Differential-Formed Shape from Multiple Images Based on One-Directional Random Small Camera Rotations

Shoei Koizumi Norio Tagawa
Graduate School of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino-shi, Tokyo 191-0065, Japan
e-mail tagawa@sd.tmu.ac.jp

Abstract

The small vibration of the eye ball, which occurs when we fix our gaze on an object, is called "fixational eye movement," and using the analogy of it for the camera motion, differential-formed and integral-formed shape recovery method was proposed. We are considering a practical system using both methods selectively and adaptively for the local texture pattern in images. In this study, we analyze the performance especially of the differential-formed method with respect to the relation between a striped-texture and a camera rotation direction. From the results, we argue that, at least for the differential-formed method, suitable one-directional camera rotations have to be applied adaptively according to the local direction of image texture.

1 Introduction

It is well known that fixational eye movements, which means an irregular involuntary motion of eyeball, arises when human gazes fixed target [1], and an actual vision system based on fixational eye movements has been proposed as the Dymanuc Retina [2] and the Resonant Retina [3]. Fixational eye movements can be interpret as an instance of stochastic resonance [4]. Shape from motion methods using random camera rotations imitating fixational eye movements of a human visual system have been proposed as both of a differential scheme [5] using the gradient equation related to the optical flow and an integral scheme [6] using the motion blur in images. The differential-formed method is effective for small camera rotations, i.e. small image motions relative to the fineness of an image texture, and for the opposite case, the integral-formed method is suitable. In [6], an analog motion blur, i.e. the motion blur caused by the random camera rotations during exposure is supposed, but when both schemes are applied complementarily, the motion blur image has to be artificially generated by averaging many still images instead of analogously blurring. Therefore, for on-line shape recovery, it is desirable that the more accurate shape can be recovered using the small number of still images for both of differential and integral methods.

Both of those methods use the random camera rotations around $X-$ and $Y-$directions, which cause two-dimensional (2-D) motion field in image sequence. This 2-D motion field are effective for the image texture having various directions uniformly, but usually there is a biased distribution of a texture direction in images. On the other hand, when the object surface has a one-directional texture, for example the stripe pattern, there may be a possibility that 1-D movements are suitable for accurate depth recovery. In addition, 1-D movements spanning to the image sequence is desirable also for computation costs. The integral method in [6] needs convolution processing to detect the point spread function of motion blur. One-directional rotation causing 1-D image movements which direction depends on an image position and the depth corresponding to this position is especially desirable in the viewpoint of computation cost, since the point spread function becomes 1-D function. Therefore, we can adopt the camera's one-directional rotation and control the rotation direction adaptively according to the texture characteristics of region of interesting in the practical system.

From the above consideration, in this study, we examine the performance of one-directional camera rotations especially for the differential method in [5]. By evaluating the depth recovery error caused by applying the differential method with one-directional camera rotations to the images having stripe pattern, we explore the possibility of the practical system in which the suitable one-directional camera rotations instead of two-directional one are applied adaptively according to the local direction of the texture.

2 Outline of Differential Method using Random Small Camera Rotations

2.1 Camera Rotation and Gradient Equation

We use perspective projection as our camera-imaging model. The camera is fixed with an (X, Y, Z) coordinate system, where the viewpoint, i.e., a lens center, is at origin O and the optical axis is along the Z-axis. By taking a focal length as a unit of geometrical representation, a projection plane, i.e. an image plane, $Z = 1$ can be used without any loss of generality. A space point (X, Y, Z) on the object is projected to the image point (x, y). In general shape from motion methods, the camera is supposed to move with translational and rotational vectors relative to the object.

The motion model used in [5] and [6] represents tremor which is the smallest component of fixational eye movements. We can set a camera's rotation center at the back of the lens center with Z_0 along the optical axis. The rotation around $Z-$axis cannot generate new information of a 3-D scene, hence we consider the rotational velocity $\vec{r} = [r_x, r_y, 0]^\top$, which can be used also for the representation of the rotational vector at origin O. On the other hand, the translational vector $\vec{u}$ is caused by the above rotation, and is formulated as follows:

$$\vec{u} = \vec{r} \times \begin{bmatrix} 0 \\ 0 \\ Z_0 \end{bmatrix} = Z_0 \begin{bmatrix} r_y \\ -r_x \\ 0 \end{bmatrix}. \quad (1)$$

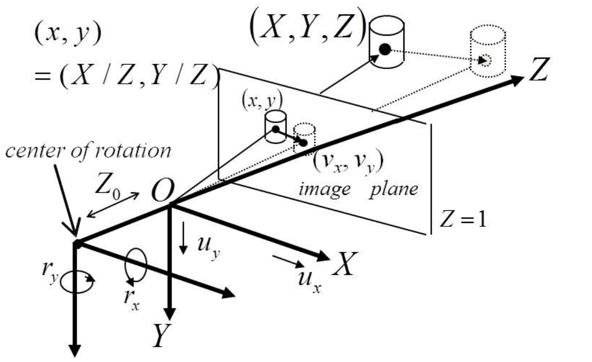


Figure 1. Coordinate system and camera motion model.

Using this representation of $\vec{u}$ and the inverse depth $d(x, y) = 1/Z(x, y)$, the optical flow $\vec{v} = [v_x, v_y]^\top$ is given as follows:

$$v_x = xyr_x - (1 + x^2)r_y - Z_0 r_y d \equiv v_x^r - r_y Z_0 d, \quad (2)$$

$$v_y = (1 + y^2)r_x - xyr_y + Z_0 r_x d \equiv v_y^r + r_x Z_0 d. \quad (3)$$

In the above equtions, d is an unknown variable at each pixel, and $\vec{u}$ and $\vec{r}$ are unknown common parameters for the whole image.

The gradient equation is the first approximation of the assumption that image brightness is invariable before and after the relative 3-D motion between a camera and an object. At each pixel (x, y), the gradient equation is formulated with the partial differentials f_x, f_y and f_t of the image brightness $f(x, y, t)$, where t denotes time, and the optical flow as follows:

$$f_t = -f_x v_x - f_y v_y. \quad (4)$$

By substituting Eqs. 2 and 3 into Eq. 4, the gradient equation representing a rigid motion constraint can be derived explicitly as follows:

$$\begin{aligned} f_t &= -(f_x v_x^r + f_y v_y^r) - (-f_x r_y + f_y r_x) Z_0 d \\ &\equiv -f^r - f^u d. \end{aligned} \quad (5)$$

In Eq. 5, f_x, f_y and f_t are observations and contain observation noise. Additionally, equation error, i.e. error caused by the first approximation in Eq. 4 generally exists. The coordinate system and the camera motion model is shown in Fig. 1.

2.2 Probabilistic Model Definition

We use M as the number of pairs of two successive frames and N as the number of pixels. In our study, $\{f_t^{(i,j)}\}_{i=1,\cdots,N;j=1,\cdots,M}$ and $\{\vec{r}^{(j)}\}_{j=1,\cdots,M}$ are treated as stochastic variables, and $\{d^{(i)}\}_{i=1,\cdots,N}$ corresponding to the inverse depth at each pixel is treated as a definite unknown variable.

It is supposed that optical flow is very small, and hence, observation errors of f_t, f_x and f_y, which are calculated by finite difference, are small. Additionally, equation error is also small, and therefore we can assume that error having no relation with f_t, f_x and f_y is added to the whole gradient equation. From this consideration, we assume that $f_t^{(i,j)}$ is a Gaussian random variable with mean 0 and variance σ_o^2, and $f_x^{(i,j)}$ and $f_y^{(i,j)}$ have no error.

$$p(f_t^{(i,j)} | d^{(i)}, \vec{r}^{(j)}, \sigma_o^2) = \frac{1}{\sqrt{2\pi}\sigma_o} \times \exp\left\{ -\frac{\left(f_t^{(i,j)} + f^{r(i,j)} + f^{u(i,j)} d^{(i)}\right)^2}{2\sigma_o^2} \right\}. \quad (6)$$

On the other hand, we also assume that r_x and r_y are independent Gaussian random variables respectively with mean 0 and variances of σ_r^2.

$$p(\vec{r}^{(j)} | \sigma_r^2) = \frac{1}{(\sqrt{2\pi}\sigma_r)^2} \exp\left\{ -\frac{\vec{r}^{(j)\top}\vec{r}^{(j)}}{2\sigma_r^2} \right\}. \quad (7)$$

From both models, the joint distribution of $\{f_t^{(i,j)}\}$ and $\{\vec{r}^{(j)}\}$ is formulated as follows:

$$\begin{aligned} & p(\{f_t^{(i,j)}\}, \{\vec{r}^{(j)}\} | \Theta) \\ & = \prod_{i=1}^{N} \prod_{j=1}^{M} p(f_t^{(i,j)} | d^{(i)}, \vec{r}^{(j)}, \sigma_o^2) \prod_{j=1}^{M} p(\vec{r}^{(j)} | \sigma_r^2), \end{aligned} \quad (8)$$

where $\Theta = \{\{d^{(i)}\}, \sigma_o^2, \sigma_r^2\}$. Additionally, the posterior distribution of $\{\vec{r}^{(j)}\}$ is

$$p(\{\vec{r}^{(j)}\} | \{f_t^{(i,j)}\}, \Theta) = \frac{p(\{\vec{r}^{(j)}\}, \{f_t^{(i,j)}\} | \Theta)}{p(\{f_t^{(i,j)}\} | \Theta)}. \quad (9)$$

The specific descriptions of Eqs. 8 and 9 are omitted.

2.3 Computation Algorithm

In order to determine Θ as a maximum likelihood estimator and to determine $\{\vec{r}^{(j)}\}$ as a MAP estimator, we apply the EM algorithm [7] by treating $\{\{f_t^{(i,j)}\}, \{\vec{r}^{(j)}\}\}$ as a complete data and $\{\vec{r}^{(j)}\}$ as a missing data.

In the EM algorithm, the E step and the M step are mutually repeated until they converge. At first, in the E step, the conditional expectation of the log likelihood of a complete data with observing $\{f_t^{(i,j)}\}$, which is called Q function, is computed. In the Q function, the estimated value $\hat{\Theta}$ is used for the parameters values in the conditional distribution. In the M step, the Q function is maximized with respect to Θ. The concrete formulation of both step can be shown in [5]

3 Evaluation of One-Directional Rotation Performance

To examine the performance of one-directional camera rotation for the one-directional texture, we use stripe pattern shown in Fig. 2(b) with 128×128 pixels. Figure 2(a) is a true depth map used for generating Fig. 2(b). Figure 2(b) consists mainly of horizontal stripes. With $M = 10$ and $\sigma_r = 0.01$, which generates a suitable-sized image motion for the gradient method with this image pattern, the recovered depth maps are shown in Fig. 3. It is noted that when the image motion is too large, the integral-formed method becomes

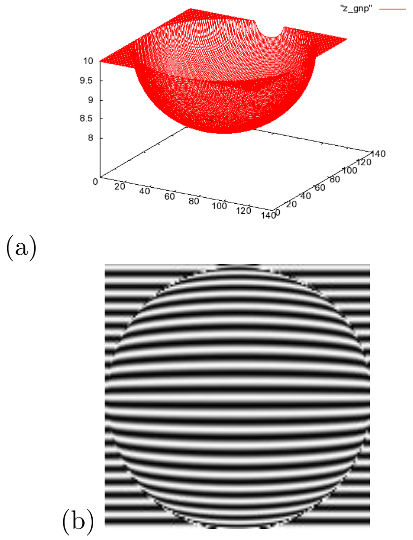


Figure 2. Example of data used in the evaluations: (a) true depth map; (b) artificial image used for making the successive images as an original image.

effective. The results with $M = 50$ and $M = 100$ are also shown in Figs. 4 and 5. From these results, the rotation around $X-$axis is apparently unsuitable for this strip pattern. On the other hand, the rotation around $Y-$axis is suitable more than 2-D rotation. The reason of this fact can be interpreted as follows. The gradient equation provides no constraints on the depth recovery for the image point where the spatial gradient of the image intensity along the optical flow takes zero. If the rotation around $X-$axis is only applied to the image shown in Fig. 2(b), the dominant component of the optical flow is v_y and there are many image points taking small value of f_y. Namely, at many image points Eq. 5 is useless for depth recovery.

From Eqs. 2 and 3, when $r_x = 0$, the optical flow is formulated as follows:

$$v_x = -(1 + x^2)r_y - Z_0 r_y d, \qquad v_y = -xyr_y. \tag{10}$$

From these equations, we can obtain that $v_x/v_y = (1 + x^2)/xy + Z_0 d/xy$. This means that the direction of the the optical flow depends on an image position and the depth corresponding to this position, but it is independent of r_y, i.e. the direction at each image position is constant during one-directional camera rotations. Additionally, we know that as the image position moves away from the center, the direction of the optical flow tends to be slant, although at the center region the direction of it is almost horizontal. In the system which we are going to develop in future, we will use the image motion parallel to the main direction of the texture in the local region for depth recovery. Therefore, to evaluate the effectiveness of the performance using the image motion parallel to the stripe as possible, we calculate the root mean square error (RMSE) by varying the calculation region size. This calculation region is defined at the center part of the image, since as the image position approaches to the center part, the image motion caused by the rotation around $Y-$axis becomes parallel to the texture direction. The evaluation results with $\sigma_r = 0.01$ are shown in Fig. 6. The horizontal axis of this figure indicates M. From Fig. 6, for the small calculation region, r_y is effective than the set of r_x and r_y. We confirmed also that the RMSE with r_x only takes approximately 1.00 independently of the size of the calculation region and M. In these simulations, there are no image noise, hence the bad influence caused by the image motion parallel to the texture direction was not revealed. We expect that if usual-leveled image noise is added, using the suitable one-directional rotation for the local texture direction is effective regardless of the number of the used images M.

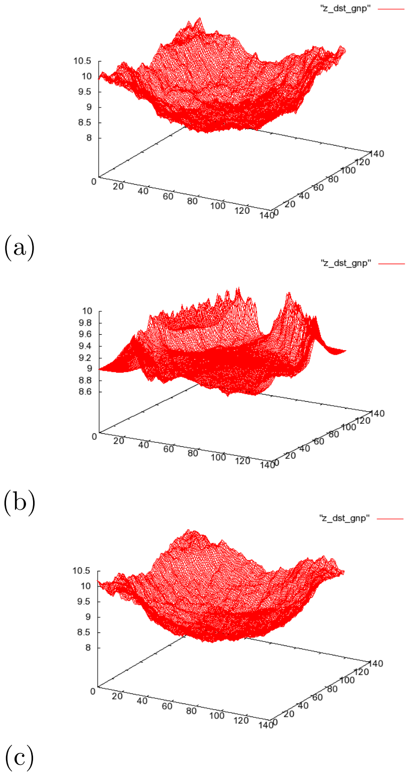
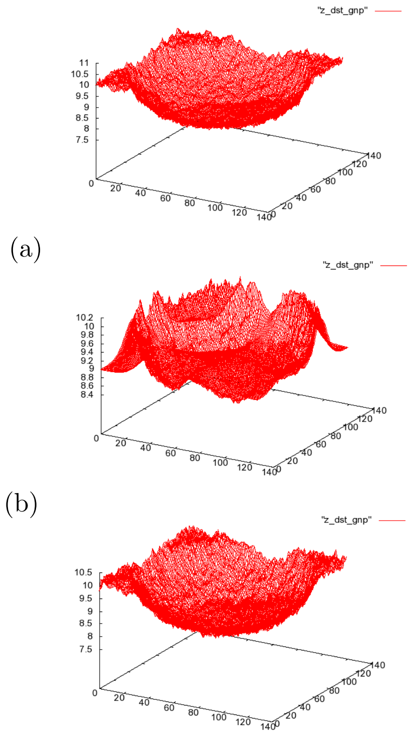


Figure 3. Recovered depth map with $M = 10$ and $\sigma_r = 0.01$: (a) r_x and r_y are used; (b) r_x is used only; (c) r_y is used only.

4 Conclusions

In this study, we make a hypothesis that in the differential method using random small camera rotations [5], there is a suitable camera rotation direction according to the direction characteristics in the image texture, and perform the numerical evaluations using the images having stripe pattern. As a result, even if there are no image noise, the image motion parallel to the texture direction is effective for an accurate depth recovery, and one-directional rotations can be actually adopted to realize such the situation.

The proposed one-dimensional rotations are effective furthermore for the integral-formed method [6], since significant reduction of computation costs can be expected in the integral-formed method. It is noted that for the integral-formed method the image motion perpendicular to the texture direction is expected to be suitable, since the motion blur is easy to occur by such the motion. Therefore, the similar examination will be carried out for the integral-formed method. Additionally, we are going to develop the practical system

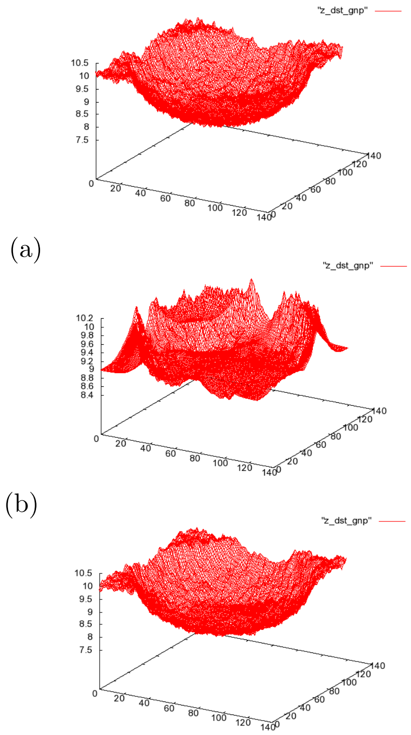


(a)

(b)

(c)

Figure 5. Recovered depth map with $M = 100$ and $\sigma_r = 0.01$: (a) r_x and r_y are used; (b) r_x is used only; (c) r_y is used only.



(a)

(b)

(c)

Figure 4. Recovered depth map with $M = 50$ and $\sigma_r = 0.01$: (a) r_x and r_y are used; (b) r_x is used only; (c) r_y is used only.

which partially recovers the depth in order of interesting region with adaptive one-directional camera rotations. In this system, the differential-formed and the integral-formed methods have to be selectivity used according to the fineness of the local texture. Hence, the decision rule has to be developed also in hurry.

References

[1] S. Martinez-Conde, S. L. Macknik, D. Hubel, "The Role of Fixational Eye Movements in Visual Perception," Nature Reviews, pp. 229–240, 2004.

[2] P. Propokopowicz, P. Cooper: "The Dynamic Retina," *Int'l J. Computer Vision*, vol. 16, pp. 191–204, 1995.

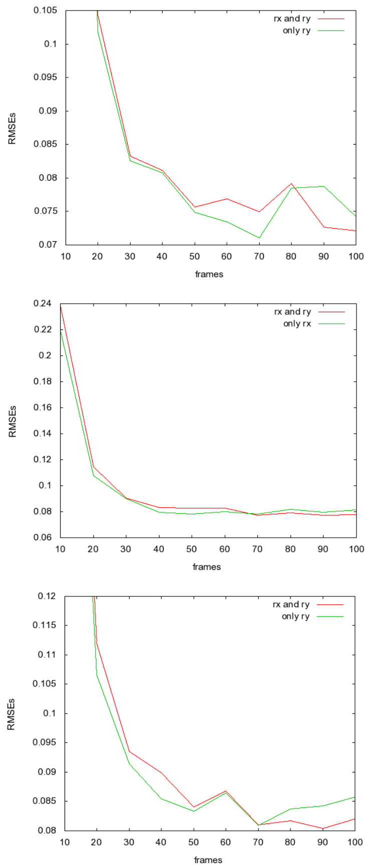
[3] M.-O. Hongler, Y. L. de Meneses, A. Beyeler, J. Jacot: "The Resonant Retina: Exploiting Vibration Noise to Optimally Detect Edges in an Image," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 25, pp. 1051–1062, 2003.

[4] L. Gammaitoni, P. Hanggi, P. Jung, F. Marchesoni: "Stochastic Resonance," *Rev. Modern Physics*, vol. 70, pp. 223-252, 1998.

[5] N. Tagawa: "Depth perception model based on fixational eye movements using Bayesian statistical inference," ICPR2010, pp. 1662–1665, 2010.

[6] N. Tagawa, Y. Iida, K. Okubo: "Depth Perception Model Exploiting Blurring Caused by Random Small Camera Motions," VISAPP2012, pp. 329–334, 2012.

[7] A. P. Dempster, N. M. Laird, D. B. Rubin: "Maximum Likelihood from Incomplete Data," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.



(a)

(b)

(c)

Figure 6. Recovered depth map with $M = 10$ and $\sigma_r = 0.01$: RMSE of recovered depth map: (a) calculation region is 12×12; (b) 36×36; (c) 60×60.