

# Direct Depth Recovery from Motion Blur Caused by Random Camera Rotations Imitating Fixational Eye Movements

Norio Tagawa, Shoei Koizumi and Kan Okubo

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

**Keywords:** Shape from Motion Blur, Random Camera Rotation, Fixational Eye Movement, Integral Formed Method.

**Abstract:** It has been reported that small involuntary vibrations of a human eyeball for fixation called "fixational eye movements" play a role of image analysis, for example contrast enhancement and edge detection. This mechanism can be interpreted as an instance of stochastic resonance, which is inspired by biology, more specifically by neuron dynamics. A depth recovery method has been proposed, which uses many successive image pairs generated by random camera rotations imitating fixational eye movements. This method, however, is not adequate for images having fine texture details because of an aliasing problem. To overcome this problem, we propose a new integral formed method for recovering depth, which uses motion blur caused by the same camera motions, i.e. many random small camera rotations. As an algorithm, we examine a method directly recovering depth without computing a blur function. To confirm the feasibility of our scheme, we perform simulations using artificial images.

## 1 INTRODUCTION

Camera vibration noise is a serious concern for a hand-held camera and for many vision systems mounted on mobile platforms such as planes, cars or mobile robots, and of course for biological vision systems. The computer vision researchers traditionally considered the camera vibration as a mere nuisance and developed various mechanical stabilizations (Oliver and Quegan, 1998) and filtering techniques (Jazwinski, 1970) to eliminate the jittering caused by the vibration.

In contrast, the Dynamic Retina (DR) (Propokopowicz and Cooper, 1995) and the Resonant Retina (RR) (Hongler et al., 2003), new devices that take advantage of vibrating noise generated by mobile platforms, were proposed for contrast enhancement and edge detection respectively. The mechanism of those devices can be interpreted as an instance of stochastic resonance (SR) (Hongler et al., 2003). SR can be viewed as a noise induced enhancement of the response of a nonlinear system to a weak input signal, for example bistable devices (Gammaitoni et al., 1998) and threshold detectors (Greenwood et al., 1999), and naturally appears in many neural dynamics processes (Stemmler, 1996).

As an example of camera vibration noise, we focus on the small vibrations of a human eyeball, which

occur when we gaze at an object and are called "fixational eye movements." It has been reported that the vibrations may work not only as an intrinsic function to preserve photosensitivity but also as an assistance in image analysis, which can be considered as a realization of a biological SR phenomenon (Martinez-Conde et al., 2004). Although DR and RR offer massive parallelism and simplicity, the depth recovery method proposed by our group through random camera rotations (Tagawa, 2010) hints more potential of fixational eye movements, i.e. depth perception potential.

The method (Tagawa, 2010) employs a differential scheme based on the gradient method for "shape from motion" (Horn and Schunk, 1981), (Simoncelli, 1999), (Bruhn and Weickert, 2005). Fixational eye movements are classified into three types: microsaccade, drift and tremor, shown in Fig. 3. This method uses the camera rotations imitating tremor, which is the smallest of the three types, to reduce the linear approximation error in the gradient method. However, if a texture in an image is fine relative to an image motion size, the method suffers from an aliasing problem, namely a large amount of error occurs in spatio-temporal differentials of image intensity used in the gradient method.

In this study, in order to avoid the above mentioned aliasing problem, we propose a new scheme

based on an integral form using the same camera rotations adopted by the method (Tagawa, 2010). Small random camera rotations during exposure can generate two-dimensional motion-blur in images. The degree of the blur is a function of a pixel position due to a perspective projection assumed in this study, and it also depends on the depth value corresponding to the each pixel. Therefore, the depth map can be recovered by analyzing the motion-blur.

Several depth recovery methods using motion-blur have been already proposed, but those use the blur caused by definite and simple camera motions. For example, blur by a translational camera motion is used in Sorel and Flusser (Sorel and Flusser, 2008), and blur by an unconstrained camera motion composed of translation and rotation is assumed in Paramanand and Rajagopalan (Paramanand and Rajagopalan, 2012). The depth recovery performance of those methods may depend on a direction of a texture in images, i.e., if the texture has a strip pattern and its direction is parallel to the motion direction in the image, there is little blur and accurate depth recovery is difficult. As against those camera motions, random camera rotations used in this study is effective for arbitrary textures. Although only to solve this problem we can use also complicated but deterministic motions, random camera rotations are easy to implement in an actual system, since there is no need to control a camera with precision.

The proposed algorithm uses a motion-blurred image and a reference unblurred image. Especially in this study, from the point view of optimization for computation, we directly estimate a depth map without computing a space-variant point spread function. It is expected that the performance of the proposed scheme depends on the degree of motion blur. For the same point spread function, i.e. the fixed deviation of the random camera rotations, fine texture is advantageous for observing the accurate blur. To confirm this property, we carry out simulations using artificial images.

## 2 PRINCIPLE OF DEPTH FROM MOTION-BLUR

### 2.1 Camera Motions Imitating Tremor

We use a perspective projection system as our camera-imaging model. A camera is fixed with an  $(X, Y, Z)$  coordinate system; a lens center corresponding to a viewpoint is at origin  $O$  and an optical axis is along the  $Z$ -axis. By taking a focal length as a unit of ge-

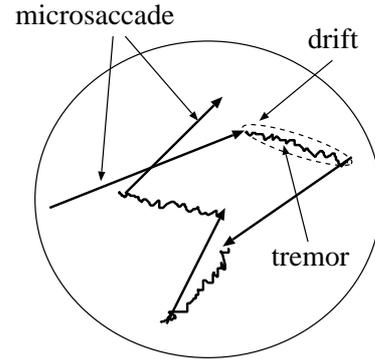


Figure 1: Illustration of fixational eye movements consist of microsaccade, drift and tremor.

ometrical representation, a projection plane, i.e. an image plane  $Z = 1$  can be used without any loss of generality. A space point  $(X, Y, Z)^T$  on an object is projected to an image point  $\vec{x} \equiv (x, y, 1)^T = (X/Z, Y/Z, 1)^T$ .

We briefly explain the motion model imitating tremor component of fixational eye movements proposed in our previous study (Tagawa, 2010). 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 can also be represented using the coordinate origin as its rotation center with the same components of the rotational vector  $\vec{r} = (r_X, r_Y, r_Z)^T$ . On the other hand, this difference between the origin and the rotation center causes a translational vector  $\vec{u} = (u_X, u_Y, u_Z)^T$  implicitly, and is formulated as follows:

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

Generally, a translational motion of a camera is needed to recover depth, and our camera motion model can cause it implicitly by only rotating a camera. This camera system can be easily controlled because of no explicit translations. This means that generally the system is developed and controlled simply. Additionally 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. The coordinate system and the camera motion model used in this study are shown in Fig. 2.

From Eq. 1, it can be known that  $r_Z$  causes no translations. Therefore, we set  $r_Z = 0$  and define  $\vec{r} = (r_X, r_Y, 0)^T$  as a rotational vector like an eyeball. In this study, to simplify the motion model,  $\vec{r}(t)$  is treated as a stochastic white process, in which  $t$  indicates time and  $\vec{r}(t)$  is measured absolutely from the

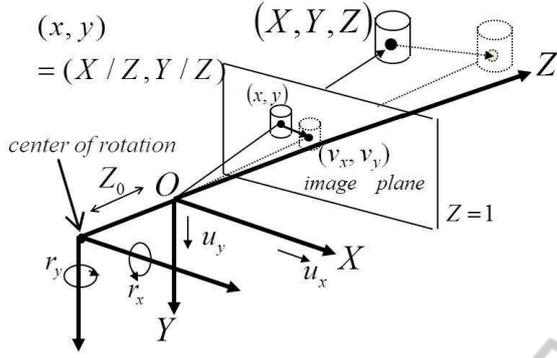


Figure 2: Coordinate system and camera motion model used in this study.

value at a reference time, i.e.  $\vec{r}(t)$  is not defined as the relative value between successive frames here. We ignore the temporal correlation of tremor which forms drift in actual fact, and we assume that the fluctuation of  $\vec{r}(t)$  at each time obeys a two-dimensional Gaussian distribution with a mean 0 and a variance  $\sigma_r^2$ , where  $\sigma_r^2$  is 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. In the actual system, we have no choice but to use a differential rotation, but for small values of the rotation angle, Eq. 1 and the other equations below hold approximately.

## 2.2 Motion Blur Associated with Depth

We can measure randomly fluctuating images with the proposed camera motion model. The previous method using this camera motion model adopted a differential formed strategy, i.e. temporal differentials of many image pairs were considered to be measurements, in which an optical flow field was implicitly analyzed (Tagawa, 2010). Since this differential formed method cannot deal with the images having fine texture detail because of an aliasing problem, in this study, we focus on an integral formed method which analyzes an accumulated image from small random rotations.

When the exposure time for imaging is sufficient, the accumulated image, i.e. the motion-blurred image,  $f_m(\vec{x})$  can be modeled as a convolution of an unblurred reference image  $f_0(\vec{x})$  with a two-dimensional point-spread function  $g_{\vec{x}}(\cdot)$  as follows:

$$f_m(\vec{x}) = \int_{\mathcal{R}} g_{\vec{x}}(\vec{x}') f_0(\vec{x} - \vec{x}') d\vec{x}' + n(\vec{x}), \quad (3)$$

where  $n(\vec{x})$  is an imaging noise,  $\mathcal{R}$  is a local support region of  $g_{\vec{x}}(\cdot)$  around  $\vec{x}$ , and  $\int g_{\vec{x}}(\vec{x}') d\vec{x}' = 1$  holds. It is expected that the degree of the motion blur in  $f_m(\vec{x})$  depends on a local depth value and is reflected in the degree of the spread of  $g_{\vec{x}}(\cdot)$ . In the following, we examine the relation between  $g_{\vec{x}}(\cdot)$  and a depth value.

Optical flow  $\vec{v} = (v_x, v_y, 0)^\top \equiv d\vec{x}/dt$  caused by a camera motion can be generally formulated using the inverse depth  $d(\vec{x}) = 1/Z(\vec{x})$  as follows:

$$\vec{v} = -\left(I - \vec{x}\vec{k}^\top\right) (\vec{r} \times \vec{x} + d(\vec{x})\vec{u}). \quad (4)$$

Hence, using Eq. 1, the optical flow caused by our camera model can be written specially as follows:

$$\vec{v} = -\left(I - \vec{x}\vec{k}^\top\right) (\vec{r} \times \vec{x}) - Z_0 d(\vec{x}) (\vec{r} \times \vec{k}), \quad (5)$$

where  $I$  indicates a  $3 \times 3$  unit matrix and  $\vec{k}$  is a unit vector indicating optical axis, i.e.  $\vec{k} = (0, 0, 1)^\top$ . This can be indicated with a component representation as follows:

$$v_x = xy r_x - (1 + x^2) r_y - Z_0 r_y d, \quad (6)$$

$$v_y = (1 + y^2) r_x - xy r_y + Z_0 r_x d. \quad (7)$$

In addition, from Eq. 2  $\vec{v}$  can be considered as a two-dimensional Gaussian random variable with a mean  $\vec{0}$  and a variance-covariance matrix

$$\begin{aligned} \vec{V}[\vec{v}] &= \sigma_r^2 \left\{ \left(I - \vec{x}\vec{k}^\top\right) \left(I - \vec{k}\vec{x}^\top\right) \right\}^2 \\ &\quad + 2\sigma_r^2 Z_0 d(\vec{x}) \left(I - \vec{x}\vec{k}^\top\right) \left(I - \vec{k}\vec{x}^\top\right) \\ &\quad + \sigma_r^2 Z_0^2 d(\vec{x})^2 \left(I - \vec{k}\vec{k}^\top\right) \\ &= \sigma_r^2 \times \\ &\quad \begin{bmatrix} x^2 y^2 + (1 + x^2 + Z_0 d)^2 & 2xy(1 + \frac{x^2 + y^2}{2} + Z_0 d) \\ 2xy(1 + \frac{x^2 + y^2}{2} + Z_0 d) & x^2 y^2 + (1 + y^2 + Z_0 d)^2 \end{bmatrix}. \end{aligned} \quad (8)$$

From these discussion, intensity at each pixel is added up a lot of neighboring pixel's intensity and a relative displacement from such a neighboring pixel per unit time can be approximately considered as the optical flow defined by Eq. 5. Therefore, it is clear that  $g_{\vec{x}}(\cdot)$  can be modeled approximately by a two-dimensional Gaussian distribution having the same variance-covariance matrix of Eq. 8. Hence, the motion-blur caused by our camera model depends on the depth map, and Eq. 3 can be used as an observation equation including the unknown variable  $d(\vec{x})$ . Hereafter, we use the representation  $g_{\vec{x}}(\cdot; d)$  to clarify that the blur is a function of depth.

### 3 ALGORITHM DESCRIPTION

With optimality in mind, we examine the direct method which directly estimate a depth map without determining  $g_{\vec{x}}(\cdot; d)$ , although this strategy generally requires a numerical search or an iterative update. We construct two algorithms of a direct method, each of which employs respectively a local optimization and a global optimization. In the following, we briefly explain both algorithms. It should be noted that since the scheme examined in this study is based on the spatial blur, high resolution recovery cannot be expected essentially. The proposed algorithms introduced in the following employ simple computations with no complicated techniques for edge preserved recovery and so on. If high resolution and high accurate recovery is needed, for example, the methods based on differential schemes have to be performed using the results obtained by the algorithms in this study as an initial values. At that time, by applying image warping processing with the initial depth values, the aliasing problem concerned in the differential scheme can be effectively avoided (Simoncelli, 1999), (Tagawa et al., 2008), (Tagawa and Naganuma, 2009).

#### A. Local Optimization Algorithm

For stable recovery, we assume that a depth value in a local region  $\mathcal{L}$  around each  $\vec{x}$  is constant. We can define the objective function with respect to the depth corresponding to each pixel based on the minimum least square criterion.

$$J_L(d(\vec{x})) \equiv \int_{\mathcal{L}} (f_m(\vec{x} - \vec{x}') - f_{conv}(\vec{x} - \vec{x}'))^2 d\vec{x}', \quad (9)$$

$$f_{conv}(\vec{x}) \equiv \int_{\mathcal{R}} g_{\vec{x}}(\vec{x}'; d) f_0(\vec{x} - \vec{x}') d\vec{x}'. \quad (10)$$

By minimizing this function defined at each  $\vec{x}$  respectively, we can recover separately the depth corresponding to each pixel. Therefore, a multivariate optimization is not needed, and we can adopt a one-dimensional numerical search.

#### B. Global Optimization Algorithm

By requiring a spatially smooth depth map, we can define the following functional to be minimized based on the regularization theory of Poggio et al. (Poggio et al., 1985).

$$J_G(d(\vec{x})) = (1 - \lambda) \int (f_m(\vec{x}) - f_{conv}(\vec{x}))^2 d\vec{x} + \lambda \int \left\{ \left( \frac{\partial d(\vec{x})}{\partial x} \right)^2 + \left( \frac{\partial d(\vec{x})}{\partial y} \right)^2 \right\} d\vec{x}, \quad (11)$$

where  $\lambda$  is a weight for adjusting the degree of the smoothness constraint requirement for a depth map, and the integration in Eq. 11 is carried out in a whole image. From the variational principle, the Euler-Lagrange equation for finding the solution of  $d(\vec{x})$  is derived using  $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$  as follows:

$$\nabla^2 d = -\frac{1 - \lambda}{\lambda} (f_m - f_{conv}) \frac{\partial f_{conv}}{\partial d}. \quad (12)$$

For discrete computation, we can approximate the smoothness constraint in Eq. 11 using  $(i, j)$  as a description of an image position.

$$\begin{aligned} & \left( \frac{\partial d(\vec{x})}{\partial x} \right)^2 + \left( \frac{\partial d(\vec{x})}{\partial y} \right)^2 \\ & \approx \frac{1}{5} \{ (d_{i+1,j} - d_{i,j})^2 + (d_{i,j+1} - d_{i,j})^2 \} \\ & + \frac{1}{20} \{ (d_{i+1,j-1} - d_{i,j})^2 + (d_{i+1,j+1} - d_{i,j})^2 \}. \end{aligned} \quad (13)$$

Using Eq. 13 and the discrete representation of Eq. 12, we can minimize Eq. 11 by the following iterative formulation with an iteration number  $n$ .

$$d_{i,j}^{(n+1)} = \bar{d}_{i,j}^{(n)} + \frac{1 - \lambda}{\lambda} (f_{m,i,j} - f_{conv}(d_{i,j}^{(n)})) \frac{\partial f_{conv}(d_{i,j}^{(n)})}{\partial d}, \quad (14)$$

$$\begin{aligned} \bar{d}_{i,j}^{(n)} &= \frac{1}{5} (d_{i+1,j}^{(n)} + d_{i,j+1}^{(n)} + d_{i-1,j}^{(n)} + d_{i,j-1}^{(n)}) \\ &+ \frac{1}{20} (d_{i+1,j+1}^{(n)} + d_{i+1,j-1}^{(n)} + d_{i-1,j-1}^{(n)} + d_{i-1,j+1}^{(n)}). \end{aligned} \quad (15)$$

### 4 NUMERICAL EVALUATIONS

The proposed algorithms suppose the definition of the motion-blurred image in Eq. 3, and this realization hardly depends on the imaging system and its control. Namely, to observe the ideal motion-blur, we have to take sufficient exposure time for imaging, and during the exposure time camera motion has to be suitably controlled with small and random properties. Now, we are developing a real imaging system, and will confirm the validity of Eq. 3 immediately. In this study, we examine the performance of the proposed algorithms with respect to the relation between the image motion size and the fineness of the texture using artificial data.

We artificially obtain motion-blurred images by digital signal processing. Firstly, we generate a huge number of images by a computer graphics technique using a true depth map and randomly sampling  $\vec{r}$  according to the Gaussian distribution in Eq. 2. An artificial motion-blurred image can be made by averaging

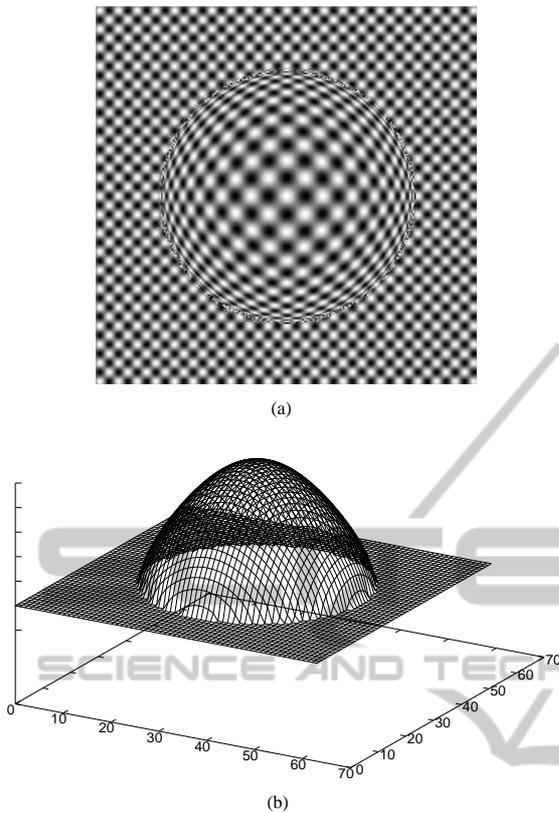


Figure 3: Example of the artificial data used in the experiments: (a) original image; (b) true inverse depth map used for generating the blurred image.

these images. In this study, the input motion-blurred image is obtained by averaging 10,000 images to imitate analog motion blur. Figure 3 shows an example of a reference image and a true inverse depth map. The image size used in the simulations is  $256 \times 256$  pixels, which corresponds to  $-0.5 \leq x, y \leq 0.5$  measured using the focal length as a unit. In Fig. 3(b), the vertical axis indicates the inverse depth  $d(\vec{x})$  using the focal length as a unit, and the horizontal axes in it indicate a pixel position in the image plane, which is marked every four pixels.

The local optimization algorithm (LOA) takes high computational cost at each pixel, and the global optimization algorithm (GOA) converges slowly. Hence, we evaluated a hybrid algorithm, in which the LOA is used sparsely in the image plane to obtain the initial values for the GOA. On the other hand, the plane indicating the background in Fig. 3(b) is used as the initial values for the LOA. Since the LOA is used for rough estimate, we used a block with  $41 \times 41$  pixels as  $\mathcal{L}$  in Eq. 9 without any special consideration and apply the LOA once to each block. On the other hand, we adaptively determined the size of  $\mathcal{R}$  in Eq. 10 according to the value of the depth updated in the opti-

mization process. Therefore,  $\mathcal{R}$  took the different size at each position in the image. We supposed a square region for  $\mathcal{R}$ , the side length of which was ten times as large as the larger of the two deviations of  $g_{\vec{x}}(\cdot; d)$ , i.e.  $x$ -deviation and  $y$ -deviation, which can be evaluated using Eq. 8.

We performed simulations with varying the size of camera rotation  $\sigma_r$ . The recovered inverse depth maps are shown in Figs. 4-6 with various values of  $\lambda$  used in the GOA. The relation between the root mean square error (RMSE) of the recovered depth map and the value of  $\lambda$  is also shown in Fig. 7. From Fig. 4, it can be easily known that small camera rotations are inadequate for depth recovery, since the motion blur in the image position is hard to be measured accurately. From Fig. 7(a), since the measured information is poor by the small rotations, the smoothness constraint indicated by  $\lambda$  is strongly needed to reduce the RMSE of the recovered depth map. On the contrary, the large rotations make the point-spread function extensive compared with a spatial variation of the target shape, and hence the Gaussian function with the variance-covariance matrix in Eq. 8 is improper and the motion blur recognized by this model becomes smoother than the true blur in the image, which causes depth recovery error. This can be seen from the RMSE values in Fig. 7(c). We can confirmed from Fig. 7(c) also that since the smoothness of the recognized motion blur tends to recover a smooth depth, the smoothness constraint in Eq. 11 is obstructive for low RMSE. Figure 8 shows the result with extremely large rotations of  $\sigma_r = 0.016$ . Such the rotations are too large for the texture of the images used in this evaluation, and it was confirmed that the RMSE is almost independent of the value of  $\lambda$ .

Additionally, we tried to recover a depth from images having a strip line pattern, which is insufficient for the method using one-directional motion blur. Figure 9 shows the original image, the images blurred with  $\sigma_r = 0.08$  and the recovered inverse depth map with  $\lambda = 0.2$ , the RMSE of which is minimum among various values of  $\lambda$ . From this result, we confirmed that our integral formed method is suitable also for such a line pattern image. The case when  $\sigma_r$  is less than about 0.004, the proposed method can not obtain good results for the texture patterns used in this study, because the image motions are too small for the texture pattern and hence there are little blur. For such the case, we can use our differential method (Tagawa, 2010) instead.

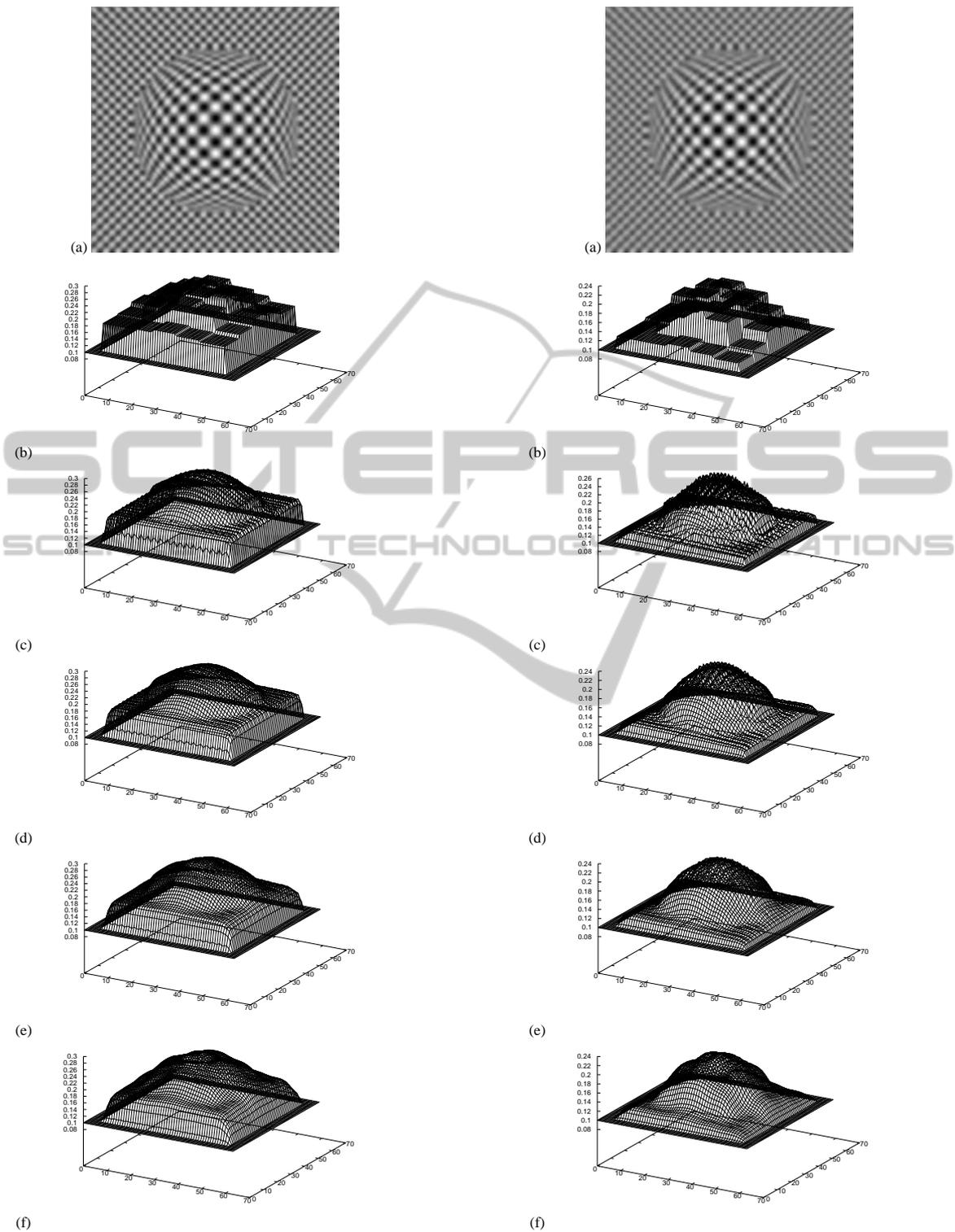


Figure 4: Example of motion-blurred image and recovered inverse depth maps with  $\sigma_r = 0.006$ : (a) motion-blurred image; (b) local optimization; (c)  $\lambda = 0.2$ ; (d)  $\lambda = 0.4$ ; (e)  $\lambda = 0.6$ ; (f)  $\lambda = 0.8$ .

Figure 5: Example of motion-blurred image and recovered inverse depth maps with  $\sigma_r = 0.008$ : (a) motion-blurred image; (b) local optimization; (c)  $\lambda = 0.2$ ; (d)  $\lambda = 0.4$ ; (e)  $\lambda = 0.6$ ; (f)  $\lambda = 0.8$ .

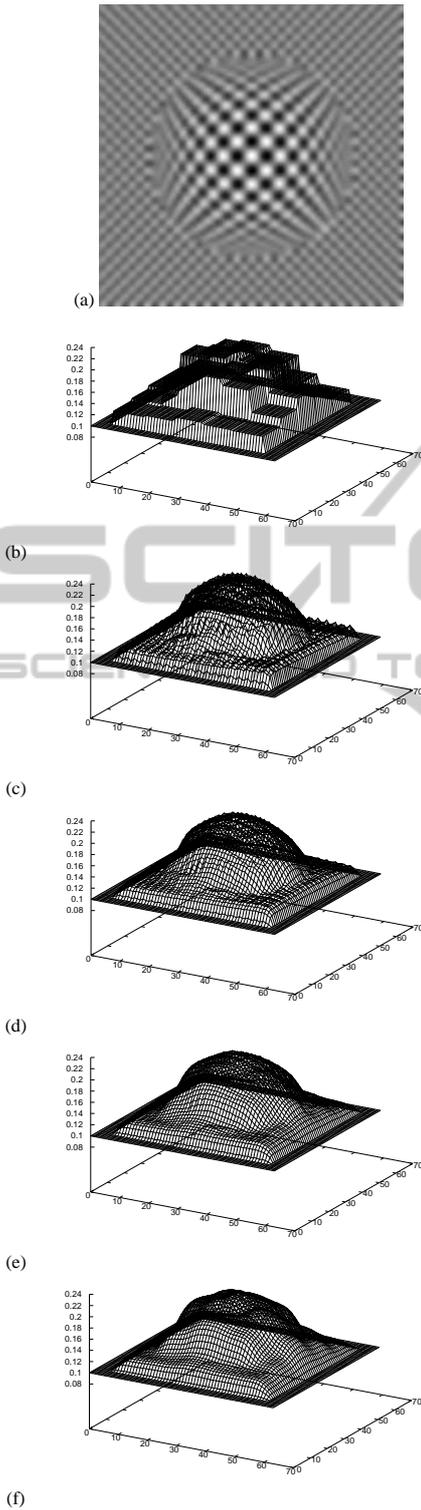


Figure 6: Example of motion-blurred image and recovered inverse depth maps with  $\sigma_r = 0.01$ : (a) motion-blurred image; (b) local optimization; (c)  $\lambda = 0.2$ ; (d)  $\lambda = 0.4$ ; (e)  $\lambda = 0.6$ ; (f)  $\lambda = 0.8$ .

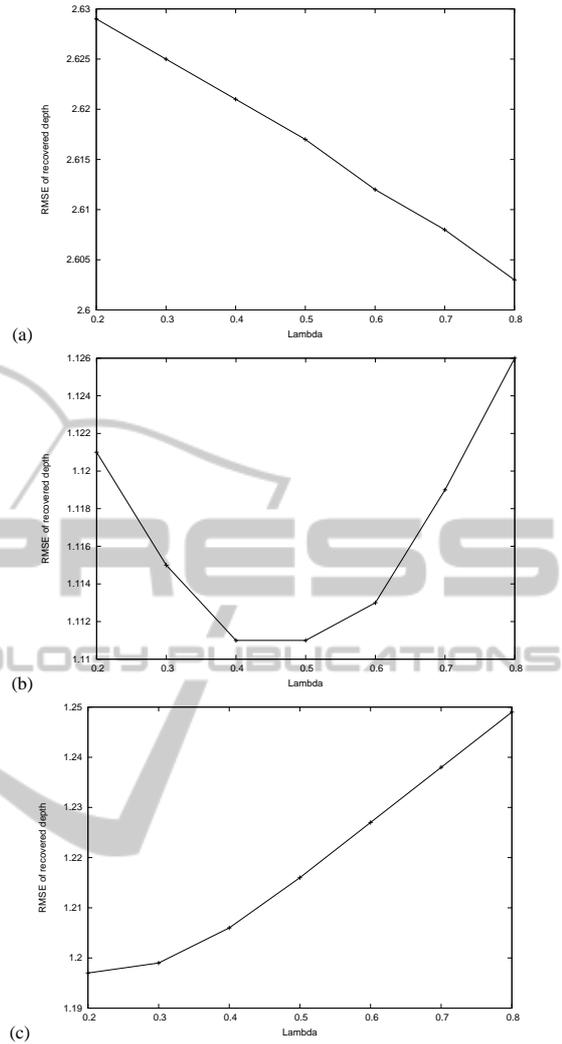


Figure 7: Relation between RMSE of recovered depth and lambda: (a)  $\sigma_r = 0.06$ ; (b)  $\sigma_r = 0.08$ ; (c)  $\sigma_r = 0.1$ .

## 5 DISCUSSIONS

### 5.1 Parameter Determination

We understand that the region  $\mathcal{L}$  in Eq. 9 should be determined according to the desirable resolution of the recovered depth map while keeping the constraint that  $\mathcal{L}$  has to be larger than the deviation of  $g_{\vec{x}}(\cdot; d)$  to avoid indefiniteness of depth recovery. On the other hand, for  $\mathcal{R}$  in Eq. 10, if there is no need to save computation cost, it is best to use a very large value because of the infinite support of Gaussian function.

In this study, since we use LOA for obtaining the initial depth for GOA, we uniformly used the square of  $41 \times 41$  pixels as  $\mathcal{L}$  shown in Sec. 4. For  $\mathcal{R}$ , to

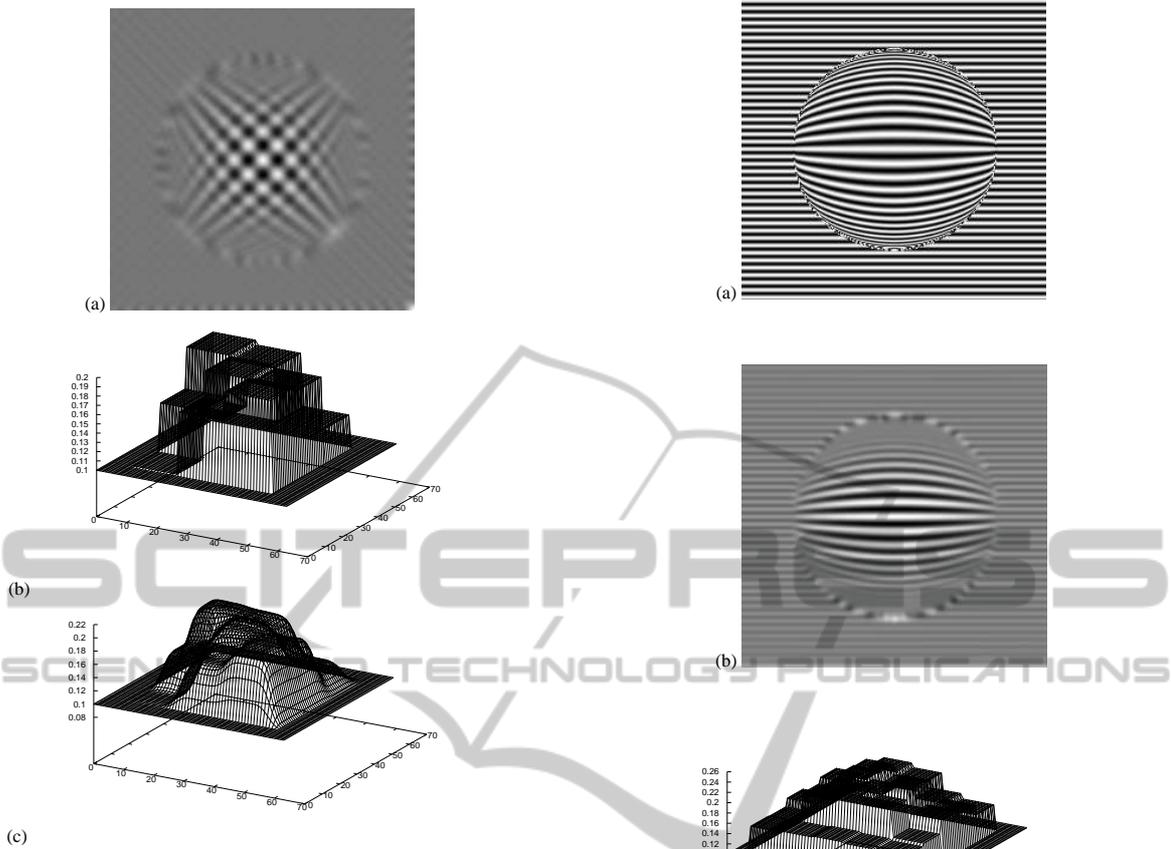


Figure 8: Example of motion-blurred image and recovered inverse depth maps with  $\sigma_r = 0.16$ : (a) motion-blurred image; (b) local optimization; (c)  $\lambda = 0.4$ , which is an example that of camera rotations are too large compared with a texture pattern.  $\mathcal{L}$  in local optimization is  $61 \times 61$  pixels.

reduce numerical errors, we used the square large enough to cover the whole support of  $g_{\bar{x}}(\cdot; d)$  concretely explained in Sec. 4. For the true depth values, the size of  $\mathcal{R}$  takes the value between  $15 \times 15$  pixels and  $19 \times 19$  pixels for  $\sigma_r = 0.005$ . The size of  $\mathcal{L}$  of  $41 \times 41$  is large enough with respect to the size of  $\mathcal{R}$  as a result. We confirmed that up to  $\sigma_r = 0.012$   $\mathcal{L}$  with  $14 \times 41$  is sufficient with respect to the size of  $\mathcal{R}$ , but for  $\sigma_r = 0.016$ , the recovered result by which is shown in Fig. 8, the size of  $\mathcal{L}$  had to be expanded as  $61 \times 61$  pixels in LOA.

For the setting of  $\lambda$ , in this study we suppose that the value should be adjusted empirically while checking the validity of the recovered depth. In future, we will adopt the EM algorithm (Dempster et al., 1977) (Tagawa et al., 2008), (Tagawa and Naganuma, 2009) to solve this problem, by which  $\lambda$  can be automatically determined from only the observed data set. In this framework, by modeling  $\lambda$  as a function of an image position using an MRF model, spatially adaptive  $\lambda$  is expected to be determined.

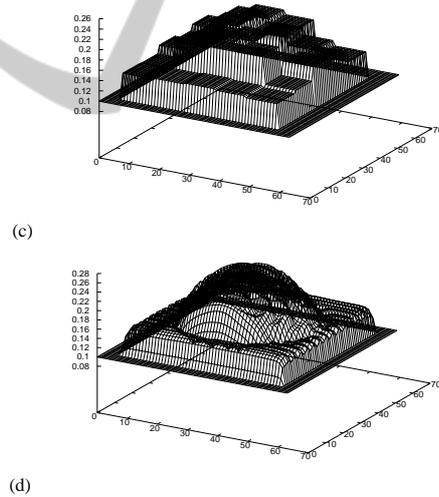


Figure 9: Results for horizontal strip pattern with  $\sigma_r = 0.08$ : (a) original image; (b) motion-blurred image; (c) local optimization; (d)  $\lambda = 0.2$  (RMSE = 1.870).

## 5.2 Relation with Depth from Defocus

At first, we refer to the bad influence of the blur caused by out of focus on our proposed method. In this study, we adopt a pinhole camera as an ideal one having no defocusing to simply explain our idea, but actual cameras can be used generally in our method. If the camera rotations are not so large, which is assumed in this study from the beginning, the degree

of the blur caused by defocusing is unchanging before and after the camera rotations. Our method uses a reference image and a blurred image for processing, and both have the same defocusing-blur, hence our method can cancel automatically the defocusing-blur.

Next, we explain the advantages of our method over the depth-from-focus method (Nayar and Nakagawa, 1994). In the depth-from-focus method, focus should be varied accurately in several different ways. However, in our method, a camera has only to be rotated randomly, i.e. accurate control of a camera is not required. Since, in our future method, the deviation of the random camera rotations will be estimated from the observed images, there will be no need to know the deviation before processing. Another advantage of our motion-blur scheme is that a lot of still images having no motion-blur, which are averaged to generate a motion-blur image, can be observed and processed if needed. Depth recovery based on either motion-blur or defocusing-blur is fitted for sufficiently fine textures, hence surfaces with originally blurred texture cannot be handled. However, if camera rotations are adopted, we can deal with such a smooth texture using the differential method (Tagawa, 2010). Hence, we can adaptively recover the depth by switching the integral method proposed in this study and the differential method according to the fineness of the surface texture.

## 6 CONCLUSIONS

We propose a new method to recover a depth map using the camera rotations imitating fixational eye movements, in particular tremor-related movements. The proposed method can compute a depth map directly from blurred image. In this study, we approximate the motion-blurred image by averaging a huge number of images artificially generated by the random camera rotations, and we are yet to examine the effectiveness of our method by real image experiments using an actual imaging system in the future work. The simulations in this study did not consider lighting condition and reflection characteristics of an imaging target. Especially, to examine an influence of specular reflection components, at first numerical evaluations have to be done thoroughly, and subsequent experiments are strongly required.

An outline of a depth map can be recovered by the method in the simulations, but its accuracy may be insufficient. The proposed method cannot be used for small image motions relative to a texture pattern. For this case, the differential method (Tagawa, 2010) is effective. On the other hand, from the funda-

mental principle of our method that the image blur is used for depth recovery, the spatial resolution of the recovered depth is not so high, no matter how careful we are on selecting the camera motion size. For this case, we can use the results from the proposed integral method as an initial depth for the method (Tagawa et al., 2008) (Tagawa and Naganuma, 2009). Hence, we plan on unifying those methods to deal with various situations. Especially to combine the differential method (Tagawa, 2010) and the integral method in this paper, we have to develop a suitable segmentation method, which divides observed images into fine texture regions and rough texture regions, taking into account the size of camera rotations. Additionally, to use both of the differential and the integral methods simultaneously, the motion-blurred image has to be generated by averaging many captured images without motion blur instead of simply capturing analog blur image using the suitable exposure time. For the case, lesser number of the image used for averaging is desirable for computational cost and real-time operation, but this requirement cannot realize the ideal motion blur supposed in this study. Therefore, we have to improve the integral method in this study to give a good performance using such the insufficient motion blur.

## REFERENCES

- 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.
- Gammaitoni, L., Hanggi, P., Jung, P., and Marchesoni, F. (1998). Stochastic resonance.
- Greenwood, P. E., Ward, L. M., and Wefelmeyer, W. (1999). Statistical analysis of stochastic resonance in a simple setting. *Physical Rev. E*, 60:4687–4696.
- Hongler, M.-O., de Meneses, Y. L., Beyeler, A., and Jacot, J. (2003). The resonant retina: exploiting vibration noise to optimally detect edges in an image. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(9):1051–1062.
- Horn, B. P. and Schunck, B. (1981). Determining optical flow. *Artif. Intell.*, 17:185–203.
- Jazwinski, A. (1970). *Stochastic processes and filtering theory*. Academic Press.
- Martinez-Conde, S., Macknik, S. L., and Hubel, D. (2004). The role of fixational eye movements in visual perception. *Nature Reviews*, 5:229–240.
- Nayar, S. K. and Nakagawa, Y. (1994). Shape from focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(8):824–831.

- Oliver, C. and Quegan, S. (1998). *Understanding synthetic aperture radar images*. Artech House, London.
- 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.
- Propokopowicz, P. and Cooper, P. (1995). The dynamic retina. *Int'l J. Computer Vision.*, 16:191–204.
- Simoncelli, E. P. (1999). Bayesian multi-scale differential optical flow. In *Handbook of Computer Vision and Applications*, pages 397–422. Academic Press.
- Sorel, M. and Flusser, J. (2008). Space-variant restoration of images degraded by camera motion blur. *IEEE Trans. Image Processing*, 17(2):105–116.
- Stemmler, M. (1996). A single spike suffices: the simplest form of stochastic resonance in model neuron. *Network: Computations in Neural Systems*, 61(7):687–716.
- Tagawa, N. (2010). Depth perception model based on fixational eye movements using byesian statistical inference. In *proc. ICPR2010*, pages 1662–1665.
- 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. and Naganuma, S. (2009). *Pattern Recognition, Structure and motion from image sequence based on multi-scale Bayesian network*. In-Tech, Croatia.