Department of Computer Science

Technical Report

Depth From Photomotion

Ruo Zhang, Ping-Sing Tsai and Mubarak Shah

CS-TR-93-04

University of Central Florida

Orlando, FL 32816

Depth From Photomotion

Ruo Zhang, Ping-Sing Tsai and Mubarak Shah

Computer Science Department

University of Central Florida

Orlando, FL 32816

shah@eola.cs.ucf.edu

Abstract

Traditional shape from shading techniques, using a single image, do not reconstruct accurate surfaces, and have difficulty with shadow areas. Traditional shape from photometric stereo techniques have the disadvantage that they need all of the input images together at once to minimize the total cost, and this process has to be restarted if new images become available.

To overcome the shortcomings of the above two techniques, we introduce a new technique called shape from photomotion. *Shape from photomotion uses a series of 2-D Lambertian input images, generated by moving a light source around a scene, to recover the depth map. In each of the input images, the object in the scene remains at a fixed position and the only variable is the light source direction. The movement of the light source causes a change in the intensity of any given point in the image. The change in intensity is what enables us to recover the unknown parameter, the depth map, since it remains constant in each of the input images. This configuration is suitable for iterative refinement through the use of the Extended Kalman Filter.*

Our novel method for computing shape is a continuous form of the photometric stereo technique. It significantly differs from photometric stereo in the sense that the shape estimate will not only be computed for each light source orientation, but also gradually refined by photomotion. Since the camera is fixed, the mapping between the depths at various light source locations is known, therefore, this method has an advantage over those which move the camera (egomotion), and keep the light source fixed. Results of this method are presented for sequences of synthetic and real images.

Contents

1 Introduction

A major task for computer vision is to derive a 3-D scene description from its 2-D images. This led to the development of the shape from X techniques, which include shape from motion, shape from texture, shape from stereo, shape from intensity, etc. Shape from intensity extracts shape information from a series of *intensity* images, assuming each is generated by a single light source. It can be further divided into three subcategories: Shape from shading, shape from photometric stereo, and shape from photometric sampling. The differences between them are in the number of images and arrangement of light sources.

Shape from shading uses a single light source, i.e., one image as input, to recover the shape information [4, 10, 17]. It has the advantage that it requires the least amount of input, however, this also introduces disadvantages. One disadvantage is that since it has less image information available, it is less accurate. At each pixel, intensity provides only one constraint, however, the description of surface shape (surface gradient, or surface normal) requires two parameters. Therefore, many shape from shading techniques introduce additional constraints, such as smoothness of surface, and use optimization methods to estimate shape. Another disadvantage is that since it employs only a single image, shape from shading will not be able to provide a complete description of a scene with shadow areas. Some methods also have problems when the scene is illuminated from the side.

To overcome some of the above problems, shape from photometric stereo was introduced [2, 3, 6, 7, 8, 13, 15, 16, 18]. The main idea behind photometric stereo is to take multiple images of a scene with different light source directions for each image, while keeping the viewing direction constant. Each image of the scene provides one constraint on the surface shape. Hence, multiple images of the same scene create an overconstrained system, which is solved for the surface shape. Shape from photometric stereo combines all of the input information together in order to minimize total cost. This method can only be used to compute shape of the areas which receive light from all of the light sources.

Another technique which is similar to photometric stereo is shape from photometric sampling [11, 12, 14]. It usually uses many light sources, instead of a few, and a sequence of images corresponding to the light sources to recover the shape information. The use of extra light sources, rather than only three sources, eliminates the inaccurate results caused

by the improper choices of the source positions in photometric stereo, and makes the results more accurate.

We introduce a new technique called *shape from photomotion*. In this technique, a series of 2-D Lambertian input images, generated by moving a light source around a scene, are used to recover the depth map. In each of the input images, the object in the scene remains at a fixed position and the only variable is the light source direction. This novel method for computing shape is a continuous form of the photometric stereo technique. It significantly differs from photometric stereo in the sense that the shape estimate is not only computed for *each* light source orientation, but also gradually *refined* by photomotion. Since the camera is fixed, the mapping between the depths at various light source locations is known, therefore, this method has an advantage over those which move the camera (egomotion, e.g. see [5]) and keep the light source fixed. Moving the light source requires no warping of depth maps.

In photomotion, the shape from shading method is employed for each light source orientation. Therefore, shape from shading can be considered as a special case of shape from photomotion, where only one light source is used. In contrast, photometric stereo can't be applied to a single image, it needs at least two images, and commonly three images are used.

2 Related Work

Shape from photometric stereo was first introduced by Woodham [18] in the early 1980's. In this method, Woodham proposed that the surface gradients could be solved for by using only two input images, if the surface albedo at each surface point were known. Furthermore, if the albedo were not known, both gradients and reflectance factors could be solved for by the addition of one more image. This allowed the reflection factor to vary from point to point on the surface. The method was simple and efficient, but it only dealt with Lambertian surfaces and was sensitive to noise. In his recent paper, Woodham [19] applied photometric stereo technique to compute optical flow.

Pentland's linear shape from shading [10] has problems with images of quadratic surface reflectance. Therefore, Pentland [9] proposed photometric motion to solve for shape and reflectance. The images needed in his approach were taken at different time frames while the object was rotated. The quadratic component of the surface reflectance function was

factored out by subtracting one image from another. The ratio of one of the images and the difference image was used to cancel out the albedo and to obtain the surface shape. Therefore, at least two images were required for the shape recovery. This approach was also extended to three-image photometric motion by considering second derivatives in the discrete form. The important difference between Pentland's and our method is that in his case the object is moving, therefore he needs warping to align the images. In our case the light source is moving, so no warping is needed, and we successively refine the depth map in each image.

Ikeuchi [6] was the first to obtain the shape of a specular surface using the photometric stereo method. In his research, he used a distributed light source obtained by uneven illumination of a diffusely reflecting planar surface and three input images. His solution involved solving a set of non-linear equations. A lookup table, made from the reflectance map, was used to perform the numerical inversion of the three reflectance maps. This method assumed a known object position, and required accurate measurements of reflected brightness.

Based on Ikeuchi's approach, Sanderson, Weiss, and Nayar [14] developed a structured highlighting approach for specular surfaces, which used an array of point sources for illumination. The simple property of specular reflection was used to solve for the surface normal. Results were only shown for smooth objects.

Tagare and deFigueiredo [16] estimated the shape of hybrid surfaces (those having both Lambertian and specular properties). An energy function was minimized with respect to the surface normal and the weights of the Lambertian and specular components. They proved that ten light sources were needed to get a unique solution. This approach was based on the assumption that the Lambertian and specular components could be pre-separated.

Nayar, Ikeuchi, and Kanade [11, 12] presented a method for recovering the shape of a hybrid surface, and relative strengths of the Lambertian and specular components, using an array of extended light sources. Their algorithm first separated the specular component from the Lambertian component. At each surface point, two surface orientations were calculated, one from the Lambertian component and one from the specular component. The final surface orientation was the weighted average of the two surface orientations. The implementation assumed that only two consecutive images contained non-zero specular components for each point, and was only suitable for 2-D, that is, it required that the light source, camera, and

object be coplanar.

Park and Tou [8] developed a normal vector equalization method for hybrid surfaces based on the simplified Torrance-Sparrow model for specular reflection by Healey and Binford [4]. Three input images were used to solve non-linear equations in order to extract the specular component. After the extraction of the specular component, the surface normal could be computed using Woodham's photometric stereo method for Lambertian surfaces. This was a straight-forward method whose only drawback was the need to solve a set of non-linear equations.

Coleman and Jain [3] solved for shape from hybrid surfaces using four light source photometric stereo. It was based on the assumption that only one of the light sources caused specularity for each surface point, therefore, they used relative deviation to determine the specular source. This reduced the problem to Woodham's photometric stereo solution for three sources.

Solomon and Ikeuchi [15] extended Coleman and Jain's solution by dividing the object into different areas, depending on the number of light sources illuminating them. The areas illuminated by four sources were solved by Coleman and Jain's method. Three source areas were solved by adding the constraint that the surface normals be unit vectors. Two source areas could only be solved if neither light source caused specularity.

Lee and Kuo [7] were the first ones to introduce parallel and cascade photometric stereo. In their recent paper, they showed that shape from shading algorithms had a problem that the accuracy of the reconstructed surface was related to the slope of the reflectance map function defined on the gradient space. They proposed two different photometric stereo concepts: Parallel and cascade. Parallel photometric stereo took all of the photometric images together to produce the best estimation of the surface. Cascade would take the images, one after the other, in a cascading manner. For each image, their shape from shading method, using triangular element surface approximation, was applied. The estimated shape from the previous image was used as input for the initial estimate of the next image. They used a two source photometric stereo method, and concluded that the best results could be obtained when the two light source directions were orthogonal to each other. Lee and Kuo's approach is close to ours, however, there are significant differences. We successively refine the shape estimate and explicitly use the confidence measurement (covariance matrix)

to represent the accuracy of the shape estimate. Our method for computing shape in each iteration is faster, simpler and more straightforward than Lee and Kuo's method.

Recently, Clark [2] proposed an active photometric stereo approach, which models the motion of the light source in infinitesimal steps. He was the first to use perspective, instead of orthographic, projection, thus, removing the need for the light source to be modeled at infinity. The computation was local, non-iterative, and directly solved for depth in a closed form equation. To measure the infinitesimal image gradients with respect to the change of the light source, seven images were needed to provide a discrete approximation. The problem with this approach is that the results shown in his paper were not very accurate, due to the discrete approximation of the infinitesimal gradients. To solve the problem of inaccuracy, a lot of images are needed in order to use least squares or median estimator to decrease the errors.

None of the above methods dealt with interreflections, the mutual illumination between surface facets. Nayar, Ikeuchi, and Kanade [13] were the first to challenge the interreflection problem using photometric stereo. Their observations were based on the fact that the erroneous shape extracted by shape from photometric stereo algorithms, in the presence of interreflections, was a little bit shallower than the real shape, therefore, it could be iteratively refined. The limitation of their algorithm was that it only dealt with Lambertian surfaces.

3 Shape From Photomotion

In this paper, we present a model for shape from photomotion, which uses a series of 2-D Lambertian input images, generated by moving a light source around a scene, to recover the depth map. In each of the input images, the object in the scene remains at a fixed position and the only variable is the light source direction. The movement of the light source causes a change in the intensity of any given point in the image. This change in intensity allows us to recover the unknown parameter, the depth map, since it remains constant in each of the input images. We use a Lambertian reflectance model, and employ the discrete approximation for p and q, and compute Z.

The main thesis of this approach is to recover whatever information is possible at a given time, then move to the next image to refine the previous estimates and also attempt to

recover information at new points for which previous estimates are not available.

Our formulation is suitable for the Extended Kalman Filter [1]. The basic process of the Kalman filter is as follows: A set of measurements of a fixed number of parameters are taken as input to estimate a number of unknown parameters, based on how good the current measurements are, and how accurate the current estimations are. The estimations from the previous iteration are used together with the new measurements in the current iteration in order to gradually refine the estimates. A major advantage of the Kalman filter is that it can be started at any point, stopped at any point, and continued at any time.

The reflectance function, of a Lambertian surface, at point (i, j) can be expressed as:

$$E_{i,j} = N_{i,j} \cdot L. \tag{1}$$

where $E_{i,j}$ is the gray level intensity, $N_{i,j}$ is the unit surface normal, and $L = (L_x, L_y, L_z)$ is the unit light source direction. The surface normal, $N_{i,j}$, can be expressed in terms of the gradient, $(p_{i,j}, q_{i,j})$, as:

$$N_{i,j} = \frac{(p_{i,j}, q_{i,j}, -1)}{\sqrt{p_{i,j}^2 + q_{i,j}^2 + 1}}.$$

Substituting for $N_{i,j}$ in equation 1 yields:

$$E_{i,j} = \frac{L_x p + L_y q - L_z}{\sqrt{p_{i,j}^2 + q_{i,j}^2 + 1}}. \tag{2}$$

Using the discrete approximations for $p_{i,j}$ and $q_{i,j}$ as follows:

$$p_{i,j} = Z_{i,j} - Z_{i,j-1},$$

$$q_{i,j} = Z_{i,j} - Z_{i-1,j},$$

where $Z_{i,j}$ is the depth at point (i, j), equation 2 becomes:

$$E_{i,j} = \frac{L_x(Z_{i,j} - Z_{i,j-1}) + L_y(Z_{i,j} - Z_{i-1,j}) - L_z}{\sqrt{(Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1}}.$$

Now our aim is to compute Z_{ij} such that the following function is minimized:

$$f(E_{i,j}, L, Z_{i,j}) = E_{i,j} - \frac{L_x(Z_{i,j} - Z_{i,j-1}) + L_y(Z_{i,j} - Z_{i-1,j}) - L_z}{\sqrt{(Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1}} = 0. \tag{3}$$

If we use superscript k to indicate the k^{th} input parameters and k^{th} output parameter, and approximate the above equation by a first-order Taylor expansion, we have:

$$f(E_{i,j}^k, L^k, Z_{i,j}^{k-1}) + \frac{\partial f}{\partial E_{i,j}}(E_{i,j} - E_{i,j}^k) + \frac{\partial f}{\partial L_x}(L_x - L_x^k) + \frac{\partial f}{\partial L_y}(L_y - L_y^k) + \frac{\partial f}{\partial L_z}(L_z - L_z^k) + \frac{\partial f}{\partial Z_{i,j}}(Z_{i,j} - Z_{i,j}^{k-1}) = 0.$$

where

$$\frac{\partial f}{\partial E_{i,j}} = 1, \tag{4}$$

$$\frac{\partial f}{\partial L_x} = -\frac{Z_{i,j} - Z_{i,j-1}}{\sqrt{(Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1}}, \tag{5}$$

$$\frac{\partial f}{\partial L_y} = -\frac{Z_{i,j} - Z_{i-1,j}}{\sqrt{(Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1}}, \tag{6}$$

$$\frac{\partial f}{\partial L_z} = \frac{1}{\sqrt{(Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1}}, \tag{7}$$

$$\frac{\partial f}{\partial Z_{i,j}} = \frac{(L_x(Z_{i,j} - Z_{i,j-1}) + L_y(Z_{i,j} - Z_{i-1,j}) - L_z)(2Z_{i,j} - Z_{i,j-1} - Z_{i-1,j})}{((Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1)^{\frac{3}{2}}} - \frac{(L_x + L_y)((Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1)}{((Z_{i,j} - Z_{i,j-1})^2 + (Z_{i,j} - Z_{i-1,j})^2 + 1)^{\frac{3}{2}}}, \tag{8}$$

and the partial derivatives are estimated at $(E_{i,j}^k, L^k, Z_{i,j}^{k-1})$. Note that in equation 3 we assumed that $Z_{i-1,j}$ and $Z_{i,j-1}$ are constants, which are given by the previous iteration. Therefore, in the Taylor series expansion, we have considered f in equation 3 to be a function of $E_{i,j}$, L and $Z_{i,j}$ only. $Z_{i-1,j}$ and $Z_{i,j-1}$ can also be treated as variables. This results in n^2 (for an $n \times n$ image) coupled linearized equations, each with three knowns. It can be shown that when this sparse system of linear equations is solved using Jacobi's iterative method, it reduces to the scheme used here in which $Z_{i-1,j}$ and $Z_{i,j-1}$ are treated as constants [17].

The depth map Z_{ij}^k at the k^{th} iteration can be computed iteratively using the following recursive Kalman filter:

$$Z_{i,j}^k = Z_{i,j}^{k-1} + K(Y - MZ_{i,j}^{k-1}), \tag{9}$$

$$K = S^{k-1}M(W + MS^{k-1}M^T)^{-1}, \tag{10}$$

$$S^k = (1 - KM)S^{k-1}, \tag{11}$$

$$Y = \frac{\partial f}{\partial Z_{i,j}}Z_{i,j}^{k-1} - f(E_{i,j}, L, Z_{i,j}^{k-1}), \tag{12}$$

$$M = \frac{\partial f}{\partial Z_{i,j}}, \tag{13}$$

where S is the 1 by 1 covariance matrix of the estimation error for the depth, and

$$W = \frac{\partial f}{\partial (E_{i,j}, L)} \Lambda \frac{\partial f}{\partial (E_{i,j}, L)}^T.$$

Λ is a 4 by 4 matrix which indicates the covariance of the input, i.e., the intensity E_{ij}, and source direction (L_x, L_y, L_z). The *input* of the system are a set of measurements: intensities and light sources, the *state* is indicated by the current estimates: depths. The previous estimates are combined with the current measurements in order to gradually refine the estimates. The refinement is done by considering how good the input measurements are and how accurate the estimates are.

Since our method is purely local, the computation of K, the Kalman gain, only involves the inverse of a 1 by 1 matrix. Currently, the initial value for S is set to 1 to indicate a poor initial guess, and the initial values for Λ are small, to describe input images with very little noise. The depth map is initially set to zero everywhere. Λ can also be derived using some noise models for the intensity and light source.

This method can be implemented using the following simple algorithm:

1. Initialize all the depth values, $Z^0_{i,j}$ to zero.
2. Compute the partial derivatives $\frac{\partial f}{\partial E_{i,j}}$, $\frac{\partial f}{\partial L_x}$, $\frac{\partial f}{\partial L_y}$, $\frac{\partial f}{\partial L_z}$, $\frac{\partial f}{\partial Z_{i,j}}$ from equations 4 to 8.
3. Compute K, Y and M from equations 10, 12, 13.
4. Compute the depth, $Z^k_{i,j}$, from equation 9.
5. Update the covariance matrix S^k from equation 11.
6. If the covariance matrix indicates improvements, update the depth.
7. Set $k = k + 1$ and goto 2.

The above process is done for each point in the image, and for each image in the sequence.

4 Segmentation

The recovery of accurate depth information requires that there must be adequate intensity information available. Once the depth has been recovered at a surface point, it should only be refined if there is adequate intensity information available, otherwise the refinement may degrade the recovered depth. This demonstrates the need for segmentation.

An example of inadequate intensity information are shadows. Shadows can be divided into self-shadows and cast shadows. Self-shadow is the part of an object which is not illuminated by direct light, and cast-shadow is the area projected by the object in the direction of direct light. In general, self-shadow is brighter than cast-shadow since it gets lights from interreflections. However, both of them create dark regions with very little intensity information available. Therefore, we can treat them in the same manner.

Segmentation is done during the processing of each image in the sequence. While processing the current image, the scene is segmented using the following four categories, depending on whether or not the area contains sufficient intensity information in the current and previous images:

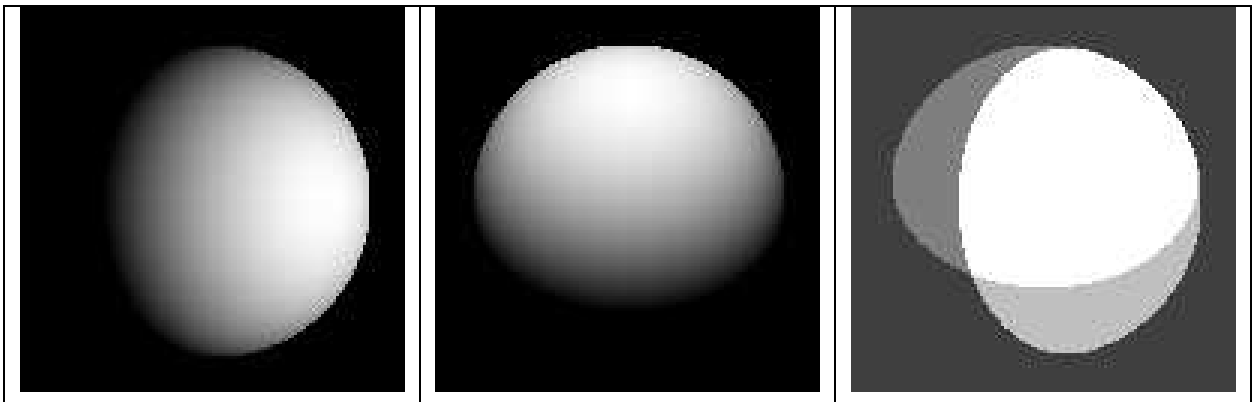
1. The areas which contain adequate intensity information in both the current image (k) and the previous image ($k-1$).

2. The areas which contain adequate intensity information in the previous image, but not the current image.

3. The areas which contain adequate intensity information in the current image, but not the previous image.

4. The areas which do not contain adequate intensity information in either one of the images.

The segmentation is done through intensity thresholding. For the first category, the estimated depth from the previous image will be refined by the current image. For the second category, the estimated depth from the previous image will remain unchanged. For the third category, the depth will be recovered for the first time using the current image. For the last category, nothing will be recovered. Figure 1 gives an example of the segmentation.



(a) (b) (c)

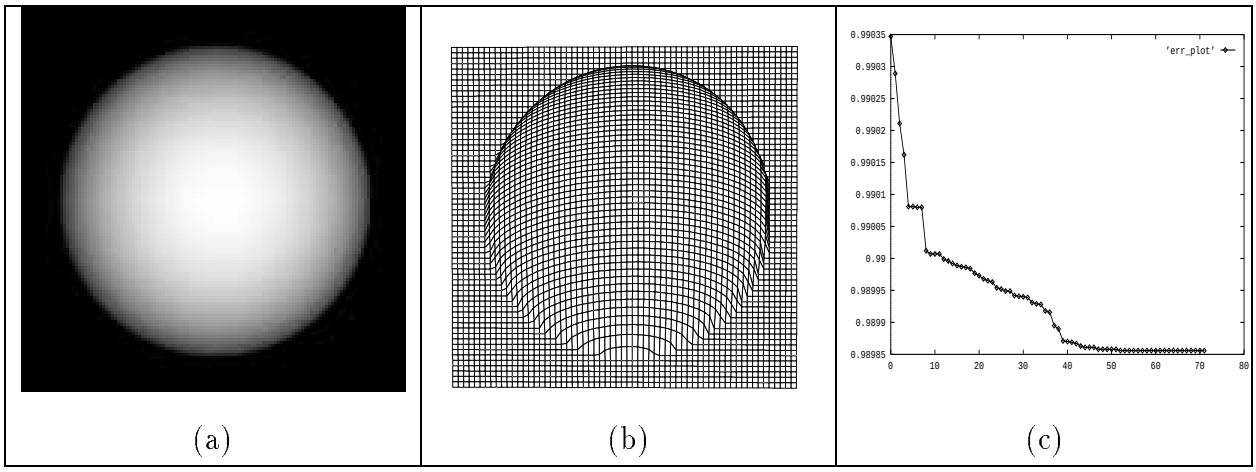
Figure 1: Sphere images for segmentation. (a) The first image (light source direction is (1, 0, 1), so the left part of the sphere is not visible). (b) The second image (light source direction is (0, 1, 1), so the bottom part of the sphere is not visible). (c) Results of segmentation. The light gray area can only be recovered in the first image, the dark gray area can only be recovered in the second image, the white area can be recovered / refined in both images, and the small dark area in the lower left corner will not be recovered in either image.

In Figure 1(c), the white area is the region of the sphere which contains adequate intensity information in both of the images, therefore, its depth is recovered by the first image, then refined by the second image. The light gray area is the region which contains adequate intensity information in the first image, so its depth is recovered by the first image and not refined by the second image. The dark gray area is the region which contains adequate intensity information only in the second image, and its depth is recovered by the second image. The small dark area in the lower left corner is the one which does not contain adequate intensity information in either one of the images, the depth in this area will not be recovered.

5 Results

5.1 Synthetic Images

The proposed method is first tested on a sequence of 72 synthetic images of a sphere. The images are generated by keeping the slant (the angle between the light source and the Z axis)



(a) (b) (c)

Figure 2: Results for the Sphere Images. (a) One of the 72 images. (b) 3-D plot of the final estimated depth map. (c) Error plot for average estimated surface normal after processing each image.

of the light source at 5 degrees, while changing the tilt (the angle between the projection of the light source in the X-Y plane and the X axis) of the light source in 5 degree steps. Figure 2 (a) gives one of the original images. Figure 2 (b) shows the corresponding 3-D plot of the recovered depth map. The error plot in 2 (c) indicates the average error in the estimated depth, compared with the true depth, after processing each image. It shows that the improvements are large in the beginning iterations, then become relatively small after the results become stable.

Here, we want to emphasize that it is not always necessary to use 72 images in our method, as it is clear that the recovered depth map is quite good after processing a few images. Our main aim in this experiment is to show that, with additional images, the shape estimate either improves or becomes stable.

The second test is performed on a set of synthetic cake images. These images were generated with the tilt of the light source changing in 90 degree increments. Each of the images contain different shadow areas caused by the layers of the cake, as shown in figure 3. After processing the first image, the shadow area has not been recovered (figure 3(e)). The subsequent images refine the original depth estimates to yield a good representation of the shape of the cake (as shown in figures 3(f)-(h)). The hole on the left side in figure 3(e), caused by the shadow area of the first image, is filled.

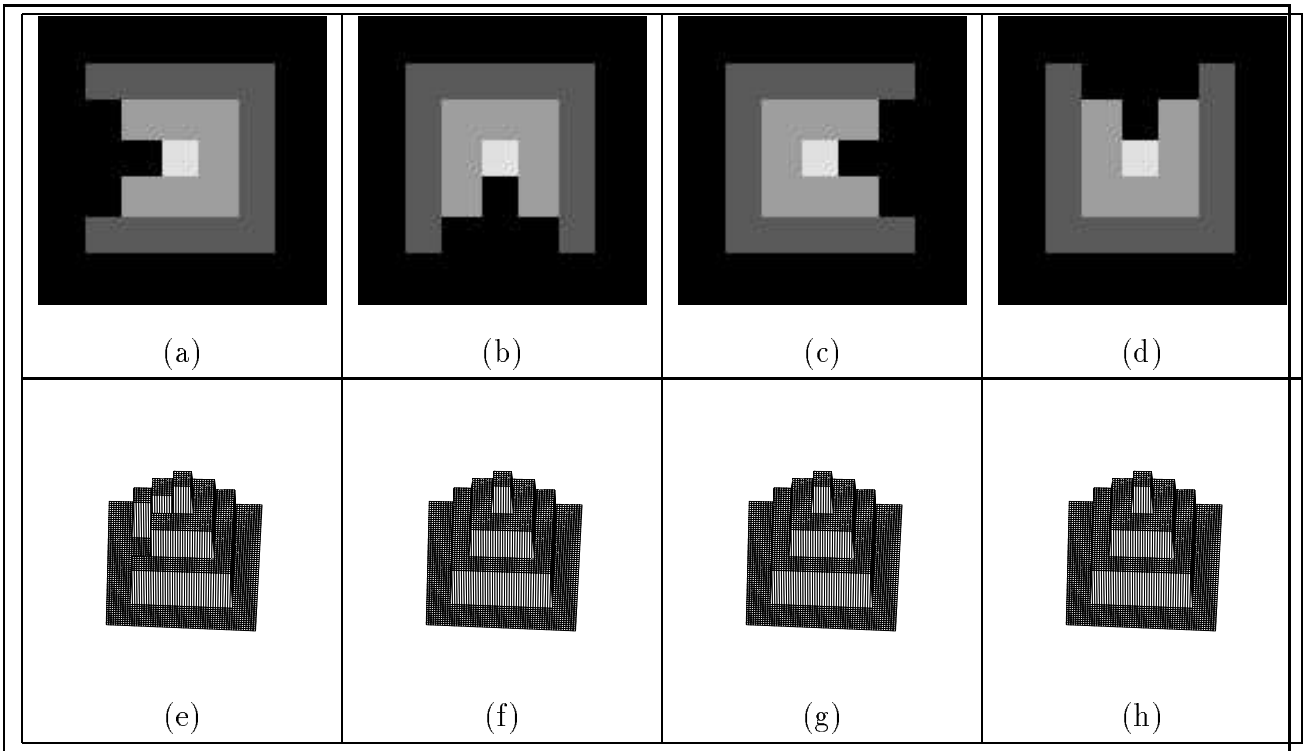


Figure 3: Results for Cake Images. (a) First image (light source direction is (1, 0, 1)). (b) Second image (light source direction is (0, 1, 1)). (c) Third image (light source direction is (-1, 0, 1)). (d) Fourth image (light source direction is (0, -1, 1)). (e) 3-D plot after processing the first image. (f) 3-D plot after processing the second image. (g) 3-D plot after processing the third image. (h) 3-D plot after processing the fourth image.

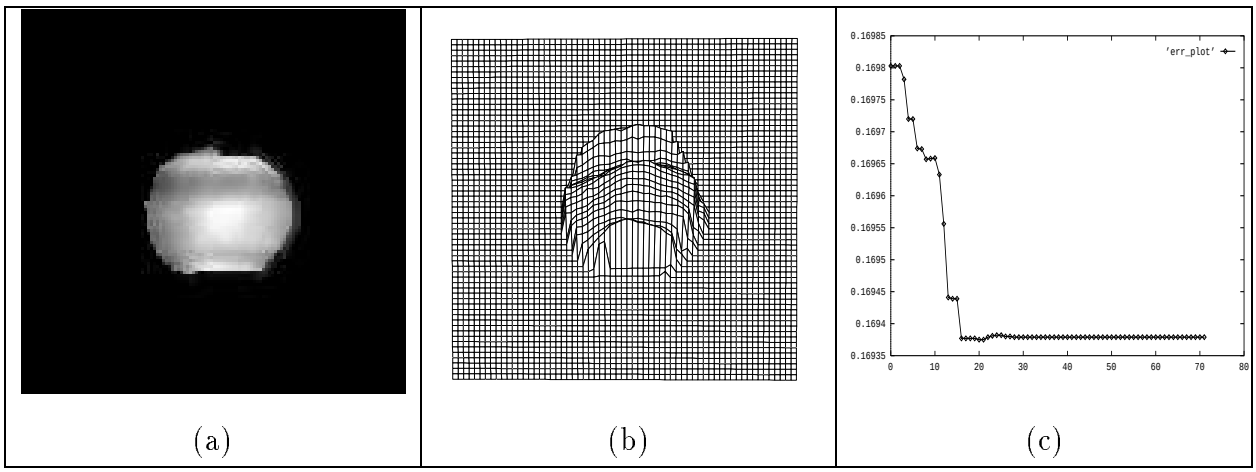


Figure 4: Results for Tomato Images. (a) One of the 72 tomato images. (b) 3-D plot of the final estimated depth map. (c) Error plot for average estimated surface normal after processing each image.

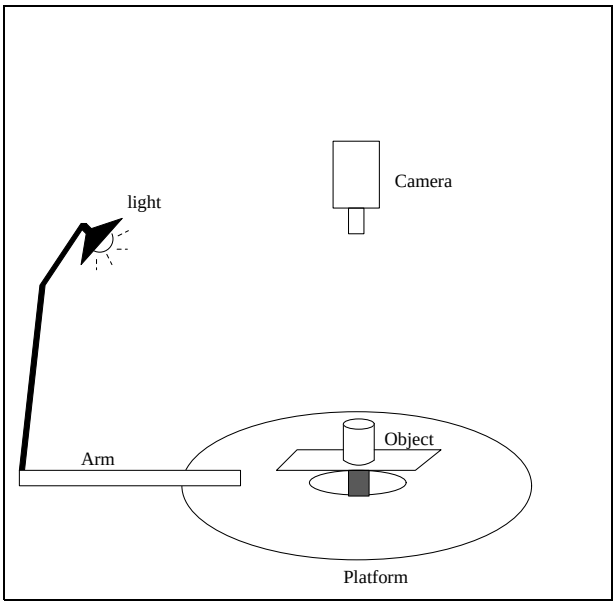


Figure 5: The experimental setup for taking real images.

This algorithm is also tested on a sequence of tomato images generated using a range image. The tilt of the light source was moved in steps of 5 degrees, and the slant remained at 5 degrees. Figure 4 shows one of the input images, the 3-D plot of the recovered depth, and the error plot. The error plot was generated by comparing the recovered depth, after processing each input image, with the original range image.

5.2 Real Images

Next, two tests are performed on real image sequences taken with a video camera. The experimental setup is shown in figure 5. A platform, with a hole in the center and an arm on the side, is used to rotate the light source around the object. The object is placed in the center hole, and a lamp is attached to the arm. The camera is directly above the object. In order to create images with different light source directions, the platform is rotated to align the arm with each of the light source directions, while the object in the center hole remains stationary.

The first sequence only contains two images (Figure 6). The objects in the scene are a wooden block and a paper box. The wooden block is placed on the top of the paper box, which creates shadows on the paper box. The first image (shown in figure 6(a)) has a shadow

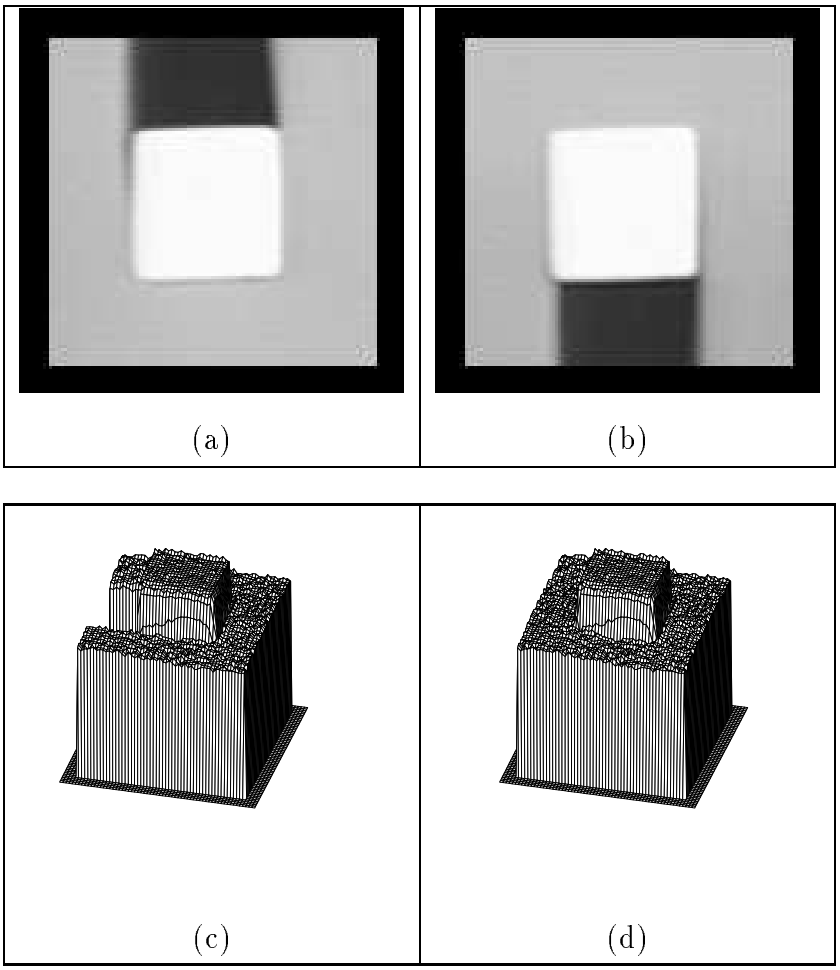
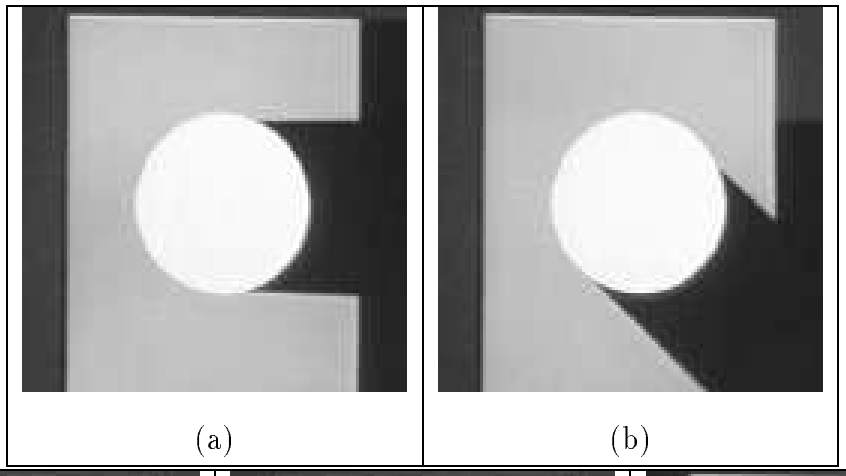


Figure 6: Results for block Images. (a) First image (light source direction is (0, -1, 1)). (b) Second image (light source direction is (0, 1, 1)). (c) 3-D plot after processing the first image. (d) 3-D plot of the final estimated depth map.



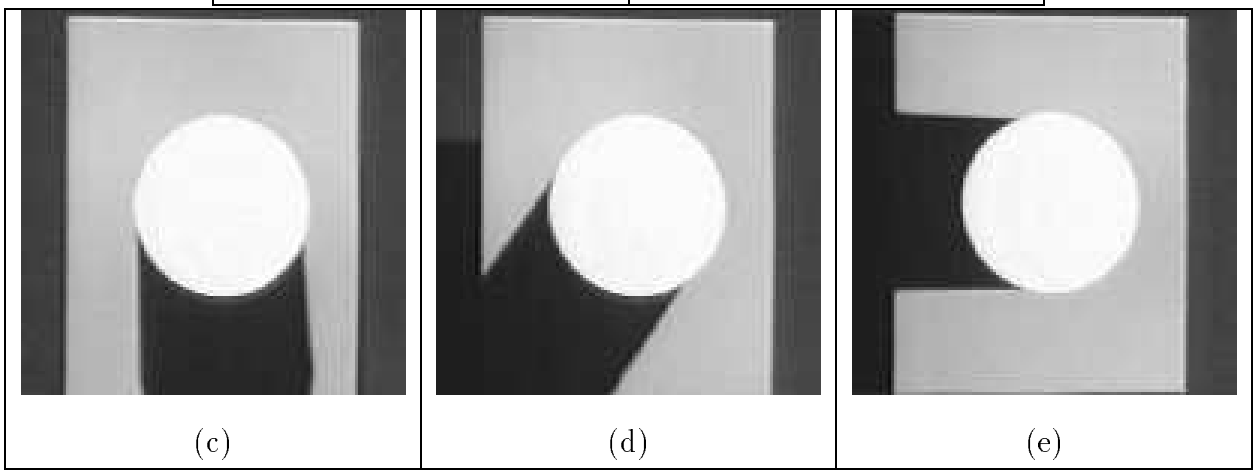


Figure 7: Results for the Cylinder Images. (a) The first image (light source direction is (-1, 0, 1). (b) The second image (light source direction is (-0.7, 0.7, 1). (c) The third image (light source direction is (0, 1, 1). (d) The fourth image (light source direction is (0.7, 0.7, 1). (e) The fifth image (light source direction is (1, 0, 1).

area on one side of the block, and the second image (shown in figure 6(b)) has a shadow area on the other side of the block. The shadow area in the first image is recovered through the second image. The rotated (to provide a good view) 3-D plots after processing each image are given in figures 6(c) and 6(d).

In another sequence, we took images of a cylinder sitting on top of a block. There were a total of five images. The light source was rotated from 0 to 180 degrees, in steps of 45 degrees. The original set of images, and the needle map describing the surface orientation after processing, are shown in figures 7 and 8. We can see that the shadow area from the first image is gradually recovered during the processing of the following images.

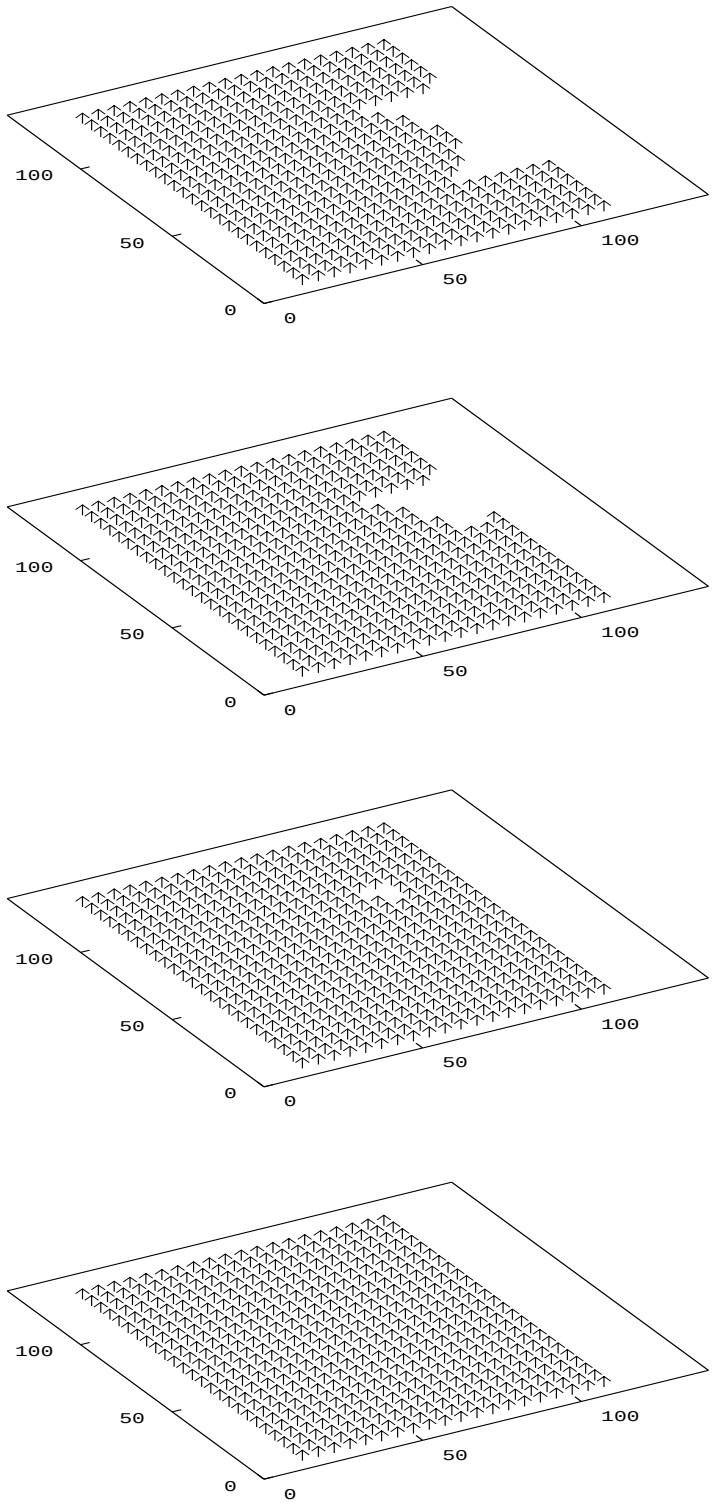


Figure 8: The needle map after processing the first, second, third, fourth image. The fifth needle map is the same as the fourth, so it is not shown.

6 Conclusion

Shape from photomotion iteratively refines the shape obtained from a sequence of images. It can take any sequence of images, in any order, as input. The process can be stopped at any iteration, and restarted from any other point without loss of accuracy. When there is only one input image, shape from photomotion is reduced to shape from shading.

So far, the movement of the light source we have considered has been arbitrary. In order to get better results and to reduce the number of input images required, we can also try to determine the optimal movement of the light source. The optimal movement of the light source is affected by the shape recovered from the current image and the shadows. It's always better if the light source is moved to illuminate as much of the shadow region as possible in the next image, while maintaining enough overlapping of bright regions. By moving light in an optimal manner, we can reduce the number of images required to a minimum, and eliminate shadow regions as much as possible. Also, we would like to extend the photomotion to specular surfaces.

References

[1] N. Ayache and O. D. Faugeras. Maintaining representations of the environment of a mobile robot. *IEEE Transactions on Robotics and Automation*, 5:804–819, 1989.

[2] J. J. Clark. Active photometric stereo. *IEEE Computer Vision and Pattern Recongnition*, pages 29–34, 1992.

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

[4] G. Healey and T. O. Binford. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42:62–86, 1988.

[5] J. Heel. *Temporal Surface Reconstruction.* PhD thesis, MIT, 1991.

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

[7] K. M. Lee and C. C. J. Kuo. Shape reconstruction from photometric stereo. In *Computer Vision and Pattern Recongnition*, pages 479–484, 1992.

[8] J. S. Park and J. T. Tou. Highlight separation and surface orientations for 3-d specular objects. *IEEE International Conference on Pattern Recongnition*, pages 331–335, 1990.

[9] A. Pentland. Photometric motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:879–890, 1991.

[10] A. P. Pentland. Local shading analysis. *IEEE Transactions on PAMI*, 6:170–187, 1984.

[11] K. Ikeuchi S. K. Nayar and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. In *IEEE Trans. on Robotics and Automation*, volume 6, pages 418–430, 1990.

[12] K. Ikeuchi S. K. Nayar and T. Kanade. Extracting shape and reflectance of hybrid surfaces by photometric sampling. In *DARPA Image Understanding Workshop*, pages 563–577, 1990.

[13] K. Ikeuchi S. K. Nayar and T. Kanade. Shape from interreflections. In *Third International Conference on Computer Vision*, pages 1–11, 1990.

[14] A.C. Sanderson, L.E. Weiss, and S.K. Nayar. Structured highlight inspection of specular surfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10:44–55, 1988.

[15] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. In *Computer Vision and Pattern Recongnition*, pages 466–471, 1992.

[16] H. D. Tagare and R. J. P. deFigueiredo. Simultaneous estimation of shape and reflectance maps from photometric stereo. *IEEE International Conference on Computer Vision*, pages 340–343, 1990.

[17] P. S. Tsai and M. Shah. A simple shape from shading algorithm. *Personal Communication*, 1992.

[18] R. J. Woodham. Photometric method for determing surface orientation from multiple images. In M. J. Brook and B. K. P. Horn, editors, *Shape from Shading*, pages 513–532. MIT Press, 1989.

[19] R. J. Woodham. Multiple light source optical flow. *IEEE International Conference on Computer Vision*, pages 42–46, 1990.