Motion from Shape from Shading

Soo Hyeon Seong and Masanobu Yamamoto
Graduate School of Science and Technology, Niigata University
8050 Ikarashi 2-nocho, Niigata-city, 950-2102, Japan
e-mail: seong@vision.info.eng.niigata-u.ac.jp
yamamoto@info.eng.niigata-u.ac.jp

Abstract

We propose a novel method for estimating the 3D motion parameters of an object using two successive images from a single camera. The method recovers the 3D shape of the object from shading information in each image, then estimates 3D motion parameters from two successive 3D shapes, and, finally determines the optical flow; whereas, the existing methods begin with optical flow estimations while assuming constant brightness during motion. Our method is useful even when the brightness of the pixel in motion does not remain constant.

1 Introduction

Human beings are able to perceive 3D motion of objects even from a sequence of 2D images of videos or movies. In order to process this vision ability by computer, the following method has been used. First, the optical flow as the correspondence between successive images is estimated from an image sequence. Next, assuming an object is rigid, the obtained optical flow is interpreted as a result of 3D motion. The 3D shape is also recovered simultaneously. This method is called the Two Stages Estimation Method. Since estimation of 3D motion in the latter stage is very sensitive to a noise, correspondence between successive images in the preceding stage must be performed correctly.

The schemes for correspondence between successive images are known as follows:
(1) The features scheme.
(2) The temporal-spatial gradient scheme.

In order to apply these schemes to algorithm, various assumptions should be needed. First, the assumption of the scheme using the features such as edge and corner on the image is that location of the point on the surface of the object corresponding to the feature point is unchangeable to the motion of the object. However, the feature points corresponding to shadow, highlight, and occluding boundary etc., cannot express the motion of the object correctly. Next, the assumption of temporal-spatial gradient scheme is that the image brightness of the point on the surface of the object is invariant before and after motion. However, Verri and Poggio [1] proved that this assumption does not satisfy in the general cases, although it does satisfy in the case that the object of Lambertian surface is displaced parallel. Much effort is devoted to improve the accuracy of the temporal-spatial gradient scheme [2, 3, 4].

In this paper, we first indicate that an example does not satisfy assumptions of the existing scheme for the correspondence between images. Next, we propose a novel method of replacing the order of the Two Stages Estimation Method, for such cases. That is, 3D shape is obtained previously and 3D motion is estimated from the obtained depth maps. The method of recovering the 3D shape from shading information in images, and the 3D motion estimation method from depth map sequences are used as tools in our method.

2 Inverse Two Stages Estimation Method

In this section, we indicate a case in which it is hard to estimate the image displacement vector from an image sequence. Figure 1 is an image sequence when a hen's egg is rotating around the vertical axis passing through the center of the egg. Here, the vertical axis is parallel to a vertical axis of an image. The relation of the camera, egg, and light source is pictured on Figure 2. The camera projection is orthographic.

First, we describe about the correspondence between images by the feature point. The feature

points such as edge and corner are detected as positions where the image brightness changes suddenly. In Figure 1, the silhouette of the egg could be detected as an edge. The image displacement vector, however, cannot be estimated from this silhouette because the silhouette position on the egg differs from before and after rotation, since the silhouette is the occluding boundary. In fact, in Figure 2, the silhouette before rotation is a projection of the occluding boundary position C. The correct position after rotation is C'. However, the silhouette after rotation is observed as a projection of the occluding boundary position D. Therefore, the silhouette as a feature point does not match to the correct position on the object during rotation.

Next, we consider about an application of the temporal-spatial gradient scheme. Let $E(x, y, t)$ be the image brightness on the point (x, y) at time t. We suppose that the point (x, y) moved to the point $(x+u, y+v)$ at next time. Where, (u, v) is the displacement vector, correspond to $\partial x/\partial t$, $\partial y/\partial t$. The image brightness can be represented as a Taylor series expansion,

$$E(x+u, y+v, t+1) = E(x, y, t) + E_x u + E_y v + E_t + O(2) \quad (1)$$

where, (E_x, E_y) is the spatial gradient, and E_t is the temporal gradient; they are obtained by partial derivative in a subscript, respectively. And, $O(2)$ contains second- and higher-order terms in δx, δy, and δt.

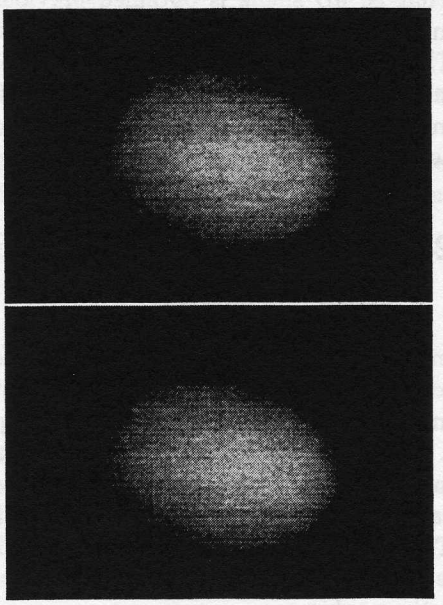

Figure 1: Images of the egg. The first image (upper) and the second image (down).

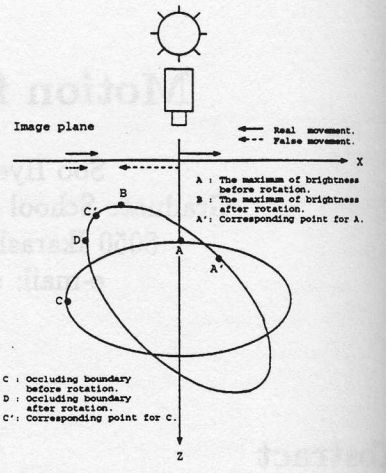

Figure 2: Problems in image correspondences.

If the image brightness does not change after rotation, $E(x+u, y+v, t+1) = E(x, y, t)$. And, $O(2)$ is usually assumed small and set to zero. Therefore, equation (1) can be rewritten as the following first-order equation for unknowns, the displacement vector (u, v).

$$E_x u + E_y v + E_t = 0 \quad (2)$$

Equation (1) is called the optical flow constraint equation [5], its solutions are referred to as optical flow. Since the surface of the egg can be modeled as a Lambertian surface, the image brightness on the surface of the object is determined by the light source direction and the surface orientation of the object as shown in Figure 2*. When the normal vector on the surface is the same as the direction of the light source, the image brightness becomes the maximum. That is, it is point A before rotation, and it is point B after rotation. If the assumption of the constancy of the image brightness at a displacement destination could be satisfied, then point A would corresponds to point B. However, the correct corresponding destination of point A is point A'. Since the image brightness of point A and point A' differs, we conclude that $E(x+u, y+v, t+1) \neq E(x, y, t)$. Therefore, the assumption of the optical flow constraint equation is not valid.

In the optical flow constraint equation, a correction term is needed for explaining an image brightness change. By the introduction of the correction term, some additional constraints are obtained [2, 3, 4]. However, these results are inapplicable

*For the explanation, the position of the light source in figure 2 differs from the position of the light source in figure 1.

to the case of the egg. For explanation of the reason, we consider another assumption which is the temporal-spatial gradient scheme. The spatial gradient of image brightness must arise by the inequality of the reflectance on the surface. The reflectance on the surface of the egg is almost constant. The spatial gradient of the image brightness on the surface of the egg is not caused by the difference of the reflectance on the surface but caused by the shading. Since a silhouette is the occluding boundary of an object and a background, there is a difference in the reflectance between the regions which surround a silhouette. However, as mentioned previously, since the position on the surface of the object corresponds to silhouette changes, the correct correspondence cannot be estimated from the temporal-spatial gradient scheme, either.

That is, the correspondence between images of the egg using the temporal-spatial gradient scheme cannot be estimated correctly. The case of the egg is generalized as follows; if the surface of an object has a uniform reflectance and smoothness, the 3D motion of the object is not applicable to the conventional method based on the correspondence between images.

For such an object, we propose a new method to estimate 3D motion. If an object has a Lambertian surface, we first recover the 3D shape from the occluding boundary information by the shape from shading algorithm. Furthermore, if an object is rigid, we estimate the 3D motion from 3D shape of the object for each time [6]. If the 3D motion parameters are obtained, the optical flow is determined easily. That is, first, the 3D shape of the object for each time is recovered from an image sequence. Next, the 3D motion is estimated from recovered 3D shapes. The optical flow is subsequently determined. We call this method an Inverse Two Stages Estimation Method because the flow of processing of the image sequence analysis has the inverse order of the conventional method.

3 Recovering the 3D Shape from Shading

Many methods for recovering the 3D shape of an object from a single image are proposed [7, 8]. In this paper, the relaxation method proposed by Ikeuchi [9] is used.

Let a coordinate system of the scene be a Cartesian (x, y, z) one. We suppose that an object is orthogonally projected on the image plane (x, y). Here, the z axis is an optical axis. By assuming that light source is far away from the object, rays are irradiated parallel on the surface of the object. Let the depth map of an object be $z = F(x, y)$. $(p, q) = (\partial F/\partial x, \partial F/\partial y)$ expresses the surface gradient, and the pq-plane is called gradient space. The surface normal of the object is expressed to $(p, q, -1)$ by using the surface gradient (p, q). The direction of the light source can also be expressed to $(p_s, q_s, -1)$ using the gradient (p_s, q_s).

If the surface of an object is the Lambertian surface, the image brightness of the surface is proportional to the cosine of the incident angle. If we normalize by setting the maximum brightness to one, the radiance will be given by the following equation.

$$R(p, q) = \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}} \quad (3)$$

It can be considered that this image brightness is the function of the surface gradient. This is called the reflectance map. If the reflectance map is given, the surface gradient will be obtained as a solution of the following image irradiance equation, where the maximum brightness of the image is normalized to one.

$$R(p, q) = E(x, y) \quad (4)$$

Since two unknowns p, q are contained in the equation (4), an image irradiance equation cannot determine the solution uniquely. If the surface of an object is smooth, then we can determine the solution using the relaxation method with an occluding boundary.

In Figure 1, there is a place which can determine the surface gradient immediately. It is the silhouette of the egg, which is called the occluding boundary. The surface normal is a direction perpendicular to viewer direction and a silhouette. However, this surface normal cannot be expressed with the gradient (p,q). This is because the gradient value becomes infinite. Here, we match the surface gradient by 1 to 1 in the unit circle on the stereographic space f and g using the following conversion equation. The reflectance map can be rewritten with $R_s(f, g)$ as the function f and g.

$$f = \frac{2p}{\sqrt{1 + p^2 + q^2} + 1}, g = \frac{2q}{\sqrt{1 + p^2 + q^2} + 1} \quad (5)$$

(f, g) can be obtained by the minimization of the following error function,

$$\int\int (E(x, y) - R_s(f, g))^2 + \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2) \, dxdy, \quad (6)$$

where, the first term is the error of an image irradiance equation, and the second term is the evaluation value which measures the smoothness of a surface. Moreover, the weight λ is constant; it multiplies and maintains the balance of the two evaluation values. Here, the surface gradient at the occluding boundary are used as the boundary conditions.

The depth map $z = F(x, y)$ is calculated as follows from the obtained (f, g). First, the surface gradient (p, q) is obtained from (f, g) by an inverse conversion of the equation (5).

$$p = \frac{4f}{4 - f^2 - g^2}, \quad q = \frac{4g}{4 - f^2 - g^2} \tag{7}$$

Next, let a point on the surface of the object be (x_0, y_0, z_0). If the point (x_0, y_0, z_0) is given, the depth map is obtained by integrating along a curve from the given starting point (x_0, y_0, z_0) [8].

$$z(x, y) = z_0(x_0, y_0) + \int_{(x_0, y_0)}^{(x,y)} (pdx + qdy) \tag{8}$$

4 Motion Estimation from Depth Map sequence

The obtained depth map is $z = F(x, y, z)$. Estimation of the 3D motion parameters are performed as follows from a depth map sequence [6].

First, the temporal-spatial gradient scheme is extended to depth map sequence at the time of an image sequence. Let a point on an object be $(x(t), y(t), z(t))$. If we suppose that an object is moved, then a point on the object is also moved. Since this point is measured as a depth map sequence, it can be calculated by the following equation.

$$z(t) = F(x(t), y(t), t) \tag{9}$$

Let a 3D velocity vector be $(u, v, w) = (dx/dt, dy/dt, dz/dt)$. Differentiating equation (9) with respect to time t, we obtain a constraint equation of a 3D velocity vector on the surface of the object,

$$w = pu + qv + r, \tag{10}$$

where $r = \partial F/\partial t$ is easily obtained from time difference of a depth map sequence.

On the other hand, the 3D velocity vector (u, v, w) on the surface of the object can be expressed using motion parameters. The motion of the object can be expressed by the linear sum of rotational displacement and translation displacement around the axis passing through the center of the object (x_c, y_c, z_c). The 3D velocity vector is given by the following equation using the rotational velocity vector $(\omega_x, \omega_y, \omega_z)$ and the translation velocity vector (T_x, T_y, T_z) as the motion parameters.

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \times \begin{pmatrix} x - x_c \\ y - y_c \\ z - z_c \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \tag{11}$$

Substituting (u, v, w) of equation (11) into equation (10), we obtain the following linear equation with the motion parameters as unknowns.

$$\begin{aligned} &pT_x + qT_y - T_z - q(z - z_c) + y(y - y_c)\omega_x \\ &+p(z - z_c) + (x - x_c)\omega_y \\ &+q(x - x_c) - 1p(y - y_c)\omega_z + r = 0 \end{aligned} \tag{12}$$

Solving the system of the linear equations (12) from a lot of points on the surface of the object, we can get the motion parameters.

Substituting the obtained motion parameters into equations (11), a 3D velocity vector is obtained. Here, (u, v) is a velocity vector observed on the image. Although a velocity vector is 1D approximation of the displacement vector, when a motion is small, it may be regarded as the displacement vector.

5 Experiment

We demonstrate the performance of the proposed method by experiment with real images. A white egg is used as an object, since the surface of egg can be considered to be Lambertian.

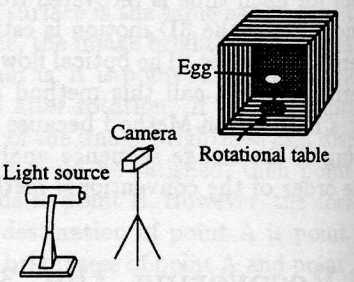

Figure 3: Experiment setup.

Figure 3 shows an experimental setup. First, a big box is set. The size of the box is $50cm$, $55cm$, and $70cm$ in height, width, and depth. The inside wall of the box is covered with black paper. Next, a rotational table is set into the box. The distance between the rotational table and entrance of the box is about $40cm$. Then, the egg is put on the

rotational table. A camera and a light source were located in the place where the distance from the object is about $2m$.

The real images were obtained by a CCD camera (SONY, Handycam CCD-TR1000)*, a workstation (SUN Sparc Station 2) and an image input board (Xvideo, Parallax Graphics).

The real images were taken using a zoom lens, by the camera placed sufficiently far compared with the size of the object. The view angle of the camera is about 5°. For this reason, when the motion of the object is small, the rate of reduction by perspective projection becomes almost constant, irrespective of changes in the distance to the object. Therefore, it can be considered that the camera model is an orthogonal projection (i.e., weak perspective projection). Moreover, we are able to consider that rays are illuminated parallel on the surface of the object. Since a fluorescent light is used as the light source, and the distance between the light source and the object is large enough to the size of the object.

The shape of the egg is an ellipsoid. The shape of the one side of long axis is sharpened, and the shape of the other side is predominantly a globe. The reflectance map was obtained from image of the globe shape side. The image is shown in Figure 4.

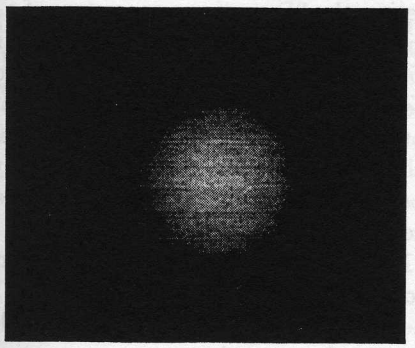

Figure 4: Image of the egg for reflectance map.

The length of the egg's long axis is about $50mm$. If a cross-section shape which intersects perpendicularly with the egg's long axis is mostly considered to be a circle, the maximum cross-section diameter is about $40mm$.

The images before and after rotation were obtained by the following operation. The egg was placed on the rotational table so that long axis would be horizontal, and intersect perpendicularly with the optical axis of the camera. A handle of the rotational table can be rotated manually and a rotation angle can be measured by reading the scale. The scale can be read to 1/100°. The rotation angle given by this experiment was 3°. The rotation axis is a vertical axis passing through the center of the egg.

The obtained image (figure 1) has 256 gray level as for image brightness, maximum brightness to 255, minimum brightness to 0. The size of an image is 638×480 pixels. The length of the egg's long axis on the image was 200 pixels, and the maximum diameter was 158 pixels. The true egg's long axis becomes $50mm$ then 0.25 mm/pixel*. Since a camera model is an orthogonal projection, the unit of the translation velocity vector (T_x, T_y, T_z) is pixel/frame, or $0.25mm$/frame.

Since the background is darkened, compared with the surface of the egg, an occluding boundary is obtained by measuring the threshold of image brightness. The left picture of Figure 5 shows the occluding boundary.

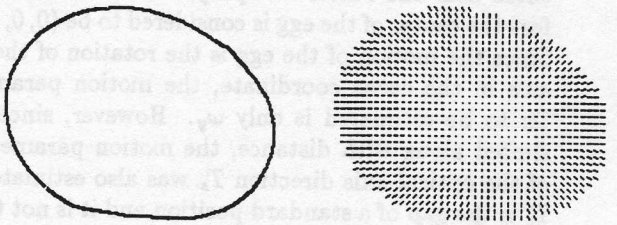

Figure 5: Occluding boundary (left) and needle diagram (right) on the egg.

The right picture of Figure 5 shows the result of (f, g), using the relaxation method with occluding boundary and the smoothness constraint. The surface gradient was obtained after 60 repetitions or more. Furthermore, (f, g) was transformed into the surface gradient (p, q), and the depth map sequence was recovered. The contour map and the depth map of the egg are shown in Figure 6.

*The camera is 1/3 inch CCD camera with 768 × 494 pixels.

*In this experiment, since the motion of an object was restricted in the direction of a horizontal axis of the image, the aspect ratio of the camera was not taken into consideration.

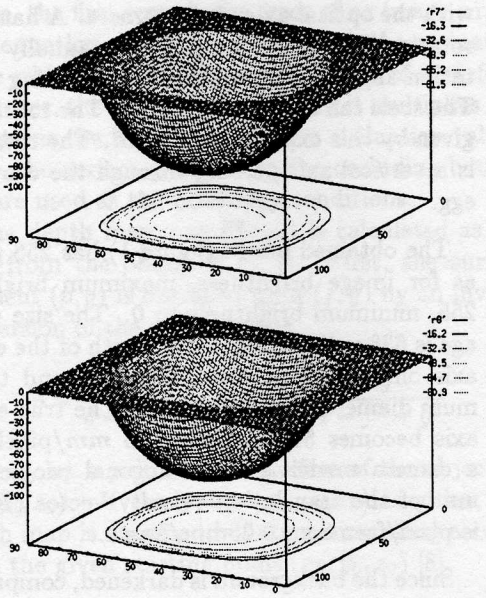

Figure 6: The depth maps of the egg in the first (upper) and the second (down) frames.

The motion parameters are obtained from the depth map sequence. The image of the egg is located near the center of a projection screen, therefore the center of the egg is considered to be $(0, 0, 0)$. Since the motion of the egg is the rotation of the y axis in the scene coordinate, the motion parameter to be estimated is only ω_y. However, since it shifted along with distance, the motion parameter of the optical axis direction T_z was also estimated. T_z is the gap of a standard position and it is not the true motion of the object. Since the true translation displacement parameter of the optical axis considers the camera model as an orthogonal projection, it cannot be measured by our method. The estimated motion parameters are shown in Table 1. The unit of T_z is $pixel/frame$.

Table 1: Estimation of motion parameters.

	T_z	ω_y
Real value	0.00000	3.00000°
Estimation	-0.00967	2.91542°

The displacement vectors field obtained the displacement parameters are pictured on Figure 7. The displacement vector is not obtained near the occluding boundary, since the depth map could not be correctly recovered near the occluding boundary.

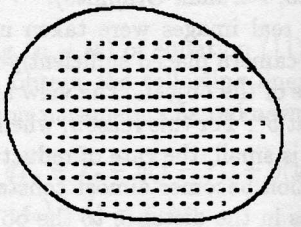

Figure 7: The displacement vectors field.

6 Conclusion

In the existing method for motion analysis, the position of the feature point on the object and the image brightness invariance, are assumed often. However, these assumptions may not be correct.

In this paper, we proposed a novel method having used constancy of shape as the assumption for estimating the 3D motion parameters of a rigid object. That is, the depth map was first estimated from shading information on the image at each time, then the 3D motion parameters were estimated from the obtained depth map sequence, and finally, the optical flow was determined. The accuracy of the proposed method was proved by the experiment with the real image.

If an object has a smooth and Lambertian surface, the proposed method is usable. The result of the proposed method will be applicable as a vision system of a snowplow.

By introduction of the more robust shape from shading algorithm, the performance of the proposed method should be improved. The next study will discuss parallelization of the processing for the real-time processing.

References

[1] A. Verri, T. Poggio: Against Quantitative Optical Flow, In Proc. First Int'l Conf. Computer Vision, London, p. 171-80, 1987

[2] N. Cornelius, T. Kanade: Adapting Optical Flow to Measure Object Motion in Reflectance

and X-ray Image Sequences, in Proceedings of ACM SIGGRAPH/SIGART Interdisciplinary Workshop on Motion, Toronto, Canada, 1933

[3] A. Nomura, H. Miike, K. Koga: Detection of Velocity Field from Varying Illumination, Trans. Institute of Electronics and Communication Engineers of Japan, J76-D-II, 9, p. 1977-86, 1993

[4] N. Mukawa: Estimation of Shape, Reflection Coefficients and Illuminant Direction from Image Sequences, Proc. 3rd ICCV, Osaka, p. 507-12, 1990

[5] B. K. P. Horn, B. G. Schunck: Determining Optical Flow, Artificial Intelligence, 17, p. 185, 1981

[6] M. Yamamoto, P. Boulanger, J.-A. Beraldin, M. Rioux, J. Domey: Direct Estimation of Range Flow on Deformable Shape from a Video Rate Range Camera , IEEE Trans. PAMI, Vol. 15, No. 1, p. 82-9, 1993

[7] B. K. P. Horn: Obtaining Shape from Shading, In Winston, P. H. ed., The Psychology of Computer Vision, McGraw-Hill, New York, 1975

[8] B. K. P. Horn: Robot Vision, MIT Press, Cambridge, 1986

[9] K. Ikeuchi: Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method , IEEE Trans. PAMI, Vol. 3, No. 6, p. 661-9, 1981