A Two-Frame Theory of Motion, Lighting and Shape

Ronen Basri Darya Frolova
Dept. of Computer Science and Applied Math
Weizmann Institute of Science
Rehovot, 76100 Israel

Abstract

This paper explores how shape, motion, and lighting interact in the case of a two-frame motion sequence. We consider a rigid object with Lambertian reflectance properties undergoing small motion with respect to both a camera and a stationary point light source. Assuming orthographic projection, we derive a single, first order quasilinear partial differential equation that relates shape, motion, and lighting, while eliminating out the albedo. We show how this equation can be solved, when the motion and lighting parameters are known, to produce a 3D reconstruction of the object. A solution is obtained using the method of characteristics and can be refined by adding regularization. We further show that both smooth bounding contours as well as surface markings can be used to derive Dirichlet boundary conditions. Experimental results demonstrate the quality of this reconstruction.

1. Introduction

Moving objects change their appearance in complex ways: object points may change both their image location as well as their intensity values from one video frame to the next, owing to a respective change in both object location relative to the camera as well as its surface orientation relative to a light source. A fundamental task in computer vision, therefore, is to provide derivations to describe how motion and lighting interact with shape to produce a motion sequence. This can potentially lead to novel reconstruction and matching algorithms that make full utilization of image information. This paper introduces such a derivation for the case of a two-frame motion sequence of a rigid object with lambertian reflectance properties.

To illustrate the interplay between motion and shading information, consider the case of a lambertian sphere with smoothly varying albedo that undergoes pure rotation. In this case motion cues extracted with the brightness constancy assumption would generally be incorrect. In particular, if the sphere has constant albedo no motion would be observed at all. At the same time, unless the albedo is known, reconstruction with shape from shading methods would not be feasible. Photometric stereo methods too are irrelevant both because lighting is stationary and because the object is moving. Nevertheless, if the lighting and motion parameters are known it should be possible to reconstruct the sphere by combining motion and lighting cues. Our formulation below achieves exactly that.

This paper addresses the case of a rigid object with Lambertian reflectance properties that undergoes small motion with respect to both the camera and a light source. We further assume the object is illuminated by a single, stationary distant point source and is observed by a stationary orthographic camera. Under such conditions, we derive expressions that relate the brightness values in two frames, and by eliminating albedo we arrive at a single, first order quasilinear partial differential equation (PDE) that relates shape, lighting, and motion. We then discuss solution methods by which the shape of the object can be recovered when lighting and motion parameters are supplied. We further explore ways to construct boundary conditions, particularly in the case of objects with smooth bounding surfaces. Experimental results demonstrate the utility of our formulation.

The paper is divided as follows. Section 2 discusses previous work. Section 3 derives the main PDE and discusses solution methods and boundary conditions. Experimental results are shown in Section 4.

2. Previous work

A large number of studies in vision utilize either shading or motion cues for reconstruction, but only few consider both cues simultaneously. Shading cues serve for shape recovery from either a single image ("Shape from shading" [6]) or from multiple images of a static scene under mere changes in lighting ("photometric stereo" [26]). In contrast, motion algorithms frequently utilize the brightness constancy assumption [7, 11] or use sparse feature

The vision group at the Weizmann Institute is supported in part by the Moross Foundation. R.B. is currently with the Toyota Technological Institute at Chicago

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

points [14, 24, 23], which are somewhat robust to differences in lighting.

Some recent methods attempt to combine shading and motion in various ways. Shading information has been introduced into stereo reconstruction [3, 4] to handle static scenes where brightness constancy (or a slightly more relaxed constraint) still applies. In this work shape from shading algorithms applied to any of the stereo pair are used to recover depth in regions between feature points. Generalizations of the brightness constancy assumption that locally model differences due to lighting were proposed in [1, 18, 5]. Finally, a number of recent reconstruction methods account for both motion and lighting cues, but in a multiframe setting [2, 9, 15, 16, 17, 22, 25, 27, 10]. For example, [22, 10] use at least four frames to determine for each hypothesized correspondence whether or not it is consistent with some lighting.

Our approach is related to a recent method proposed in [16], which also combines motion and intensity cues to recover shape. (A similar approach was suggested by the same authors in the context of bilateral symmetric objects in [21]). This approach, given three images of a moving object, uses a propagation process to determine shape. In this process a point correspondence across the three frames determines the surface normal at a point, and the surface normal in turn determines the next correspondence. The setting is general, and can be used under large motion, perspective projection, and changes in lighting between the frames. However, it requires at least three images (or uniform albedo). Our method, in contrast, applies in a more restricted setting (small motion, orthographic projection, and stationary lighting), but it requires only two views, and perhaps most significantly, it provides a succinct mathematical formulation of the problem in the form of a *single* differential equation. This clarifies the structure and numerical properties of the problem, suggests effective solution methods, and provides new ways to derive boundary conditions.

3. Two-Frame Analysis

We consider a rigid object with Lambertian reflectance properties undergoing small motion with respect to both the camera and a stationary light source. In what follows we derive expressions that under these conditions relate motion, lighting, and shape in two frames. We further discuss solution methods to recover the shape of the object when both motion and lighting parameters are given. Note that knowledge of the parameters of the rigid motion is not sufficient to determine the shape of the object, since the magnitude of motion at every point depends on the depth of that point. Our method therefore uses both lighting and motion cues to determine the magnitude of motion, and hence the shape of the object.

3.1. Preliminaries

Given an image pair $I(x,y)$ and $J(x,y)$ of a rigid, lambertian object we begin by associating a coordinate frame with the images so that they differ by a mere rotation about the Y-axis. This is achieved in a process similar to that presented in [20]. Assume the motion of the object is known, under the orthographic (or weak perspective) projection a point $P \in \Re^3$ whose projection onto I is given by $\mathbf{p}$ would be found at $\mathbf{q} = s\tilde{R}P + \mathbf{t}$ in J, where $s > 0$ is a uniform scaling factor, $\tilde{R}$ is a 2×3 matrix containing the top two rows of a 3D rotation matrix R, and $\mathbf{t}$ is a 2-vector of translation. We define our coordinate frame by first translating J by $-\mathbf{t}$. (Alternatively, we can eliminate translation by identifying a pair of corresponding points $\mathbf{p}_0$ and $\mathbf{q}_0$ with the origin of I and J respectively.) Next we eliminate s by scaling J by a factor of $1/s$. Finally, we rotate both images so as to make their epipolar lines parallel to the horizontal axis. This can be done by factoring the rotation R into a product of three consecutive axial rotations, about the Z, Y, and Z axes, i.e., $R = R_z(\psi)R_y(\theta)R_z(\phi)$. Rotating I by an angle ϕ and J by an angle $-\psi$ rectifies the epipolar lines. Following this operation J is related to I by a mere 3D rotation of angle θ about the Y-axis:

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}.$$

Since we assume this rotation to be small we use below a first order approximation with $\sin\theta \approx \theta$ and $\cos\theta \approx 1$.

Note that this change of coordinates merely serves to simplify notations. Similar expressions to the ones introduced below can be derived also for non-rectified images.

3.2. Main derivation

Denote the object's surface (in the coordinate frame of I) by $z(x,y)$. The surface normal at each point is expressed by $\mathbf{n}(x,y) = 1/\sqrt{z_x^2 + z_y^2 + 1}(-z_x, -z_y, 1)^T$, where $z_x = \partial z/\partial x$ and $z_y = \partial z/\partial y$. We assume the object is illuminated by a single point source at infinity ("directional source") and express the direction of the light source by a vector $\mathbf{l} = (l_1, l_2, l_3)^T$. The image intensities determined by the Lambertian law are given by $I = \rho \mathbf{l}^T \mathbf{n}$, written explicitly as:

$$I(x,y) = \frac{\rho}{\sqrt{z_x^2 + z_y^2 + 1}} \left(-l_1 z_x - l_2 z_y + l_3\right), \quad (1)$$

where we use $\rho(x,y)$ to denote the surface albedo scaled by light intensity.

Following a rotation about the Y-axis, a point (x,y) in I appears at a new location (x',y) in J, and its surface normal rotates as well. The intensities in J can then be expressed as a dot product of the light with the rotated normal, or equivalently as a dot product of the original normals

with the light source rotated in the opposite direction, i.e., $J(x',y) = \rho \mathbf{l}^T(R_y(\theta)\mathbf{n}) = \rho(R_y^T(\theta)\mathbf{l})^T\mathbf{n}$. Using the first order approximation to $R_y(\theta)$ and dropping any higher order terms this is written explicitly as:

$$J(x',y) = \frac{\rho}{\sqrt{z_x^2+z_y^2+1}}\left(-(l_1+l_3\theta)z_x - l_2 z_y - l_1\theta + l_3\right). \tag{2}$$

Subtracting (1) from (2) we obtain

$$J(x',y) - I(x,y) = -\frac{\rho\theta}{\sqrt{z_x^2+z_y^2+1}}\left(l_3 z_x + l_1\right). \tag{3}$$

Next, by combining (3) with (1) we can eliminate both the albedo and the norm terms, obtaining a relation between shape, motion and lighting

$$(-l_1 z_x - l_2 z_y + l_3)(J(x',y) - I(x,y)) = -(l_3 z_x + l_1)\theta I(x,y). \tag{4}$$

The relation above is still unsatisfactory since I and J are evaluated at different points, requiring a solution to the correspondence problem, which itself depends on the surface $z(x,y)$. Notice that due to the first order approximation to $R_y(\theta)$ we have that $x' \approx x - \theta z$. Therefore, in the case of small motion we can replace $J(x',y)$ with a Taylor series:

$$J(x-\theta z, y) = \sum_{k=0}^{\infty} \frac{1}{k!}\frac{\partial^k J(x,y)}{\partial x^k}(-\theta z)^k. \tag{5}$$

In this paper we use a first order Taylor expansion: $J(x',y) \approx J(x,y) - \theta z J_x(x,y)$ (as is commonly done also with constant brightness [7, 11]), but the same derivations can be provided with any number of terms of 5. A higher order Taylor expansion can be necessary, for example, to handle high frequencies in albedo.

This approximation will be accurate when the difference $|x'-x| \approx |\theta z|$ is small, which is satisfied in particular when both the rotation angle θ and the depth value z are small, or when θ is small and z is bounded. Notice that under the orthographic projection $z = 0$ is the depth of the centroid of the object, and so we can expect z to be small under the usual orthographic assumptions, i.e., whenever the object is far from the camera relative to its size.

Plugging this into (4) we obtain

$$(-l_1 z_x - l_2 z_y + l_3)(J(x,y) - z\theta J_x(x,y) - I(x,y)) = -(l_3 z_x + l_1)\theta I(x,y). \tag{6}$$

This is our final relation, which can be written succinctly as a PDE in $z(x,y)$ with coefficients that depend on depth, lighting, and motion

$$a z_x + b z_y = c, \tag{7}$$

where

$$\begin{aligned} a(x,y,z) &= l_1(I_\theta - z J_x) - l_3 I \\ b(x,y,z) &= l_2(I_\theta - z J_x) \\ c(x,y,z) &= -l_3(I_\theta - z J_x) - l_1 I, \end{aligned}$$

and $I_\theta = (J - I)/\theta$.

3.3. Solution by Characteristics Curves

Eq. (7) is a first order partial differential equation in $z(x,y)$. Such an equation is called *quasilinear* since it is linear in the derivatives of z, while its coefficients depend on z. A useful interpretation of this equation can be obtained [28] by writing (7) in the form of a scalar product, $(a,b,c)(-z_x,-z_y,1)^T = 0$. This form implies that at every point the vector of coefficients is perpendicular to the surface normal, and so it must lie in the tangent plane to the surface $z(x,y)$. Consequently, the field of coefficient vectors traces curves on the sought surface. These are the characteristic curves, which in the case of a quasilinear equation lie in 3D. The characteristic curves are defined by $(dx,dy,dz)^T \times (a,b,c)^T = 0$, where $(dx,dy,dz)^T$ is an arbitrary infinitesimal vector parallel to the coefficient vector, and can be constructed by solving the following system of ordinary differential equations $\frac{dx}{ds} = a$, $\frac{dy}{ds} = b$, $\frac{dz}{ds} = c$, where s is a dummy variable representing parametrization along the characteristics.

While the characteristic curves lie in 3D, in the case of (7) each characteristic traces a planar curve on the surface, and this family of characteristic curves lie in parallel planes. This can be verified by checking that

$$l_1 l_2 a - (l_1^2 + l_3^2)b - l_3 l_2 c = 0 \tag{8}$$

for all (x,y,z), implying that all vectors of coefficients $(a,b,c)^T$ are perpendicular also to the constant vector $(l_1 l_2, -(l_1^2+l_3^2), -l_3 l_2)^T$. Note that in general the family of planes traced by the characteristics will not coincide with the epipolar planes, unless $l_2 = 0$.

With the appropriate Dirichlet boundary conditions Eq. (7) can be solved using the method of characteristics. In particular, let (x_0,y_0) denote a point on the boundary whose (known) depth value is given by $z_0(x_0,y_0)$. Then, using (x_0,y_0,z_0), the coefficients (a_0,b_0,c_0) at (x_0,y_0) can be computed. Moving from the boundaries slightly in direction (a_0,b_0,c_0), i.e., to $(x_1,y_1,z_1) = (x_0,y_0,z_0) + \gamma(a_0,b_0,c_0)$, where $\gamma > 0$ is a small constant, brings one to a new point that, to a first order approximation, lies on the surface. At this new point new coefficients, (a_1,b_1,c_1), can be computed. This process can be repeated for each characteristic until the entire surface is reconstructed. Due to (8), intersections of different characteristics are not common.

Solutions with the method of characteristics have the disadvantage that each characteristic is solved independently, leading to greater sensitivity to errors. One way to address this problem is by introducing regularization. In our experiments below we show reconstructions obtained with the method of characteristic without using additional regularization.

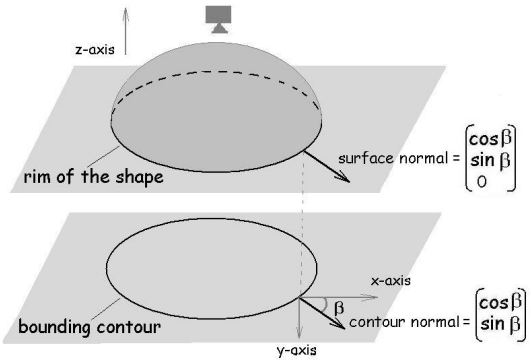

Figure 1. *The figure shows the projection of the rim of the shape onto the image plane. Normal vectors at the rim points have the same directions as normals vectors to the bounding contour in the image plane*

3.4. Boundary conditions

Solutions with characteristics require Dirichlet boundary conditions, with an initial depth value $z(x, y)$ specified at some point along each characteristic. One way to obtain depth values for boundary conditions is by detecting distinctive surface markings. Such markings will generally be robust to changes in illumination, and so their locations in the image can be identified across frames. But even if the object lacks sufficiently many markings, depth values can also be obtained along the bounding contours of a smooth shape by further exploiting the particular form of (7).

For smooth objects the bounding contour is the projection of the *rim* of the shape (Figure 1). Under the orthographic projection the surface normal at the rim is parallel to the image plane and its orientation coincides with the normal to the projected bounding contour. Using a spherical coordinate representation for the surface normal: $\mathbf{n} = (\cos\beta\sin\alpha, \sin\beta\sin\alpha, \cos\alpha)$ we have $\alpha = \pi/2$ on the boundary and therefore $\mathbf{n} = (\cos\beta, \sin\beta, 0)$.

For the points at which $n_z = 0$ z_x and z_y diverge, but their ratio is finite and satisfies $z_y/z_x = \tan\beta$. Equation (7) therefore takes the simpler form: $az_x + bz_y = 0$.

Replacing in (7) the ratio z_y/z_x by $\tan\beta$ we obtain: $a\cos\beta + b\sin\beta = 0$.

Writing the explicit expressions for a and b and rearranging we obtain a simple linear equation in z which can be solved almost everywhere

$$z = \frac{(l_1 I_\theta - l_3 I)\cos\beta + l_2 I_\theta \sin\beta}{(l_1\cos\beta + l_2\sin\beta)J_x}. \qquad (9)$$

Despite the elegance of (9) using it to obtain an estimate of depth at the boundary is tricky. The intensity values of the rim points are not directly accessible since points on the rim face perpendicularly from the camera. Moreover, some of the points that lie on the rim in I may face away from the camera after rotation, making the estimation of I_θ and J_x problematic. One way to overcome this problem is by extrapolating the desired intensity values from nearby observed values, as in [19]. An alternative method is presented below.

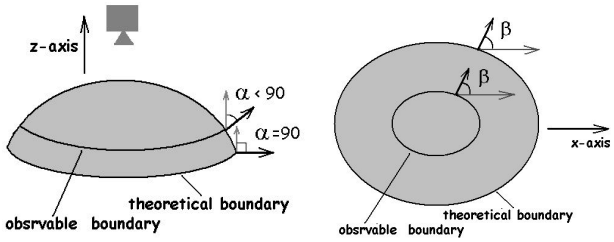

Figure 2. *Computing boundary conditions. The figure shows a surface (left) viewed from above (right). The rim of the surface forms the "theoretical boundary," but intensities are only visible in internal points. We assume that by moving within the image into the inside of the shape we reach surface points whose normal is coplanar with their respective point on the rim.*

To specify boundary conditions we assume the object is smooth and use intensities near the silhouette in the spirit of (9). Since we cannot directly access the intensities of silhouette points we use instead the following method. We make the simplifying assumption that if we move within the image from a boundary point in a direction opposite to the surface normal at the point (i.e., in the direction $\pi + \beta$ relative to the X-axis) and inspect the surface at nearby points then the normal vectors of such points take the form: $\mathbf{n} = (\cos\beta\sin\alpha, \sin\beta\sin\alpha, \cos\alpha)$ with the same β as on the "theoretical boundary". (In other words, we assume that the normal vectors at such nearby points will be coplanar with the normal at the respective boundary point, see Figure 2.) We select the nearest visible point to the boundary, and use the standard lambertian equation to determine $\mathbf{n}$:

$$I = \rho\mathbf{l}\mathbf{n} = \rho(l_1\cos\beta\sin\alpha + l_2\sin\beta\sin\alpha + l_3\cos\alpha). \qquad (10)$$

This equation now has two unknowns, ρ and α (since β is determined by the boundary orientation). To obtain an estimate of these two values we first compute for every value of α in the range $\pi/3 \le \alpha \le \pi/2$ a corresponding value of albedo, and choose ρ to be the mean of these values. Using the estimated albedo we then recompute α. Once α is estimated the derivatives of z are determined, $z_x = -\cos\beta\tan\alpha$ and $z_y = -\sin\beta\tan\alpha$, and by plugging them into (7) we obtain a linear equation in z:

$$Az = B, \qquad (11)$$

where

$$\begin{aligned} A(x,y) &= J_x(l_1 z_x + l_2 z_y + l_3) \\ B(x,y) &= I_\theta(l_1 z_x + l_2 z_y + l_3) - I(l_3 z_x + l_1). \end{aligned}$$

Due to the noise present in images, direct solution of (11) may lead to errors. So instead of solving (11) directly we solve a minimization problem with the cost function:

$$F(x, y, z, \frac{\partial z}{\partial \vec{u}}) = (Az - B)^2 + \lambda(\frac{\partial z}{\partial \vec{u}})^2, \qquad (12)$$

with a regularizing constant $\lambda > 0$. Here $\vec{u}$ is the tangent direction of the boundary contour, and $\partial z/\partial \vec{u}$ is the directional derivative of z.

3.5. Special case: frontal lighting

A particularly simple formula is obtained in the case of a frontal light source, e.g., in flash photography. In this case lighting can be expressed in the form of $\mathbf{l} = (0, 0, 1)^T$, and the coefficients in (7) become $a = -I$, $b = 0$, and $c = -I_\theta + zJ_x$. As a consequence, all quasilinear terms vanish, and (7) transforms to a system of linear ordinary differential equations (ODEs), each along a different epipolar line

$$Iz_x + J_x z = I_\theta. \qquad (13)$$

This system can be solved [28] by introducing an integrating factor $\mu(x, y) = \exp(\int_{x_0}^{x}(J_x/I)dx)$. Dividing both sides of (13) by I and multiplying by μ we obtain

$$\frac{d}{dx}(z\mu) = \frac{I_\theta}{I}\mu. \qquad (14)$$

Integrating with respect to x and dividing by μ we obtain

$$z = \frac{1}{\mu}\left(\int \frac{I_\theta}{I}\mu dx + C\right). \qquad (15)$$

The value of the constant C is determined by the boundary conditions.

4. Experiments

We have conducted several simulations and real experiments to demonstrate the utility of our formulation for shape reconstruction. In all experiments an object was rotated with respect to both a camera and a light source. The light source orientation and the rigid transformation were either set (in simulations) or measured (in the real experiments) manually. 3D shape was recovered by solving (7) using the method of characteristics (Section 3.3). In case of frontal lighting structure was recovered by solving (13) directly by the method of integrating factors (15). No regularization or smoothing was applied. Once the shape was recovered we returned to (1) to also recover the albedo.

4.1. Simulations

In the first set of experiments we applied our method to a lambertian sphere with both constant and spatially varying albedo. In the case of a constant, uniform albedo we used only one image, since any rotation of the sphere would not modify the image. We can therefore apply our method to this image with any small rotation. Figure 3(left) shows the original image and the 3D reconstruction obtained. We show in addition a subset of the characteristic curves and recovered depth on the boundary. It can be noticed that all characteristics lie in parallel planes, as is predicted by (8). Figure. 3(right) shows the application of our formulation to a pair of images of a sphere with non-uniform albedo. For albedo we used $\rho(x, y) = 0.1 + (x^2 + y^2)/2$. Shape from shading methods would require in this case that the albedo would be provided. Our method, in contrast, exploits the second image to eliminate albedo and to recover the surface. The reconstruction error, measured by $\|\hat{z} - z\|_2^2/\|z\|_2^2$, where $\hat{z}(x, y)$ denotes the estimated depth values and $\|.\|_2$ denotes the l_2 norm of a function, was 4.13% for the uniform sphere and 3.75% for the non-uniform sphere.

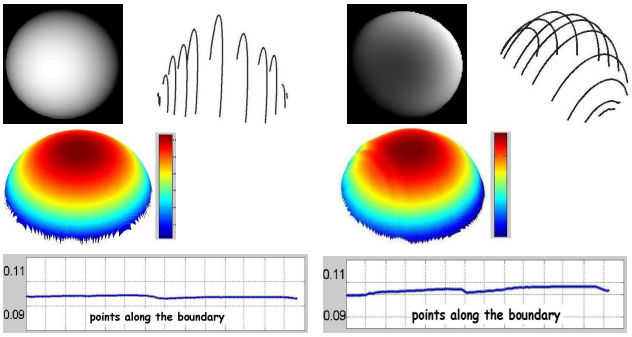

Figure 3. *Reconstruction of a synthetic sphere with uniform (left) and non-uniform smooth (right) albedo. Light directions are [-0.3, 0.2, 0.93] and [0.5, -0.3, 0.8] respectively. Top: an image used for reconstruction (one of two in the case of the non-uniform albedo) and characteristic curves; Second row: 3D reconstruction with a characteristics solution; Bottom: recovered depth along the boundary.*

In a sequence of additional experiments we reconstructed the shape of a camel toy (Figure 4). We used a model obtained using a handheld laser scanner. The first experiment shows the camel (with uniform albedo) illuminated by frontal lighting. To recover the shape of the camel from this image we used the derivation presented in Section 3.5 and solved a set of linear ordinary differential equations (13) using the method of integrating factor (15). The next three experiments show recovery with light source at generic directions, with the final experiment albedo is painted on the camel. The ground truth shape is present for comparison. These figures demonstrate that overall a faithful reconstructions is obtained.

For comparison, we applied a conventional stereo method described in [13] to these images (Figure 4) and the error of reconstruction was 38.5% at best, due to the brightness constancy assumption (compared to 5-10% with our method).

4.2. Experiments with real images

In the next experiments we reconstructed the shape of a hippo toy, a whistler, and a Pinocchio doll. While the hippo and Pinocchio have complex shapes, the whistler is rugged

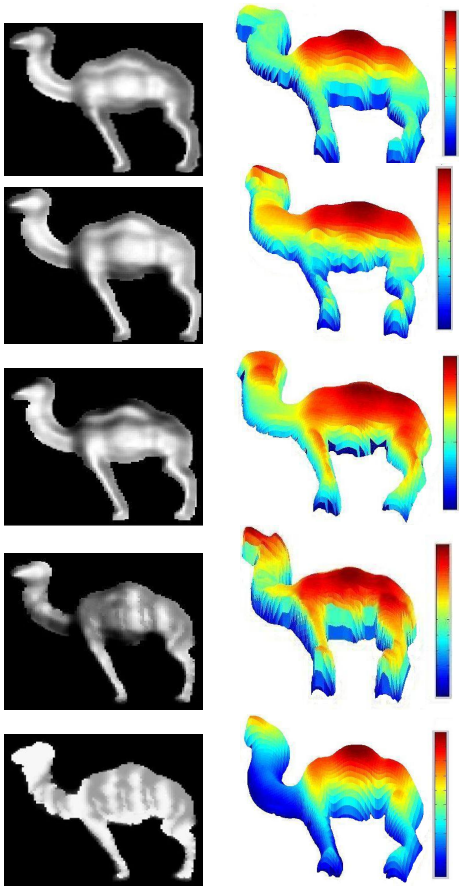

Figure 4. *Four synthetic reconstructions of a camel toy. Top: a uniform albedo camel illuminated with frontal lighting and the reconstruction obtained using the integrating factor method described in Sec. 3.5. The error of the reconstruction is 5%. The next two rows show the same camel illuminated with generic lighting (light directions respectively are [0.3; 0.1; 0.95] and [-0.2; 0.4; 0.89]). The errors of these reconstructions respectively are 9% and 10%. The following row shows a camel with non-uniform albedo illuminated from [0.5; 0.3; 0.8]; the error of the reconstruction is 8%. The bottom row shows the albedo pattern used in the last experiment and the ground truth shape (laser scan was taken from [12]).*

and has a complex albedo pattern. To measure rotation we placed the objects on a turn-table and marked rotations at 1 degree spacing. We used in these experiments rotations of up to 6 degrees. To determine the light source direction we took a picture of a ball with uniform albedo. Light source direction was indicated by the peak location in this image. In all cases we reconstructed the shapes by relying only on gray level image pairs. Once the shapes were reconstructed we used the color images to recover the albedo in the red, green, and blue channels separately. Similar results were obtained when we repeated these experiments with small mis-estimations of lighting direction and rotation angle.

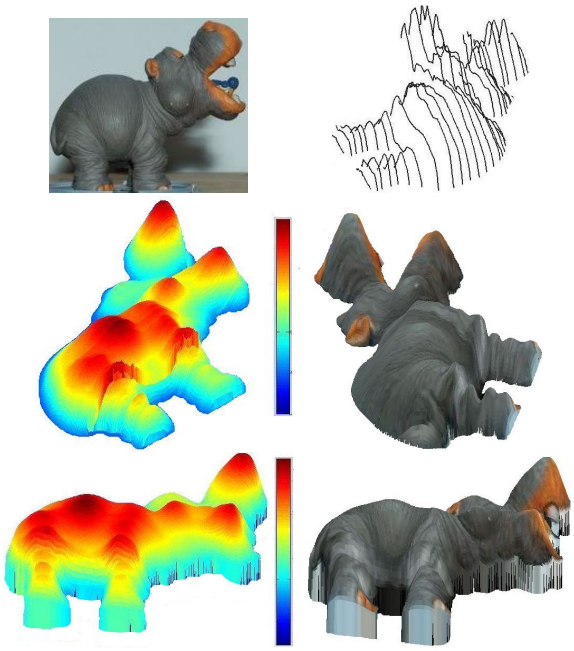

Figure 5. *Real experiment with a hippo toy (rotation angle is 5 degrees; estimated light direction is [0.1; -0.2; 0.97]). From top to bottom: The first of two images used for reconstruction and characteristic curves; two views of the reconstructed shape, without and with the recovered albedo.*

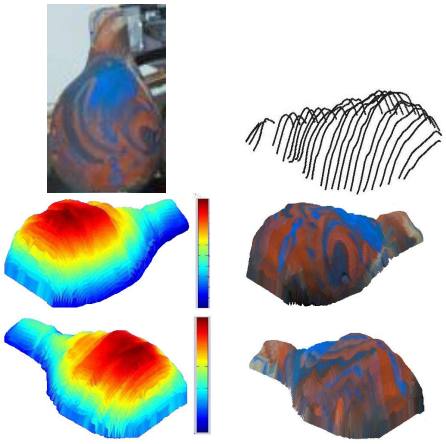

Figure 6. *Real experiment with ceramic whistle (rotation angle is 3 degrees; estimated light direction is [-0.3; -0.2; 0.93]). From top to bottom: The first of two images used for reconstruction and characteristic curves; and two views of the reconstructed shape, without and with the recovered albedo.*

5. Conclusion

We have examined the interplay between shape, lighting, and motion in the case of a rigid, lambertian object that undergoes a small motion with respect to both a stationary camera and a light source. Under these conditions we derived a single, quasilinear partial differential equation in the depth values that can be used, given motion and lighting pa-

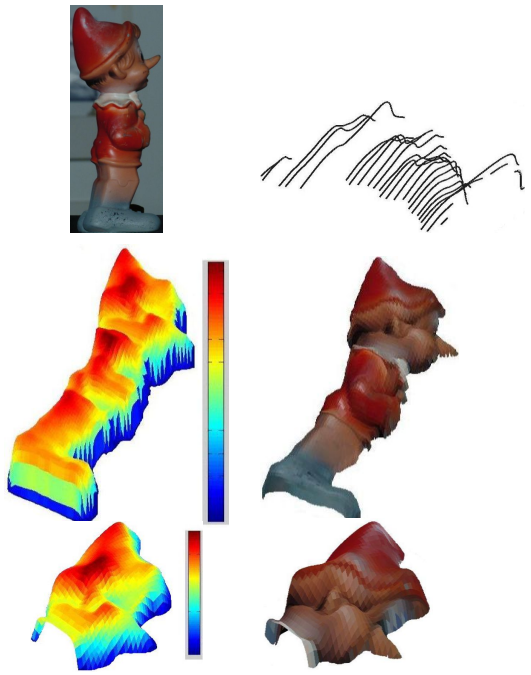

Figure 7. *Real experiment with Pinocchio toy (rotation angle is 1 degree; estimated light direction is [-0.2; 0.2; 0.95]). From top to bottom: The first of two images used for reconstruction and characteristic curves; the reconstructed shape, without and with the recovered albedo, and a zoom in into the recovered head of Pinocchio.*

rameters, to recover the shape of the object from an image pair. We further discussed solution methods by using characteristic curves, and used our formulation to derive boundary conditions at or near the object's bounding silhouette.

Our approach relies on several limiting assumptions, requiring advance knowledge of both motion and lighting. One particular difficulty is the requirement to provide the angle of out-of-plane rotation (denoted throughout this paper by θ), as it is known that the angle of out-of-plane rotation cannot be recovered from a mere two orthographic images of an object [24, 8, 20]. Despite these limitations, we believe that the simplicity of this formulation can serve to better understand how lighting and motion interact with shape to produce an image. We hope our work will further facilitate research to generalize this formulation to handle realistic settings such as perspective projection, non-stationary camera or light source, multiple light sources, and specular reflectance, and to develop new methods to recover motion and lighting parameters under realistic conditions.

References

[1] M. J. Black, D. J. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *CVIU*, 78(1):8–31, 2000. 2

[2] R. Carceroni and K. Kutulakos. Multi-View scene capture by surfel sampling: From video streams to Non-Rigid 3D motion, shape & reflectance. In *ICCV*, pages 60–67, June 2001. 2

[3] J. Cryer, P. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, July 1995. 2

[4] P. Fua and Y. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *IJCV*, 16:35–56, Sept. 1995. 2

[5] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *PAMI*, 23(6):661–673, June 2001. 2

[6] B. Horn and M. Brooks. *Shape from Shading*. MIT Press, 1989. 1

[7] B. Horn and B. Schunck. Determining optical flow. *AI*, 17(1-3):185–203, August 1981. 1, 3

[8] T. Huang and C. Lee. Motion and structure from orthographic projections. *PAMI*, 11(5):536–540, May 1989. 7

[9] H. Jin, D. Cremers, D. Wang, A. Yezzi, E. Prados, and S. Soatto. 3-d reconstruction of shaded objects from multiple images under unknown illumination. *IJCV*, 2008. 2

[10] N. Joshi and D. Kriegman. Shape from varying illumination and viewpoint. In *ICCV*, 2007. 2

[11] T. Kanade and B. Lucas. An iterative image registration technique with an application to stereo vision. In *IJCAI:674-679*, 1981. 1, 3

[12] I. Kemelmacher and R. Basri. Indexing with unknown illumination and pose. In *CVPR*, volume I, pages 909–916, 2005. 6

[13] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *ICCV*, 2001. 5

[14] H. Longuet Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Sept. 1981. 2

[15] A. Maki, M. Watanabe, and C. Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. *IJCV*, 48(2):75–90, July 2002. 2

[16] Y. Moses and I. Shimshoni. 3d shape recovery of smooth surfaces: Dropping the fixed viewpoint assumption. In *ACCV*, pages I:429–438, 2006. 2

[17] N. Mukawa. Estimation of shape, reflection coefficients and illumination direction from image sequences. In *ICCV:507-512*, 1990. 2

[18] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *PAMI*, 20(9):961–979, Sept. 1998. 2

[19] P. Nillius and J. Eklundh. Automatic estimation of the projected light source direction. In *CVPR*, page I:10761083, 2001. 4

[20] I. Shimshoni, R. Basri, and E. Rivlin. A geometric interpretation of weak-perspective motion. *PAMI*, 21(3):252–257, March 1999. 2, 7

[21] I. Shimshoni, Y. Moses, and M. Lindenbaum. Shape reconstruction of 3d bilaterally symmetric surfaces. *IJCV*, 39(2):97–110, Sept. 2000. 2

[22] D. Simakov, D. Frolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *ICCV*, pages 1202–1207, 2003. 2

[23] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography. *IJCV*, 9(2):137–154, 1992. 2

[24] S. Ullman. *The Interpretation of Visual Motion*. MIT Press, 1979. 2, 7

[25] M. Weber, A. Blake, and R. Cipolla. Towards a complete dense geometric and photometric reconstruction under varying pose and illumination. *IVC*, 22(10):787–793, September 2004. 2

[26] R. Woodham. Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1):139–144, Jan. 1980. 1

[27] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: unifying multiview stereo, photometric stereo, and structure from motion. In *ICCV*, Oct. 2003. 2

[28] D. Zwillinger. *Handbook of Differential Equations*. Academic Press, 1992. 3, 5