

# Selective Use of Optimal Image Resolution for Depth from Multiple Motions based on Gradient Scheme

Norio Tagawa and Shoei Koizumi

*Graduate School of System Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, Japan*

**Keywords:** Direct Shape Recovery from Motion, Gradient Method, Random Camera Motions, Fixational Eye Movements.

**Abstract:** The gradient-based depth from motion method is effective for obtaining a dense depth map. However, the accuracy of the depth map recovered only from two successive images is not so high, and hence, to increase the depth information by tracking corresponding image points through an image sequence is often performed by using, for example, the Kalman filter-like technique. Alternatively, multiple image pairs generated by random small camera rotations around a reference direction can be used for gaining much information of depth without such the tracking procedure. In the framework of this strategy, in this study, to further improve the accuracy, we propose a selective use of the optimal image resolution. The appropriate resolution image is required to have a linear intensity pattern which is the most important supposition for the gradient method often used for dense depth recovery based on the theory of “shape from motion.” The performance of our proposal is examined through numerical evaluations using artificial images.

## 1 INTRODUCTION

The gradient-based depth from motion methods have been vigorously studied to recover a dense depth map (Horn and Schunk, 1981), (Simoncelli, 1999), (Bruhn and Weickert, 2005), (Tagawa et al., 2008), (Brox and Malik, 2011), (Ochs and Brox, 2012). However, the accuracy of the depth map recovered from two successive images is not enough, and hence some methods track corresponding points in an image sequence to use multiple viewpoint. The accurate tracking is also difficult and the various techniques have been studied, for example, based on the Kalman filter (Paramanand and Rajagopalan, 2012) and the particle filter. We proposed a tracking method, too, which adopts the Bayesian label assignment instead of explicit tracking (Ikeda et al., 2009). If possible, the accurate depth recovery with no use of the tracking is desired.

The accuracy of the gradient method hardly depends on the equation error of the gradient equation. The gradient equation is a first order approximation of the intensity invariant constraint before and after the relative motion between a camera and an object, and in general the second and more higher order terms causes the equation error. The amount of the error depends on the relative relation between the size of the image motion called an optical flow and the spatial

frequency of an image intensity pattern. This means that the appropriate spatial frequency exists at each pixel respectively according to the size of the optical flow. Therefore, we can select the optimal image resolution including the effective frequency and use it for the gradient equation. However, if the images have little variations of the spatial frequency, the optimal frequency component will not necessarily be extracted at each pixel according to the specific optical flow determined by the depth at that pixel and the relative camera motion. To avoid the problem, we should analyze many intensity pairs, i.e., many optical flows for each 3-D point on a target object.

On the other hand, the depth recovery method using random camera rotations imitating fixational eye movements of a human’s eye ball (Martinez-Conde et al., 2004) has been proposed (Tagawa, 2010). In this method, since a camera is assumed to rotate randomly around the reference direction with a small rotation angle, the gradient method is applied simultaneously to a lot of image pairs without the image point tracking. In the usual framework of the gradient method, the optical flow is detected based on the gradient equation in the first step, and next, the depth map is recovered from the optical flow. This two step procedure is not suitable for expanding the gradient scheme for multiple image pairs, and the direct method is adopted in (Tagawa, 2010), in which

the depth map is directly recovered without the optical flow detection. In this study, we propose the selective use of the optimal image resolution in the framework of the method in (Tagawa, 2010)

In the following, the outline of the method in (Tagawa, 2010) is explained in Sec. 2, and the proposed method for the optimal resolution selection and its effectiveness confirmed by numerical evaluations are referred in Sec. 3. We show the conclusions of this study in Sec. 4.

## 2 OUTLINE OF DEPTH FROM MULTIPLE IMAGE PAIRS

### 2.1 Camera Motions Imitating Tremor

The camera coordinate system and the camera motion model imitating tremor, which is one of the fixational eye movements and is smallest one (Martinez-Conde et al., 2004), in this study are the same as those in (Tagawa, 2010) and are shown in Fig. 1. We use a perspective projection system as our camera-imaging model. A space point  $(X, Y, Z)^\top$  on an object is projected to an image point  $\vec{x} \equiv (x, y, 1)^\top = (X/Z, Y/Z, 1)^\top$ .

On the analogy of a human eyeball, we can set a camera's rotation center at the back of a lens center with  $Z_0$  along an optical axis, and we assume that there is no explicit translational motions of a camera. This rotation with the rotational vector  $\vec{r} = (r_X, r_Y, r_Z)^\top$  can also be represented using the coordinate origin as its rotation center with the same rotational vector  $\vec{r}$ . On the other hand, this difference between the origin and the rotation center causes a translational vector  $\vec{u} = (u_X, u_Y, u_Z)^\top$  implicitly, and is formulated as follows:

$$\begin{bmatrix} u_X \\ u_Y \\ u_Z \end{bmatrix} = \begin{bmatrix} r_X \\ r_Y \\ r_Z \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ Z_0 \end{bmatrix} = Z_0 \begin{bmatrix} r_Y \\ -r_X \\ 0 \end{bmatrix}. \quad (1)$$

With this system,  $Z_0$  can be simply known beforehand, hence an absolute depth can be recovered, although a general camera motion enables us to get only relative depth.

From Eq. 1, it can be known that  $r_Z$  causes no translations. Therefore, we set  $r_Z = 0$  and define  $\vec{r} = (r_X, r_Y, 0)^\top$  as a rotational vector like an eyeball.  $\vec{r}(t)$  can be treated as a stochastic white Gaussian process, in which  $t$  indicates time measured from a reference time. The fluctuation of  $\vec{r}(t)$  at each time is assumed to be a two-dimensional Gaussian distribution with a mean 0 and a variance  $\sigma_r^2$ , where  $\sigma_r^2$  is

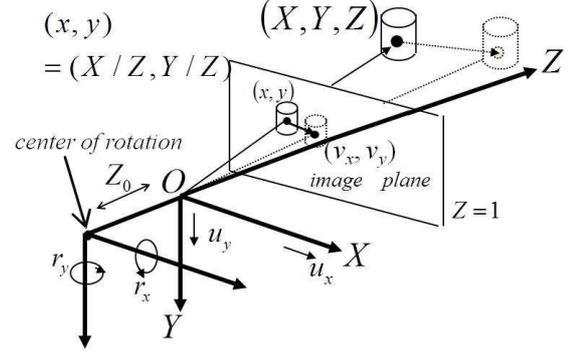


Figure 1: Coordinate system and camera motion model used in this study.

assumed to be known.

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

In the above description, we define  $\vec{r}$  as a rotational velocity to make a theoretical analysis simple. For small values of the actual rotation angle, Eq. 1 and the other equations below approximate finite camera motions.

### 2.2 Depth from Multiple Image Pairs based on Gradient Method

Using the camera motion model described above and the inverse depth  $d(x, y) \equiv 1/Z(x, y)$ , the optical flow  $\vec{v} \equiv [v_x, v_y]^\top$  is formulated as follows:

$$v_x = xy r_x - (1 + x^2)r_y + yr_z - Z_0 r_y d \equiv v_x^r - r_y Z_0 d, \quad (3)$$

$$v_y = (1 + y^2)r_x - xy r_y - xr_z + Z_0 r_x d \equiv v_y^r + r_x Z_0 d. \quad (4)$$

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

At each pixel position  $(x, y)$ , the gradient equation is formulated with the partial differentials  $f_x$ ,  $f_y$  and  $f_t$  of the image brightness  $f(x, y, t)$  and the optical flow as follows (Horn and Schunk, 1981):

$$f_t = -f_x v_x - f_y v_y, \quad (5)$$

where  $t$  denotes time. By substituting Eqs. 3 and 4 into Eq. 5, the gradient equation representing a rigid motion constraint can be derived explicitly

$$\begin{aligned} f_t &= -(f_x v_x^r + f_y v_y^r) - (-f_x r_y + f_y r_x) Z_0 d \\ &\equiv -f^r - f^u d. \end{aligned} \quad (6)$$

$M$  is the number of pairs of two successive frames and  $N$  is the number of pixels. We assume that  $f_t^{(i,j)}$

has a Gaussian random error corresponding to the equation error, and  $f_x^{(i,j)}$  and  $f_y^{(i,j)}$  have no error.

$$p(f_t^{(i,j)} | d^{(i)}, \vec{r}^{(j)}, \sigma_o^2) = \frac{1}{\sqrt{2\pi}\sigma_o} \times \exp \left\{ -\frac{\left( f_t^{(i,j)} + f_r^{(i,j)} + f_u^{(i,j)} d^{(i)} \right)^2}{2\sigma_o^2} \right\}, \quad (7)$$

where  $i = 1, \dots, N$  and  $j = 1, \dots, M$ , and  $\sigma_o^2$  is an unknown variance.

Since multiple frames vibrated by irregular rotations  $\{\vec{r}^{(j)}\}$  are used for depth recovery without tracking of the corresponding points in the images, the recovered  $d^{(i)}$  at each pixel takes an average value of the neighboring region defined by vibration width in image. Therefore,  $\{d^{(i)}\}$  should be assumed to have local correlation in the image. In this study, to simplify the stochastic modeling of  $\{d^{(i)}\}$ , we adopt the following equation as the depth model.

$$p(\vec{d} | \sigma_d^2) = \frac{1}{(\sqrt{2\pi}\sigma_d)^N} \exp \left\{ -\frac{\vec{d}^\top \vec{L} \vec{d}}{2\sigma_d^2} \right\}, \quad (8)$$

where  $\vec{d}$  is a  $N$ -dimensional vector composed of  $\{d^{(i)}\}$  and  $\vec{L}$  indicates the matrix corresponding to the 2-dimensional Laplacian operator. By assuming this probabilistic density, we make a recovered depth map smooth. The use of the prior distribution of the values to be estimated is interpreted as a regularization scheme in the signal processing viewpoint (Poggio et al., 1985). In this study, the variance  $\sigma_d^2$  is controlled heuristically in consideration of smoothness of a recovered depth map. Hereafter, we use the definition  $\Theta \equiv \{\sigma_o^2, \sigma_r^2\}$ .

Based on the probabilistic models of  $\vec{r}^{(j)}$ ,  $f_t^{(i,j)}$  and  $d^{(i)}$  defined above, we can statistically estimate the depth map. By applying the MAP-EM algorithm (Dempster et al., 1977),  $\{\vec{d}, \Theta\}$  can be estimated as a MAP estimator based on  $p(\vec{d}, \Theta | \{f_t^{(i,j)}\})$ , which is formulated by marginalizing the joint probability  $p(\{\vec{r}^{(j)}\}, \vec{d}, \Theta | \{f_t^{(i,j)}\})$  with respect to  $\{\vec{r}^{(j)}\}$ , in which the prior of  $\Theta$  is formally regarded as a uniform distribution. The concrete formula of  $p(\vec{d}, \Theta | \{f_t^{(i,j)}\})$  is shown in Eq. 15 in the APPENDIX. Additionally,  $\{\vec{r}^{(j)}\}$  can be estimated as a MAP estimator based on  $p(\{\vec{r}^{(j)}\} | \{f_t^{(i,j)}\}, \hat{\Theta}, \hat{\vec{d}})$ , in which  $\hat{\cdot}$  means a MAP estimator. The concrete formula of it is also introduced in the APPENDIX by Eq. 17. From the formulations, the direct MAP estimation of  $\vec{d}$  is realized to be difficult, but the MAP-EM algorithm can solve it stably through an indirect iterative scheme each iteration of which consists of

the E-step and the M-step. In the concrete update procedure of  $\vec{d}$ , we use the One-Step-Late (OSL) technique (Green, 1990) to avoid complicated computation of  $\cdot$ . The details of the algorithm are shown in (Tagawa, 2010).

### 3 ACCURATE RECOVERY BY OUTLIER REDUCTION

#### 3.1 Selection of Optimal Resolution for Gradient Equation

The gradient equation in Eq. 5 is a linear approximation of the intensity invariant constraint before and after the relative camera motion. In general, there are the second and more higher order terms which are considered as the equation error included in the observation of  $f_t$ , which is defined as a simple difference between successive images like conventional many methods, and cause the recovery error of the depth map. The amount of these unwanted terms depend on the relative relation between the spatial frequency of the intensity pattern and the size of the optical flow at each pixel. Therefore, in this study, we try to improve the accuracy of the recovered depth by selecting and using the suitable spatial frequency component and discarding the other components as an outlying data for each image pair.

After multi-resolution decomposition shown in Fig. 2, the proposed strategy consists of two steps. In the first step, to detect the spatial frequency by which drastically large amount of the equation error is observed in the gradient equation. As one of the indices, we use the consistency of the spatial gradient vectors between two successive frames at each pixel,  $\vec{f}_s^{(i,j)} \equiv [f_x^{(i,j)}, f_y^{(i,j)}]^\top$  and  $\vec{f}_s^{(i,j-1)} \equiv [f_x^{(i,j-1)}, f_y^{(i,j-1)}]^\top$ . It should be noted that low resolution images are likely to generate small equation error as compared with high resolution images but high resolution images can be used to recover high resolution depth map. Hence, from low resolution to high resolution at each pixel, we search the resolution in which the directions of the spatial gradient between the successive frames are reverse using the sign of the inner product  $\vec{f}_s^{(i,j)\top} \vec{f}_s^{(i,j-1)}$ . The image components whose resolution is lower than the resolution in which the sign of  $\vec{f}_s^{(i,j)\top} \vec{f}_s^{(i,j-1)}$  is negative are selected as the candidates for the most appropriate resolution.

In the second step, the amount of the nonlinear terms included in the observation of  $f_t$  is estimated, and using it the most appropriate resolution is de-

tected and is used to recover the depth at each pixel.  $f_t$  is exactly represented as follows:

$$f_t = -f_x v_x - f_y v_y - \frac{1}{2} \{ f_{xx} v_x^2 + f_{yy} v_y^2 + 2f_{xy} v_x v_y \} + \dots \quad (9)$$

If the first step is performed well, the nonlinear term can be considered small, and in this case the second order term in Eq. 9 can be estimated at each pixel  $i$  as follows:

$$-\frac{1}{2} \{ (f_x^{(i,j)} - f_x^{(i,j-1)}) v_x^{(i,j)} + (f_y^{(i,j)} - f_y^{(i,j-1)}) v_y^{(i,j)} \}. \quad (10)$$

Spontaneously, we can define two measures for estimating the amount of the equation error.

$$J_1 = \frac{|(f_x^{(i,j)} - f_x^{(i,j-1)}) v_x^{(i,j)} + (f_y^{(i,j)} - f_y^{(i,j-1)}) v_y^{(i,j)}|}{2|f_x^{(i,j)} v_x^{(i,j)} + f_y^{(i,j)} v_y^{(i,j)}|}, \quad (11)$$

$$J_2 = \frac{|(f_x^{(i,j)} - f_x^{(i,j-1)}) v_x^{(i,j)} + (f_y^{(i,j)} - f_y^{(i,j-1)}) v_y^{(i,j)}|}{2\sqrt{f_x^{(i,j)2} + f_y^{(i,j)2}}}. \quad (12)$$

$J_1$  measures the nonlinearity as a ratio for the amount of the first term, which can be interpreted a signal-to-noise ratio. We can know the merit of  $J_1$  from Eq. 11 that  $J_1$  can be estimated using only the direction of the true optical flow, namely the amplitude of the optical flow is not required to be known.  $J_2$  measures the nonlinearity with the dimension of the optical flow, and this amount is proportional to the recovery error of  $d$ .

Additionally, to estimate the higher order terms including the second order term the following two measures can be used.

$$J_3 = \frac{|f_t^{(i,j)} - f_{t0}^{(i,j)}|}{|f_x^{(i,j)} v_x^{(i,j)} + f_y^{(i,j)} v_y^{(i,j)}|}, \quad (13)$$

$$J_4 = \frac{|f_t^{(i,j)} - f_{t0}^{(i,j)}|}{\sqrt{f_x^{(i,j)2} + f_y^{(i,j)2}}}, \quad (14)$$

where  $f_{t0}^{(i,j)}$  is a true value of  $f_t$ . For the candidate resolutions selected in the first step, the most appropriate resolution for depth recovery is determined by comparing the value of one selected from  $J_k$  ( $k = 1, 2, 3, 4$ ). It is noted that the exact values of these measures cannot be computed, since these include the variables to be determined. Therefore, only those estimates are provided.

### 3.2 Numerical Evaluation

To confirm the effectiveness of the proposed method, and especially compare the efficiency of  $J_k$  ( $k =$

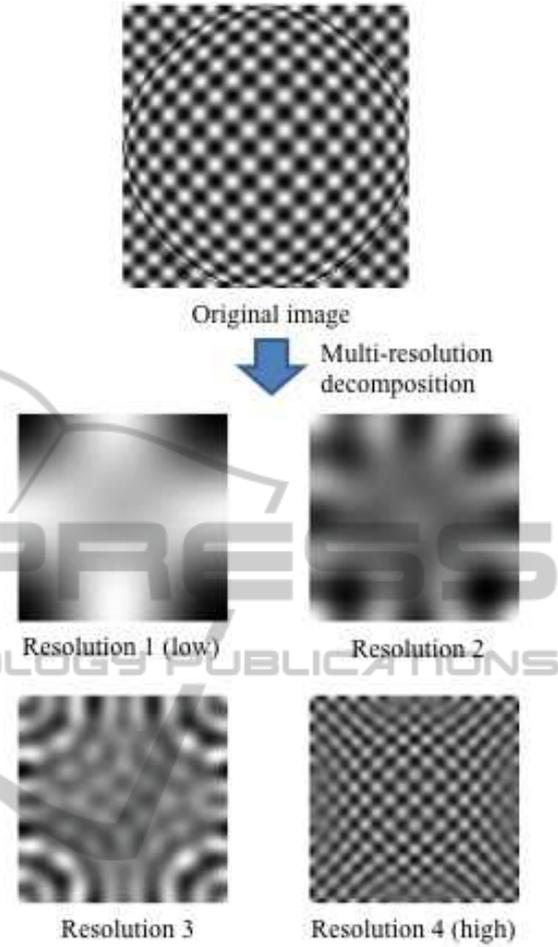


Figure 2: Example of multi-resolution decomposition.

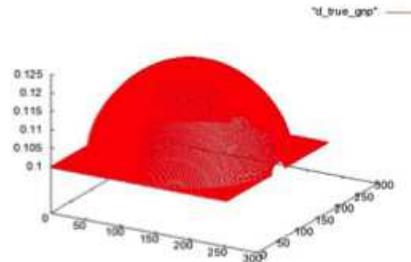


Figure 3: True depth map used for evaluation.

1, 2, 3, 4), we conducted numerical evaluations using artificial images. Figure 3 shows the true inverse depth map for evaluation. The vertical axis indicates the inverse depth  $d$  using the focal length as a unit, and the horizontal axes represent the pixel position in the image plane. The reference image generated by a computer graphics technique is shown in Fig. 2. The images viewed with random camera motions are generated using the reference image, the true depth map and the random camera rotations

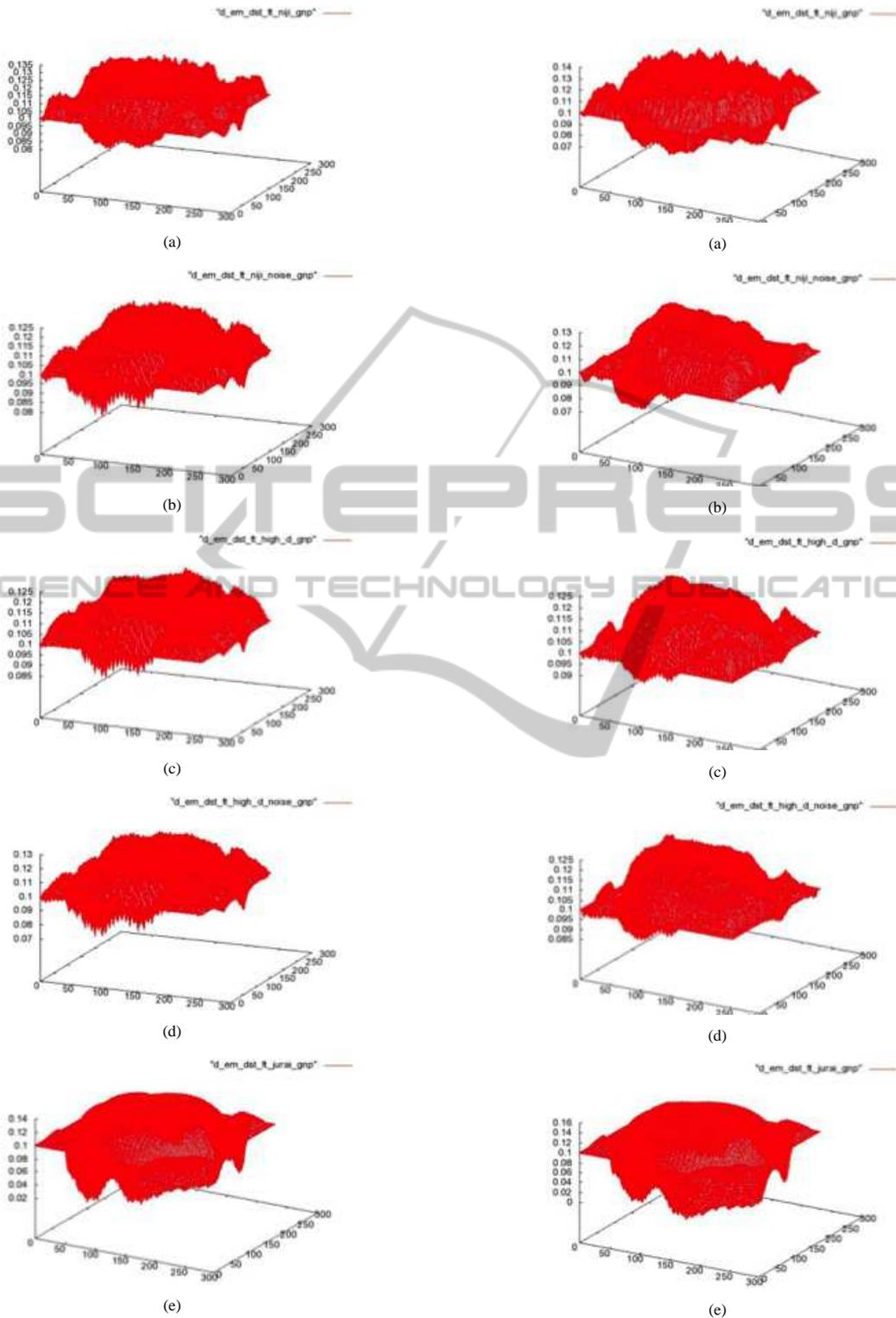


Figure 4: Recovered depth map with  $\sigma_r^2 = 0.006^2$ : appropriate resolution is selected using (a)  $J_1$ , (b)  $J_2$ , (c)  $J_3$ , (d)  $J_4$ , and (e) result by conventional method.

Figure 5: Recovered depth map with  $\sigma_r^2 = 0.008^2$ : appropriate resolution is selected using (a)  $J_1$ , (b)  $J_2$ , (c)  $J_3$ , (d)  $J_4$ , and (e) result by conventional method.

sampled as a Gaussian independent random variable defined in Eq. 2. The image size adopted in these evaluations is  $256 \times 256$  pixels, which corresponds to  $-0.5 \leq x, y \leq 0.5$  measured using the focal length as a unit. All images were decomposed into four resolution layers shown in Fig. 2. We recovered the depth map from a image set consisting of 100 images each having random movements. Image pairs consist of each image and the reference image were used to compute  $f_x$ ,  $f_y$  and  $f_t$  in the gradient equation. The most appropriate resolution was determined by the method explained in the above section at each pixel in each image pair, and was used as an observation for the MAP-EM algorithm. A plane  $Z = 9$  was used as an initial value of the depth map. The parameter  $\sigma_d^2$  determining the degree of smoothness of the recovered depth was fixed as  $\sigma_d^2 = 1.0 \times 10^{-4}$ , and the accuracy of the recovered depth map was evaluated by varying  $\sigma_r^2$  determining the amplitude of the random camera rotations.

In this evaluation, firstly we use the true value of the optical flow for computing  $J_k$  ( $k = 1, 2, 3, 4$ ) to confirm the essential effectiveness of the selective use of the spatial frequency. Figure 4 shows the recovered inverse depth map for  $\sigma_r^2 = 0.006^2$ . Likewise, Fig. 5 shows the results for  $\sigma_r^2 = 0.008^2$ , i.e. the results for the large motions compared with Fig. 4. In the figures, the conventional method (Tagawa, 2010) indicates that the image intensity is used as is, namely selective use of the optimal resolution is not performed. Figure 6 represents the relation between the depth recovery error and the amplitude of the random camera rotations. From both of the root mean square error (RMSE) and the relative error, we can confirm that the proposed outlier reduction is in effect especially for the large camera rotations. For relatively small camera rotations, the proposed all measures can be used to improve the accuracy of In the four measures,  $J_1$  shows good performance regardless of size of the camera motion.

Successively, we evaluate the actual case in which the true optical flow is unknown. At each pixel in each image pair, the minimum least square solution of the optical flow is derived using the candidate resolutions selected by the first step in the proposed method, and is used for compute  $J_k$  ( $k = 1, 2, 3, 4$ ). The relation between the depth recovery error and the amplitude of the random camera rotations are shown in Fig. 7. As a result, the performance of  $J_1$  was still in effect, which is expected to be caused by the fact that  $J_1$  depends only on the direction of the optical flow.

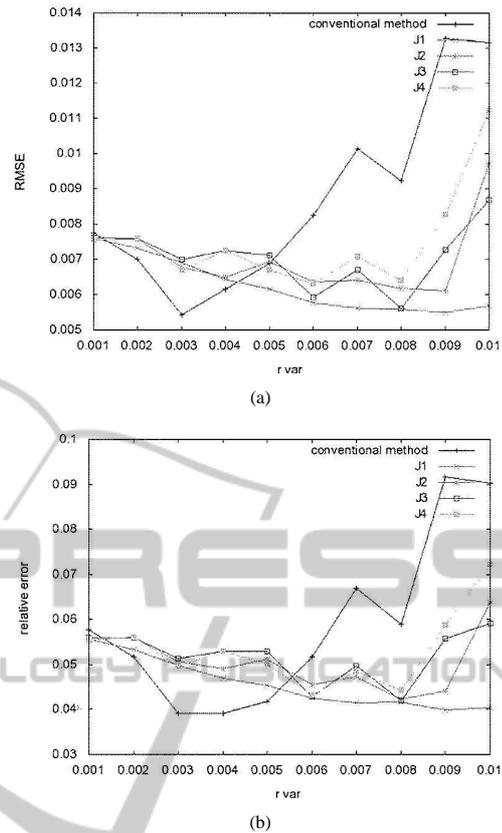


Figure 6: Recovery error vs. amplitude of camera rotation evaluated using true values of optical flow: (a) RMSEs, (b) relative errors.

## 4 CONCLUSIONS

In this study, we examine the method to improve the accuracy of the depth recovery based on the relative camera motions. We focused on the linear approximation error of the gradient equation, and proposed the selective use of the optimal image resolution. We defined different four measures to estimate the approximation error, and the effectiveness was confirmed by integrating the proposed selection method into the conventional depth recovery method using random camera rotations. In the future, real image experiments have to be carried out to indicate the actual performance of our method.

We are now proceeding with the further examination to improve the accuracy of the recovered depth, in which we adopt Eq. 10 for the representation of the second order term included in the observed  $f_t$  for the resolution selected by the measure  $J_1$ . By this representation, the gradient equation corresponding to the second order approximation of the intensity invariant constraint can be formulated as a linear equation

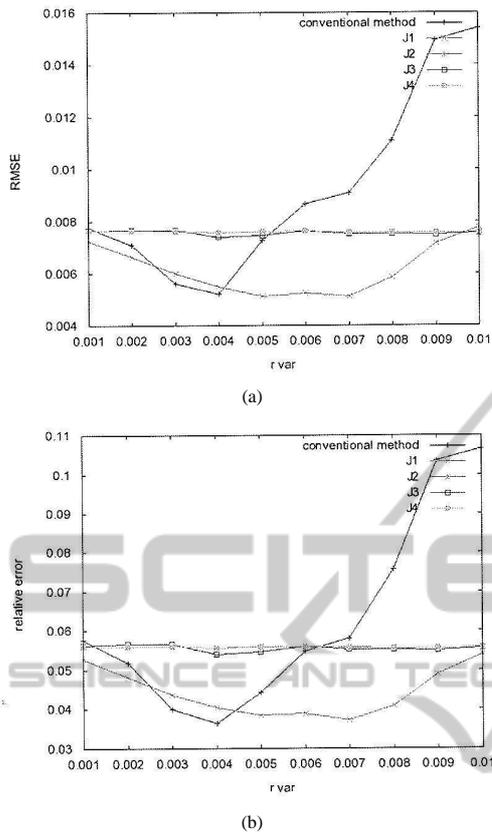


Figure 7: Recovery error vs. amplitude of camera rotation using estimates of optical flow: (a) RMSEs, (b) relative errors.

about an optical flow, which is easy to be handled. It is noted that this formulation is nothing but redefining a spatial gradient of intensity patterns using the intensity of two successive images.

In the framework using random camera rotations, the integral-formed method based on motion blur has also proposed. This method is effective for the large image motions compared with the image intensity patterns, and hence, it can be used for the case in which the differential-formed method in this study is not suitable (Tagawa et al., 2012), (Tagawa et al., 2013). Unification of both schemes is a future problem, too.

## REFERENCES

Brox, T. and Malik, J. (2011). Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Trans. Pattern Anal. Machine Intell.*, 33(3):500–513.

Bruhn, A. and Weickert, J. (2005). Locas/kanade meets horn/schunck: combining local and global optic flow methods. *Int. J. Comput. Vision*, 61(3):211–231.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data. *J. Roy. Statist. Soc. B*, 39:1–38.

Green, P. J. (1990). On use of the em algorithm for penalized likelihood estimation. *J. Roy. Statist. Soc. B*, 52:443–452.

Horn, B. P. and Schunk, B. (1981). Determining optical flow. *Artif. Intell.*, 17:185–203.

Ikeda, M., Okubo, K., and Tagawa, N. (2009). Implicit tracking of multiple objects based on bayesian region label assignment. In *proc. VISAPP2009*, pages 503–506.

Martinez-Conde, S., Macknik, S. L., and Hubel, D. (2004). The role of fixational eye movements in visual perception. *Nature Reviews*, 5:229–240.

Ochs, P. and Brox, T. (2012). Higher order motion models and spectral clustering. In *proc. CVPR2012*, pages 614–621.

Paramanand, C. and Rajagopalan, A. N. (2012). Depth from motion and optical blur with unscented kalman filter. *IEEE Trans. Image Processing*, 21(5):2798–2811.

Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.

Simoncelli, E. P. (1999). Bayesian multi-scale differential optical flow. In *Handbook of Computer Vision and Applications*, pages 397–422. Academic Press.

Tagawa, N. (2010). Depth perception model based on fixational eye movements using bayesian statistical inference. In *proc. ICPR2010*, pages 1662–1665.

Tagawa, N., Iida, Y., and Okubo, K. (2012). Depth perception model exploiting blurring caused by random small camera motions. In *proc. Int. Conf. on Computer Vision, Theory and Applications*, pages 329–334.

Tagawa, N., Kawaguchi, J., Naganuma, S., and Okubo, K. (2008). Direct 3-d shape recovery from image sequence based on multi-scale bayesian network. In *proc. ICPR08*, page CD.

Tagawa, N., Koizumi, S., and Okubo, K. (2013). Direct depth recovery from motion blur caused by random camera rotations imitating fixational eye movements. In *proc. Int. Conf. on Computer Vision, Theory and Applications*, pages 177–186.

## APPENDIX

The posterior distribution of  $\vec{d}$  and  $\Theta$  is derived using the Bayes' theorem and the uniform prior probability  $p(\Theta) = \text{Const.}$  as follows:

$$\begin{aligned}
 p(\vec{d}, \Theta | \{f_t^{(i,j)}\}) &= \frac{p(\vec{d}, \Theta, \{f_t^{(i,j)}\})}{p(\{f_t^{(i,j)}\})} \\
 &\propto \int \cdots \int p(\{f_t^{(i,j)}\}, \{\vec{r}^{(j)}\}, \vec{d}, \Theta) d\{\vec{r}^{(j)}\} \\
 &= \int \cdots \int p(\{f_t^{(i,j)}\} | \{\vec{r}^{(j)}\}, \vec{d}, \sigma_o^2) p(\{\vec{r}^{(j)}\} | \sigma_r^2) p(\vec{d} | \sigma_d^2) \\
 &\quad \times p(\Theta) d\{\vec{r}^{(j)}\} \\
 &\propto \int \cdots \int \prod_{i=1}^N \prod_{j=1}^M p(f_t^{(i,j)} | d^{(i)}, \vec{r}^{(j)}, \sigma_o^2) \prod_{j=1}^M p(\vec{r}^{(j)} | \sigma_r^2) \\
 &\quad \times p(\vec{d} | \sigma_d^2) d\{\vec{r}^{(j)}\} \\
 &= \frac{1}{(2\pi)^{N(M+1)/2+M} \sigma_o^{MN} \sigma_r^{2M} \sigma_d^N} \\
 &\quad \times \int \cdots \int \exp \left\{ -\frac{\sum_i^N \sum_j^M (f_t^{(i,j)} + \vec{w}^{(i,j)\top} \vec{r}^{(j)})^2}{2\sigma_o^2} \right. \\
 &\quad \left. - \frac{\sum_{j=1}^M \vec{r}^{(j)\top} \vec{r}^{(j)}}{2\sigma_r^2} \right\} d\{\vec{r}^{(j)}\} \exp \left\{ -\frac{\vec{d}^\top \vec{L} \vec{d}}{2\sigma_d^2} \right\}, \quad (15)
 \end{aligned}$$

$$\begin{aligned}
 \vec{w}^{(i,j)} &\equiv \begin{pmatrix} f_x^{(i,j)} x^{(i)} y^{(i)} + f_y^{(i,j)} (1 + y^{(i)2}) \\ -f_x^{(i,j)} (1 + x^{(i)2}) - f_y^{(i,j)} x^{(i)} y^{(i)} \\ f_x^{(i,j)} y^{(i)} - f_y^{(i,j)} x^{(i)} \end{pmatrix} \\
 &\quad + Z_0 d^{(i)} \begin{pmatrix} f_y^{(i,j)} \\ -f_x^{(i,j)} \\ 0 \end{pmatrix}. \quad (16)
 \end{aligned}$$

The posterior distribution of  $\{\vec{r}^{(j)}\}$  is also derived as follows:

$$\begin{aligned}
 p(\{\vec{r}^{(j)}\} | \{f_t^{(i,j)}\}, \Theta, \vec{d}) &= \frac{p(\{\vec{r}^{(j)}\}, \{f_t^{(i,j)}\} | \Theta, \vec{d})}{p(\{f_t^{(i,j)}\} | \Theta, \vec{d})} \\
 &\propto p(\{f_t^{(i,j)}\} | \{\vec{r}^{(j)}\}, \vec{d}, \sigma_o^2) p(\{\vec{r}^{(j)}\} | \sigma_r^2) \\
 &= \frac{1}{\sqrt{(2\pi)^{2M} \prod_i^M \det \vec{V}_r^{(j)}}} \\
 &\quad \times \exp \left\{ -\frac{1}{2} \sum_{j=1}^M (\vec{r}^{(j)} - \vec{r}_m^{(j)})^\top \vec{V}_r^{(j)-1} (\vec{r}^{(j)} - \vec{r}_m^{(j)}) \right\}, \quad (17)
 \end{aligned}$$

$$\vec{r}_m^{(j)} = -\frac{1}{\sigma_o^2} \vec{V}_r^{(j)} \sum_{i=1}^N f_t^{(i,j)} \vec{w}^{(i,j)}, \quad (18)$$

$$\vec{V}_r^{(j)} = \left( \frac{1}{\sigma_o^2} \sum_{i=1}^N \vec{w}^{(i,j)} \vec{w}^{(i,j)\top} + \frac{1}{\sigma_r^2} \vec{I} \right)^{-1}. \quad (19)$$