In IEEE CVPR-93, New York City, NY, June, 1993.

Shape From Photomotion

Ruo Zhang, Ping-Sing Tsai and Mubarak Shah *
Computer Science Department
University of Central Florida
Orlando, FL 32816

Abstract

We introduce a new technique called shape from photomotion. *It uses a series of 2-D Lambertian images, generated by moving a light source around a scene, to recover the depth map. In each of the 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. Our method 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.*

1 Introduction

Shape from shading uses a single image to recover the shape information. 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. 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. To overcome some of the above problems, shape from photometric stereo was introduced. 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. The information in the images is combined together in order to minimize total cost. This method can only be used to compute the shape of areas which receive light from all of the sources.

We introduce a new technique called *shape from photomotion*. In this technique, a series of 2-D Lambertian 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), and keep the light source fixed, because no warping of depth maps computed with different camera locations is needed.

Lee and Kuo [1] were the first ones to introduce parallel and cascade photometric stereo. Parallel photometric stereo took all of the photometric images together to produce the best estimation of the surface. Cascade took the images, one after the other, in a cascading manner. Their shape from shading method, using triangular element surface approximation, was applied for each image. The estimated shape from the previous image was used as input for the initial estimate of the next image. The difference between Lee and Kuo's approach and ours is that 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.

2 Shape From Photomotion

In the following section, we assume a Lambertian reflectance model, and employ the discrete approximation for surface gradients p and q. Let $E_{i,j}$ be the gray level intensity, $N_{i,j}$ be the unit surface normal, and $L = (L_x, L_y, L_z)$ be the unit light source direction, and $Z_{i,j}$ be the depth at point (i, j). Our aim

*Send email to shah@sono.cs.ucf.edu for an extended version of this paper.

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

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

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:

$$f(E^k_{i,j}, L^k, Z^{k-1}_{i,j}) + \frac{\partial f}{\partial L_x}(L_x - L^k_x) + \frac{\partial f}{\partial L_y}(L_y - L^k_y) + \frac{\partial f}{\partial L_z}(L_z - L^k_z) + \frac{\partial f}{\partial Z_{i,j}}(Z_{i,j} - Z^{k-1}_{i,j}) = 0.$$

The partial derivatives are estimated at $(E^k_{i,j}, L^k, Z^{k-1}_{i,j})$.

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

$$\begin{aligned} Z^k_{i,j} &= Z^{k-1}_{i,j} + K(Y - MZ^{k-1}_{i,j}), \\ K &= S^{k-1}M(W + MS^{k-1}M^T)^{-1}, \\ S^k &= (I - KM)S^{k-1}, \\ Y &= \frac{\partial f}{\partial Z_{i,j}} Z^{k-1}_{i,j} - f(E_{i,j}, L, Z^{k-1}_{i,j}), \\ M &= \frac{\partial f}{\partial Z_{i,j}}, \end{aligned}$$

where I is the identity matrix, 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. Since our method is purely local, the computation in the Kalman filter only involves the inverse of a 1 by 1 matrix.

3 Segmentation

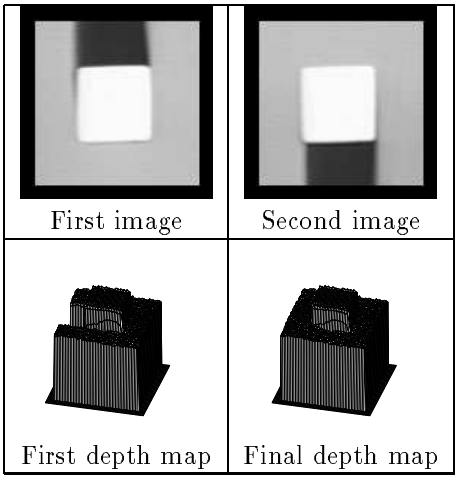
The recovery of accurate depth information requires that there be adequate intensity information available. Once the depth has been recovered at a surface point, it should only be refined if there is adequate intensity information available, otherwise the refinement may degenerate the recovered depth.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

Segmentation is done during the processing of each of the images 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: The areas which contain adequate intensity information in both the current image (k) and the previous image ($k-1$); the areas which contain adequate intensity information in the previous image, but not the current image; the areas which contain adequate intensity information in the current image, but not the previous image; the areas which do not contain adequate intensity information in either one of the images.

Using the results from segmentation, we decide whether the depth in an area should remain unchanged, be recovered, or be refined.

4 Results

The results are shown here for a sequence of two images. The images were taken using a video camera. 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 to create shadows on the paper box. The first image has a shadow area on one side of the block, and the second image 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 3-D plots after processing each image are given in order to provide a good view.



References

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