*, **82**(1):57–81, 2001.
- [11] D. Jacobs, P. Belhumeur, R. Basri. "Comparing Images Under Variable Illumination," *CVPR*:610–617, 1998.
- [12] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, 1993.
- [13] B. Kim, P. Burger, "Depth and Shape from Shading Using the Photometric Stereo Method," *CVGIP*, **54**(3):416–427. 1991.
- [14] J. Koenderink, A. van Doorn, "The Generic Bilinear Calibration-Estimation Problem," *IJCV*, **23**(3):217–234, 1997.
- [15] Q.T. Luong, P. Fua, Y.G. Leclerc, "The radiometry of multiple images," *PAMI*, forthcoming.
- [16] Y. Moses, *Face recognition: generalization to novel images*, Ph.D. Thesis, Weizmann Inst. of Science, 1993.
- [17] S.K. Nayar, K. Ikeuchi, T. Kanade, "Determining Shape and Reflectance of Hybrid Surfaces by Photometric Sampling," *IEEE Trans. on Robotics and Automation*, **6**(4):418–431, 1990.
- [18] R. Onn, A.M. Bruckstein, "Integrability Disambiguates Surface Recovery in Two-Image Photometric Stereo," *IJCV* **5**(1):105–113, 1990.
- [19] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability, including attached shadows, in a single image of a convex Lambertian object," personal communication.
- [20] R. Ramamoorthi, P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of convex Lambertian object." *JOSA A*, Oct 2001:2448–2459
- [21] A. Shashua, "On Photometric Issues in 3D Visual Recognition from a Single 2D Image", *International Journal of Computer Vision*, **21**(1-2):99–122, 1997.
- [22] H.Y. Shum, K. Ikeuchi, R. Reddy, "Principal Component Analysis with Missing Data and Its Application to Polyhedral Object Modeling," *PAMI*, **17**(9):854–867, 1995.
- [23] C. Tomasi, T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *IJCV*, **9**(2):137–154, 1992.
- [24] R.J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Optical Engineering*, **19**(1):139–144, 1980.
- [25] R.J. Woodham, Y. Iwahori, R.A. Barman, "Photometric Stereo: Lambertian Reflectance and Light Sources with Unknown Direction and Strength," *UBC TR-91-18*, August 1991.
- [26] H.D. Tagare, R.J.P. de Figueiredo, "A Theory of Photometric Stereo for a Class of Diffuse Non-Lambertian Surfaces," *PAMI* **13**(2):133–152, 1991.
- [27] T. Wiberg, "Computation of Principal Components when Data are Missing", *Proc. Second Symp. Computational Statistics*:229–236, 1976.
- [28] A. Yuille, D. Snow, R. Epstein, P. Belhumeur, "Determining Generative Models of Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability", *IJCV*, **35**(3):203–222, 1999.