Shape from Color

Mark S. Drew
School of Computing Science
Simon Fraser University
Vancouver, B.C.
Canada V5A 1S6
(604) 291-4682
e-mail mark@cs.sfu.ca

Tecnical Report CSS/LCCR TR 92-07

Abstract

We present a method for recovering the shape of a single Lambertian surface, with unknown but uniform reflectance, from a single RGB image of the object. The method depends upon the linearity of the relationship between RGB sensor responses and surface normal directions that stems from the Lambertian model. The algorithm follows closely upon work on the photometric stereo method. However, that method requires as input three separate black and white intensity images, each taken separately, whereas the method set out here needs only one color image. For the method to succeed, the lighting color must vary with direction from the surface. We test the method by including three separate point light sources all contributing to the illumination of the scene. The method correctly recovers surface normals, up to an overall orthogonal transformation.

1 Introduction

Recently, work by Petrov and others on the mathematical structure of color images, based on group theory, has become more widely known [6]. That work shows that, for a given color image, there is not an unlimited set of possibilities for the underlying surface normals that shape the shading field giving rise to the color and intensity of the image. In particular [6] proves that, for a single Lambertian surface illuminated by light that varies spectrally with direction from the surface, the possible solutions for the set of surface normals are nonunique and can all be characterized by a set of group orbits. Brill [1] showed, using a related mathematical analysis, that surface normals could be retrieved from a single color image once the nonuniqueness is fixed by knowing, a priori, the surface normals for three distinguished points.

This work would seem to indicate that the recovery of shape from a single color image is possible. This paper shows that this is indeed the case. However, in order to have a scheme that is independent of a priori knowledge it is necessary to somehow fix on a particular solution out of the infinite number possible. Recently, Woodham et al. [8] explored a very similar problem within the domain of photometric stereo [7]. They showed that, in a situation in which three separate black and white intensity images, taken separately, are used for input, the group structure again appears. Linearity can be used to show that each normalized surface normal, constrained to lie on a sphere, is transformed by the Lambertian model into a set of three numbers in color space that can be viewed as forming a vector with tip lying on the surface of an ellipsoid. A least squares fit to the 6 unknown parameters of the ellipsoid leaves 3 arbitrary constants for a solution to three sets of light source strength–direction vectors. These unspecified parameters are shown to correspond to the the three parameters of the rotation group $SO(3)$.

In that model, then, the three light sources are determined up to an overall orthogonal transformation. Here, we do not use three images taken with three light sources separately. As a result, we cannot in fact recover the light source directions. Instead, what is recovered is a 3 by 3 matrix characterizing the way in which the lighting color varies with direction as well as how the surface responds to the light. Color is crucial for this analysis. As a concomitant circumstance, the entire set of surface normals is also recovered, up to an orthogonal transformation. In section 2 we detail the relationship between a linear model and surface normals, and in section 3 apply the model to the case of a light source with directionally varying color illuminating a Lambertian surface. In section 4 we test the model using a set of three point sources illuminating a colored object.

2 Linear Shading Models

In [8], Woodham et al. outline the familiar situation of 3–image photometric stereo, in which three intensity images are taken of an object. For a Lambertian surface with uniform albedo, distant lighting, and distant viewing, the image irradiance can be written simply as

$$E(x, y) = \alpha \cos(\theta) , \tag{1}$$

where α combines both the surface albedo and the source strength, and the term $\cos(\theta)$ is the shading field. Suppose we call the surface normal at any point $\boldsymbol{x}$. Then if we let $\boldsymbol{a}$ be the direction to the light source multiplied by α, $E(x, y)$ is simply $\boldsymbol{a}^T \boldsymbol{x}$. Call this number y.

Now suppose we take three separate images of the object, each one using a different light source direction and strength. Then $\boldsymbol{a}$ is different for each image. If we group the resulting values y_1, y_2, y_3 in a kind of "vector"

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \tag{2}$$

(although these components do not form a vector in the sense of three components in a coordinate system) then the three measured images can be summarized by writing

$$\boldsymbol{y} = \boldsymbol{A} \, \boldsymbol{x} , \tag{3}$$

where $\boldsymbol{A}$ is formed by stacking the three $\boldsymbol{a}$ vectors row–wise:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \boldsymbol{a}_3 \end{pmatrix} . \tag{4}$$

Now denote the inverse of $\boldsymbol{A}$ by $\boldsymbol{B}$, so that

$$\boldsymbol{x} = \boldsymbol{B} \, \boldsymbol{y} . \tag{5}$$

Any normal vector must lie on the sphere $\boldsymbol{x}^T \boldsymbol{x} = 1$; substituting, we have

$$\boldsymbol{y}^T \boldsymbol{B}^T \boldsymbol{B} \, \boldsymbol{y} \equiv \boldsymbol{y}^T \boldsymbol{C} \, \boldsymbol{y} = 1 . \tag{6}$$

Since $\boldsymbol{C}$ is $\boldsymbol{B}^T \boldsymbol{B}$, it is 3×3 and symmetric positive definite. Therefore it consists of 6 parameters.

The innovative work of Woodham et al. [8] recognizes that the quadric (6) therefore comprises an ellipsoid, fixed by 6 parameters. The three images take surface normals into a 3-tuple lying on an ellipsoid

$$c_{11} y_1^2 + c_{22} y_2^2 + c_{33} y_3^2 + 2c_{12} y_1 y_2 + 2c_{13} y_1 y_3 + 2c_{23} y_2 y_3 = 1 . \tag{7}$$

To approximate the matrix entries c_{ij}, then, one can simply look for the best least squares fit to the measured y data resulting from the three images. However, $\boldsymbol{C}$ has only 6 unique entries, so

$\boldsymbol{A}$ cannot be uniquely determined from such a fit. Nonetheless, $\boldsymbol{A}$ must be determined up to a rotation, or more generally an orthogonal transformation if reflections are considered, because (6) is invariant with respect to this group.

The least squares fit follows by setting up an auxiliary matrix $\boldsymbol{M}$ with columns given by all measurements of y values that are coefficients of the c_{ij} in eqn. (7). Suppose we sample the image at N pixels. Then $\boldsymbol{M}$ has columns

$$\left(y_1^2\,,y_2^2\,,y_3^2\,,2y_1y_2\,,2y_1y_3\,,2y_2y_3\right)\,,$$

where now we mean $N \times 1$ tuples y_i. Thus $\boldsymbol{M}$ is $N \times 6$. The best least squares answer for the solution $\boldsymbol{z} = (c_{11}\,,c_{22}\,,c_{33}\,,c_{12}\,,c_{13}\,,c_{23})^T$ is determined (see, e.g., [5]) by the pseudoinverse $\boldsymbol{M}^+$. Now $\boldsymbol{A}$ is determined up to a rotation since, letting $\boldsymbol{D} = \boldsymbol{C}^{-1}$, one finds

$$\boldsymbol{D} = \boldsymbol{C}^{-1} = \boldsymbol{A}\,\boldsymbol{A}^T\,. \tag{8}$$

Therefore any assignment scheme for $\boldsymbol{A}$ that adheres to (8) will do, and $\boldsymbol{A}$ cannot be determined absolutely without injection of further knowledge.

The scheme chosen in [8] is as follows:

1. Set $a_{11} = \sqrt{d_{11}}$, $a_{12} = a_{13} = 0$.
2. Set $a_{21} = d_{21}/a_{11}$, $a_{22} = \pm\sqrt{d_{22} - {a_{21}}^2}$, $a_{23} = 0$. The $\pm$ sign is chosen to make $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$ a right-handed system.
3. Set $a_{31} = d_{31}/a_{11}$, $a_{32} = (d_{32} - a_{21}a_{31})/a_{22}$, $a_{33} = \pm\sqrt{d_{33} - {a_{31}}^2 - {a_{32}}^2}$, .

This generates a lower-triangular matrix $\boldsymbol{A}$. Thus for any measured $\boldsymbol{y}$ one can generate an approximation to the underlying $\boldsymbol{x}$.

In the next section, we show how this formalism may be adopted to the case of a single color image.

3 Color Shading Models

There are three crucial features that render the above analysis solvable: the linearity of eqn. (3), and the invertibility and $\boldsymbol{x}$ independence of $\boldsymbol{A}$. Clearly, if both these conditions can be met within some other paradigm than photometric stereo the same least squares algorithm can be used to arrive at an approximation to shape.

It is possible to meet the linearity condition by considering the vector $\boldsymbol{\rho}$ formed from sensor responses in an RGB color space [6]. For within a color Lambertian model shading is given exactly as in eqn. (1) except that the albedo is replaced by an RGB vector $\boldsymbol{b}$ giving the color of the illuminant reflected by the surface. Invertibility is lost in this model, however, if a single light source spectrum is used, because vectors $\boldsymbol{\rho}$ from each pixel are parallel.

To correct this deficiency, consider a situation in which a uniformly colored object is illuminanted by three lights, each having a different spectral power distribution. Barring accidental metamers giving parallel $\boldsymbol{\rho}$ vectors, this situation produces a set of vectors $\boldsymbol{\rho}$ that are formed in a linear way from surface normals and yet possess the invertibility property.

To see this, suppose the color and strength of reflected light from the first light is $\boldsymbol{b}_1$. Here, the color of the light and the reflectance of the surface are inextricably confounded unless some color constancy scheme is called into play (cf. [2, 4]). Suppose now that there are three point source lights present simultaneously, with unknown normalized direction vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$. Let the colors generated from the surface be, respectively, $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$. These vectors $\boldsymbol{b}$ describe both RGB direction as well as the strength of the source–surface reflected color signal. Figure 1 shows the situation. Here we have a single spherical Lambertian surface, illuminated simultaneously by lights of three different colors and strengths from three different directions. Figure 1 (a–c) show the effects of the three different lights separately, and Figure 1(d) is the sum of these. Because the particular colors are not critical these images are shown in black and white. Nevertheless, all four are color images; Figure 1(a–c) are each single color but it is important to note that Figure 1(d) varies in color over the surface of the object.

The resulting camera response from a pixel corresponding to normalized surface normal $\boldsymbol{x}$ is

$$\boldsymbol{\rho} = (\boldsymbol{a}_1^T \boldsymbol{x})\, \boldsymbol{b}_1 + (\boldsymbol{a}_2^T \boldsymbol{x})\, \boldsymbol{b}_2 + (\boldsymbol{a}_3^T \boldsymbol{x})\, \boldsymbol{b}_3 \,. \tag{9}$$

Now combining RGB vectors $\boldsymbol{b}_i$ into a matrix $\boldsymbol{B}$

$$\boldsymbol{B} = (\boldsymbol{b}_1,\ \boldsymbol{b}_2,\ \boldsymbol{b}_3) \;, \tag{10}$$

one may rewrite eqn. (9) as

$$\boldsymbol{\rho} = \boldsymbol{F}\, \boldsymbol{x} \equiv \boldsymbol{B}\, \boldsymbol{A}\, \boldsymbol{x} \;. \tag{11}$$

The situation of three lights of different spectral characteristics coming from three different directions will certainly produce a 3×3 matrix $\boldsymbol{F}$ that is independent of $\boldsymbol{x}$ and invertible. Nonetheless, the linear invertible color model (11) is not limited to this situation. However, in order for invertibility to be maintained, the matrix $\boldsymbol{B}\,\boldsymbol{A}$ capturing light source RGB and spatial directional properties must characterize an extended source or set of sources that produces variations in color for all or at least a good many of the normal vectors in the scene. As well, linearity of the model must be maintained. This means [1] that $\boldsymbol{\rho}$ must arise from a cosine model. For the model (11), invertibility of $\boldsymbol{F}$ is ensured if it is not rank-reduced. This is turn implies that neither $\boldsymbol{A}$ nor $\boldsymbol{B}$ has rank less than 3. For matrix $\boldsymbol{A}$, this means the light directions must be linearly independent. Similarly, matrix $\boldsymbol{B}$ must correspond to colors in RGB space that are not coplanar or coincident. These two conditions guarantee directionally varying color over the object surface.

Even if there is some degeneracy of the data the least squares solution method will still find the best approximate solution. As well, even if the lighting produces severely degenerate data it may be possible to employ an inversion method based on the Singular Value Decomposition to find a solution [3].

The particular model (11) satisfies the requirement in Petrov's analysis [6] that the spectrum be directionally varying. Petrov's work, and that of Brill [1], show that surface normals should in principle be recoverable from a single RGB image. The least squares method can be used to select a particular solution out of all those available. To use the method, exactly the same equations, (6–8), can be used, but with $\boldsymbol{y}$ replaced by $\boldsymbol{\rho}$. However, instead of recovering the matrix of light directions $\boldsymbol{A}$, the algorithm will recover an approximation of the product matrix $\boldsymbol{F}$.

This means that the present method is more limited than the three–image method of [8] in that light direction vectors and light strengths are not recovered. Because color and coordinate space components are mixed in $\boldsymbol{F}$ the light source vectors are unknown unless the surface is gray or its color is known.

On the other hand, the present method does succeed in recovering surface normals and thus shape. For once an approximation for $\boldsymbol{F}$ is known one can invert (11) to obtain the surface normal $\boldsymbol{x}$ from the measured RGB values:

$$\boldsymbol{x} = \boldsymbol{F}^{-1} \boldsymbol{\rho} . \tag{12}$$

Of course, since one recovers $\boldsymbol{F}$ only up to an overall orthogonal transformation the orientation of the surface remains unknown without further knowledge.

To study how well the model performs, in the next section we consider the simple surface of Fig. 1, illuminated by three lights.

4 Simulation Test

A single color image of the colored sphere of Fig. 1, illuminated by three distant point sources with distinct colors, yields an image in each RGB channel. For the sphere of Fig. 1, we used normalized light source directions

$$\boldsymbol{a}_1 = (0.5567, 0.2408, 0.7950), \boldsymbol{a}_2 = (-0.5567, 0.2408, 0.7950), \boldsymbol{a}_3 = (0.0, 0.0, 0.1). \tag{13}$$

The light source color combines with the surface color in the vectors $\boldsymbol{b}$, one for each light source. These vectors also include the strength of the light source. For Fig. 1, we chose

$$\boldsymbol{b}_1 = (1.7321, 1.7321, 1.7321), \boldsymbol{b}_2 = (0.4364, 1.7457, 0.8729), \boldsymbol{b}_3 = (0.5571, 0.8356, 1.1142). \tag{14}$$

Using the resulting RGB values $\boldsymbol{\rho}$ in place of values $\boldsymbol{y}$ in eqn. (7), we determine the best least squares solution for the matrix $\boldsymbol{C}$ in

$$\boldsymbol{\rho}^T \boldsymbol{C}^T \boldsymbol{C} \boldsymbol{\rho} = 1 . \tag{15}$$

Using the scheme in section 2 for pinning down a particular matrix $\boldsymbol{A}$, we recover from the image an approximation of the color–direction matrix $\boldsymbol{F}$. Let that approximation be denoted $\boldsymbol{G}$.

Matrix $\boldsymbol{G}$ is the best approximation of the matrix that takes a surface normal into a color space vector $\boldsymbol{\rho}$. To invert the transformation we need $\boldsymbol{G}^{-1}$. Here we find

$$\boldsymbol{G}^{-1} = \begin{pmatrix} 0.4063 & 0.0 & 0.0 \\ -1.4097 & 0.9659 & 0.0 \\ -6.3508 & -4.2943 & 9.4816 \end{pmatrix} \tag{16}$$

Without any further knowledge of the arbitrary rotation left unfixed by the method, surface normals are recovered from the above matrix and from the image via

$$\boldsymbol{x}_{rot} \simeq \boldsymbol{G}^{-1}\boldsymbol{\rho} \ , \tag{17}$$

where $\boldsymbol{x}_{rot}$ includes a subscript indicating that it is correct only up to a rotation.

Since here in fact we know the correct matrix $\boldsymbol{F}$ we can substitute for $\boldsymbol{\rho}$ and determine the rotation $\boldsymbol{R}$ by which $\boldsymbol{G}$ is rotated from $\boldsymbol{F}$. In general this would not be possible. Here, we find

$$\boldsymbol{x}_{rot} = \boldsymbol{G}^{-1}\boldsymbol{\rho} = \boldsymbol{G}^{-1}\boldsymbol{F}\boldsymbol{x} \equiv \boldsymbol{R}\boldsymbol{x} \ , \tag{18}$$

and thus

$$\boldsymbol{x} = \boldsymbol{R}^{-1}\boldsymbol{x}_{rot} = \boldsymbol{R}^{T}\boldsymbol{x}_{rot} \ . \tag{19}$$

The matrix $\boldsymbol{R}$ comes out to be

$$\boldsymbol{R} = \begin{pmatrix} 0.2931 & 0.2122 & 0.9269 \\ -1.0241 & 0.0728 & 0.2621 \\ -0.0128 & -0.9651 & 0.2518 \end{pmatrix} . \tag{20}$$

This matrix is only approximately orthogonal. It is close, however, to a rotation about the axis (0.6376, 0.4708, 0.6098) by an angle 101.02°.

To see how well eqn. (17) performs we first recover $\boldsymbol{x}_{rot}$ and then compare to the correct $\boldsymbol{x}$ values using eqn. (19). It is meaningful to compare the correct $\boldsymbol{x}$ with the recovered one by considering the angle between the two. For all pixels, we find that this error angle has a mean of 6.47°, a standard deviation of 11.39°, and a median value 3.20°. Figure 2 shows a histogram of the error results. In the main, the algorithm is seen to perform very well. Figure 3 depicts the original and recovered surface normals. They are seen to be virtually indistinguishable.

5 Conclusions

We have adapted previous work on shape recovery using the photometric stereo method to the case of a single color image. The fact that such a shape–from–color algorithm is possible was adumbrated by previous work on the group structure of color images. The present method can recover surface normals provided the input image satisfies the two constraints of linearity and invertibility. Shape descriptors are recovered only up to an unknown orthogonal transformation.

Because the method is based on a least squares approximation it will be robust with respect to noise. There may however be another source of error if the linearity condition is violated by shadowing. Consider once again the three–light situation. Illumination from one or more of the light sources may not reach some pixels in the image. When that is the case the linear model of eqn. (11) no longer holds because the shading with respect to some of the lights is zero. I.e., the measured $\boldsymbol{\rho}$ omits one of the lights when the angle to that light exceeds $90°$. To see how much this problem affected the simulation we carried out the least squares analysis using only non-shadowed pixels. The results were very close to those shown above.

6 Acknowledgements

This work was supported in part by a grant from the Centre for Systems Science of Simon Fraser University.

References

[1] M.H. Brill. Photometric models in multispectral machine vision. In B.E. Rogowitz, M.H. Brill, and J.P. Allebach, editors, *Human Vision, Visual Processing and Digital Display II*, volume 1453, pages 369–380. SPIE, 1991.

[2] M.S. Drew and B.V. Funt. Natural metamers. *CVGIP:Image Understanding*, 1992. In press.

[3] G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins U. Press, 1983.

[4] J. Ho, B.V. Funt, and M.S. Drew. Separating a color signal into illumination and surface reflectance components: Theory and applications. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 12:966–977, 1990.

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

[6] A.P. Petrov. Color and Grrassman–Cayley coordinates of shape. In B.E. Rogowitz, M.H. Brill, and J.P. Allebach, editors, *Human Vision, Visual Processing and Digital Display II*, volume 1453, pages 342–352. SPIE, 1991.

[7] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.

[8] R.J. Woodham, Y. Iwahori, and R.A. Barman. Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength. Technical Report TR 91-18, University of British Columbia Department of Computing Science, 1991.

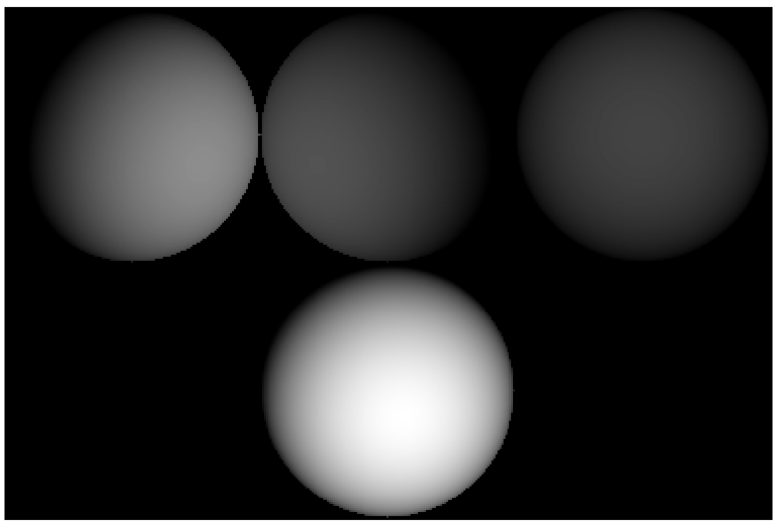

Figure 1: A colored sphere illuminated by three differently colored lights, each in a different direction.
(a–c) The image contribution from each light separately. Each image has a single color.
(d) The sum of (a–c). This image has varying color over the surface of the object.

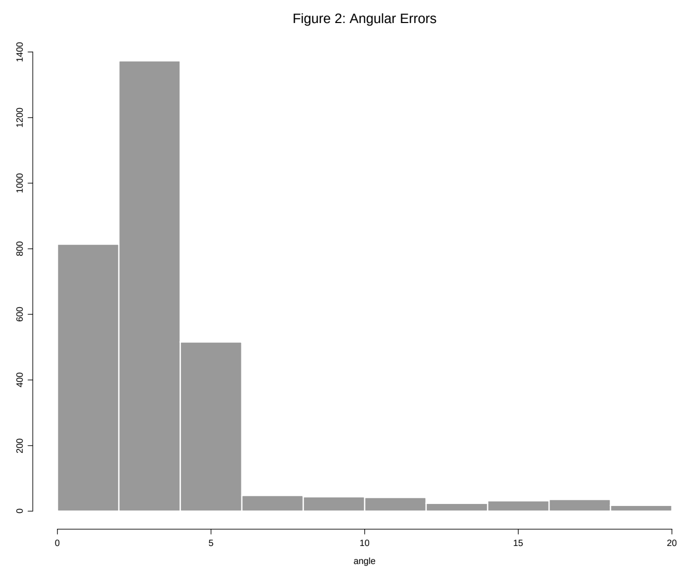

Figure 2: Histogram of the angular errors between the actual and recovered surface normals.

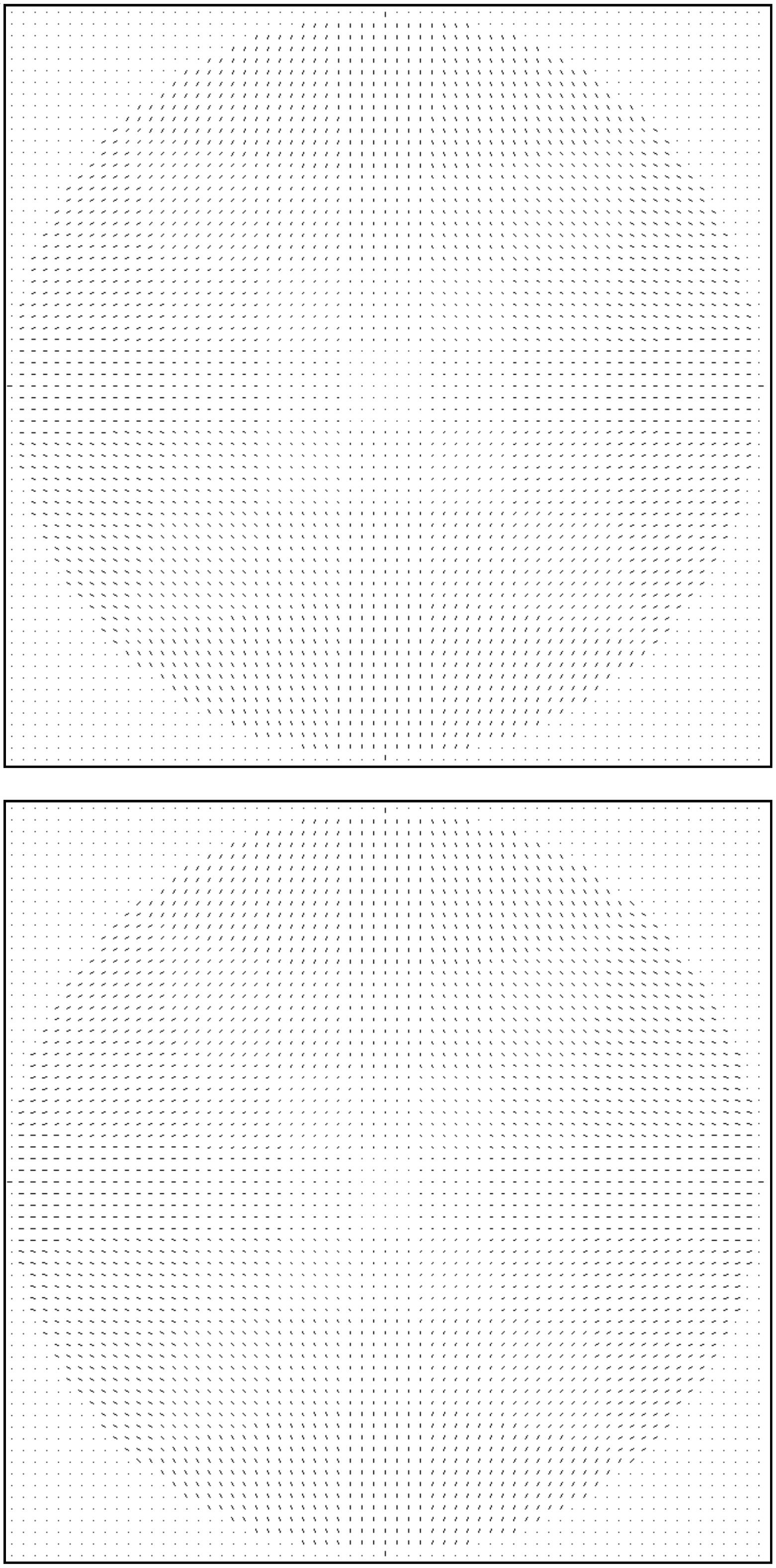

Figure 3: Surface normal vectors for (a) input image and (b) normals recovered by the algorithm.