We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Stereoscopic Calculation Model Based on Fixational Eye Movements

Norio Tagawa

Abstract

Fixational eye movement is an essential function for watching things using the retina, which has the property of responding only to changes in incident light. However, since the rotation of the eyeball causes the translational movement of the crystalline lens, it is possible in principle to recover the depth of the object from the moving image obtained in this way. We have proposed two types of depth restoration methods based on fixation tremor; differential-type method and integral-type method. The first is based on the change in image brightness between frames, and the latter is based on image blurring due to movement. In this chapter, we introduce them and explain the simulations and experiments performed to verify their operation.

Keywords: motion stereoscopic, fixational eye movements, differential-type method, integral-type method, optical flow, gradient equation, image blur

1. Introduction

When humans stare at a target, an irregular involuntary movement called fixational eye movements occur [1]. The human retina can maintain reception sensitivity by finely vibrating the image of the target on the retina, so in order to see something, first fixation motion is required. It has been reported that the vibrations may work not only as such the intrinsic function to preserve photosensitivity but also as an assistance in image analysis, the mechanism of which can be interpreted as an instance of stochastic resonance (SR) [1]. SR is inspired by biology, more specifically by neuron dynamics [2], and based on it, the Dynamic Retina (DR) [3] and the Resonant Retina (RR) [4], which are new vision devices taking advantage of random camera vibrations, were proposed for contrast enhancement and edge detection respectively. It has been reported that the movement of the retinal image due to fixation eye movements can be an unconscious clue to depth perception, and an actual vision system based on fixational eye movements has been proposed [5].

On the other hand, binocular stereopsis is vigorous and plays an essential role in depth perception of a human vision system [6]. In general, binocular stereopsis detects relatively large disparities, hence it can recognize high accurate depth. However this causes an occlusion problem, and a lot of solutions of it have been proposed. Wang et al. have proposed a local detector for occlusion based on deep learning [7]. In [8], a robust depth restoration method has been proposed that integrates line-field imaging technology that simultaneously observes multiple angle views with stereo vision. Therefore, we expect that primitive depth

information detected by fixational eye movements can be used to solve occlusion for binocular stereopsis. There is a concern that the accuracy of depth restoration by a small camera motion is lower than that of stereo vision. Even so, it is expected that erroneous correspondence due to the existence of occlusion can be reduced by using the depth information from fixational eye movements for the correspondence problem in stereo vision.

In monocular stereoscopic vision, “structure from motion (SFM)” has been the most widely studied, and many remarkable results have been reported. SFM has various calculation principles. To achieve spatially dense depth recovery with high computational efficiency, a method based on the gradient equation that expresses the constraint between the spatiotemporal derivative values of image intensity and the movement on the image is effective [9–11]. It should be noted that for such a gradient method, there is an appropriate size of movement to recover the correct depth. Since the gradient equation holds perfectly for small motions, the error in the equation cannot be ignored for very large motions. On the contrary, in the case of small movement, the motion information is buried in the observation error of the spatiotemporal derivative in intensity.

Adaptation of the frame rate is required to make the motion size suitable for the gradient method. We have proposed a method that does not require a variable frame rate based on multi-resolution decomposition of images, but it requires high computational cost [12]. Therefore, we focus on small movements with an emphasis on avoiding equation errors in the gradient method. Then, in order to solve the above signal-to-noise ratio (S/N) problem that occurs with small movements, many observations are collected and used all at once [13, 14]. In such a strategy, it is desirable that the direction and size of the motion take different values. From the above discussion, we examined a depth perception model based on fixational eye movement and gradient method. Fixational eye movements are divided into three types: microsaccades, drifts, and tremors. As the first report of our attempt, we focused on tremor, the smallest of the three types. In the next step, we plan to use drift and microsaccade analogies for further progress. Using a lot of images captured with random small motions of camera, which consists of three-dimensional (3-D) rotations imitating fixational eyeball motions [1], many observations can be used at each pixel, i.e. many gradient equations can be used to recover the each depth value corresponding to the each pixel. Since the difference between the center of the three-dimensional rotation and the lens center generates a translational motion of the lens center, depth information can be obtained from these images. Simulations with artificial images confirm that the proposed method works effectively when the observed noise is an actual sample of a theoretically defined noise model.

However, if the wavelength of the main luminance pattern is small compared to the size of the motion in the image, aliasing will occur and the gradient equation will be useless. In other words, the methods of [13, 14] cannot be applied. To avoid the above problem, we proposed a new scheme based on the integral form that also used the analogy of fixational eye movement [15, 16]. Add up the many images generated by the above method to get one blurry image. The degree of blur is a function of pixel position and also depends on the depth value of each pixel. That is, the difference in the degree of image blur indicates the depth information. Based on the proposed scheme, the spatial distribution of the image blur is effectively estimated using the blurred image and the original image without blur. By modeling the small 3-D rotation of the camera as a Gaussian random variable, the depth map can be calculated analytically from this blur distribution.

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 [17], and blur by an unconstrained camera motion composed of translation and rotation is assumed in [18]. The depth recovery performance of these methods depends on the orientation of the texture in the image. That is, if the texture has a strip pattern whose direction is parallel to the direction of motion in the image, there is less blurring and accurate depth recovery is difficult. Unlike such a specific camera motion, the random camera rotation used in this study works well for any texture. Deterministic camera motion can be used just to solve this problem, but since it does not require precise control of the camera, it is easy to implement random camera rotation in a real system.

We proposed two algorithms based on the integral scheme. The first algorithm detects a point spread function (PSF) that represents image blur and then analyzes it to restore depth [15]. The second algorithm directly calculates the depth without detecting the PSF [16]. These algorithms use a motion-blurred image and a reference un-blurred image. It is expected that the performance of the proposed scheme depends on the degree of motion blur. For the same PSF, i.e. the fixed deviation of the random camera rotations, fine texture is advantageous for observing the accurate blur. This characteristic is the opposite of the method based on the differential scheme based on the gradient equation.

The features of our methods described above can be summarized as follows.

1. In a camera motion model that simulates eyeball rotation, the translation of the lens center is secondarily generated by eye rotation, reducing the number of unknown parameters. This camera motion model also allows you to recover depth as an absolute quantity instead of a relative quantity.
2. The movement of the camera is subtle because it simulates the tremor component of the eyeball. Therefore, by using a large number of image pairs, it is possible to improve the accuracy of depth recovery while avoiding occlusion.
3. In order to use multiple image pairs at the same time, we have adopted a direct framework that explicitly uses the inverse depth, which is a common parameter for them.

In the following, we will first explain the camera model and the imaging system, and then explain the differential scheme and the integral scheme, and the algorithms based on each. Due to the limitation of the number of pages, the integer type explains only the direct method. The function and characteristics of each algorithm are shown as the result of computer simulation with an emphasis on quantitative comparison with the true value. Finally, one of the algorithms in the differential scheme is applied to the real image, and the result is explained.

2. Camera motions imitating tremor and projection model

In this research, we use a perspective projection system as a camera imaging model. The camera is fixed in the (X, Y, Z) coordinate system. The lens center corresponding to the viewpoint is at the origin O , and the optical axis is along the Z axis. By taking the focal length as a unit of geometric representation, we can assume image plane $Z = 1$ without loss of generality. The point $(X, Y, Z)^T$ on the object in 3-D space is projected on the image point $\boldsymbol{x} \equiv (x, y, 1)^T = (X/Z, Y/Z, 1)^T$.

A brief description of the camera motion model that mimics the tremor component of fixational eye movements proposed in the previous study [12]. According to

the analogy of the human eyeball, the center of camera rotation is set behind the lens center by Z_0 along the optical axis, and there is no explicit translational motion of the camera. This rotation vector $\mathbf{r} = (r_x, r_y, r_z)^T$ also expresses a rotation vector centered on the coordinate origin, which is the lens center, using the same component. On the other hand, this difference between the coordinate origin and the center of rotation results in the translation vector $\mathbf{u} = (u_x, u_y, u_z)^T$, which is formulated as follows:

$$\begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ Z_0 \end{bmatrix} = Z_0 \begin{bmatrix} r_y \\ -r_x \\ 0 \end{bmatrix}. \quad (1)$$

In general, the translational motion of a camera lens is essential to recover depth, and our camera motion model can implicitly achieve that translation simply by rotating the camera. This facilitates camera control. In addition, the system can recover absolute depth by pre-calibrating Z_0 . We show the coordinate system and camera motion model used in this study in **Figure 1**.

From Eq. (1), it can be known that r_z causes no translations. Therefore, we can set $r_z = 0$ and redefine $\mathbf{r} = (r_x, r_y, 0)^T$ as a rotational vector like an eyeball. Using Eq. (1) and the inverse depth $d(x, y) = 1/Z(x, y)$, image motion called “optical flow” $\mathbf{v} = [v_x, v_y]^T$ is given as follows:

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

$$v_y = (1 + y^2) r_x - xy r_y + Z_0 r_x d. \quad (3)$$

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

In this study, we treat $\mathbf{r}(t)$ as a white stochastic process to simplify the motion model, and t indicates time. $\mathbf{r}(t)$ is defined as the rotation speed with respect to the camera orientation of $t = 0$, not the derivative between consecutive frames. In the actual fixational eye movement, the temporal correlation of tremor that forms the drift component is ignored. We assume that each fluctuation of $\mathbf{r}(t)$

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

follows a two-dimensional (2-D) Gaussian distribution with a mean of $\mathbf{0}$ and a variance–covariance matrix of $\sigma_r^2 \mathbf{I}$ with an identity matrix $\mathbf{I}$.

$$p(\mathbf{r}(t)|\sigma_r^2) = \frac{1}{(\sqrt{2\pi}\sigma_r)^2} \exp \left\{ -\frac{\mathbf{r}(t)^T \mathbf{r}(t)}{2\sigma_r^2} \right\}. \quad (4)$$

In this study, $\mathbf{r}$ is defined as the rotational speed for ease of theoretical analysis. In a real system, we have no choice but to use finite rotation, but if the angle of rotation value is small, then the formulations are almost valid.

3. Differential-type method

3.1 Gradient equation for rigid motion

The general 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. Assuming that the brightness values before and after 3-D motion are equal, the image brightness after 3-D motion are expressed by Taylor expansion, and terms of degree 2 and above are ignored. As a result, at each pixel (x, y) , the gradient equation is formulated with the partial differentials f_x , f_y and f_t , where t denotes time, of the image brightness $f(x, y, t)$ and the optical flow, as follows:

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

By substituting Eqs. (2) and (3) into Eq. (5), the gradient equation representing a rigid motion constraint can be derived explicitly.

$$f_t = -\left(f_x v_x^r + f_y v_y^r\right) - \left(-f_x r_y + f_y r_x\right) Z_0 d \equiv -f^r - f^u d. \quad (6)$$

3.2 Probabilistic model for differential-type method

Let M be the number of pairs of two frames and N be the number of pixels. In our study, $\left\{f_t^{(i,j)}\right\}_{i=1,\dots,N;j=1,\dots,M}$ and $\left\{\mathbf{r}^{(j)}\right\}_{j=1,\dots,M}$ are treated as random variables, and $\left\{d^{(i)}\right\}_{i=1,\dots,N}$ corresponding to the inverse depth of each pixel is treated as stochastic variable and recovered individually for each pixel. However, multiple frames $\left\{\mathbf{r}^{(j)}\right\}$ that vibrate due to irregular rotation are used for processing, but no pixel tracking is done. Therefore, the recovered $d^{(i)}$ at each pixel does not correspond exactly to the value of this pixel, but takes the mean of the adjacent regions defined by the vibration width of the image. As a result, the recovered $d^{(i)}$ correlates with the values in the adjacent regions. Therefore, from the beginning, $d^{(i)}$ should be treated as a variable with such a correlation. Based on tremor, $d^{(i)}$, which is estimated to correlate with the neighborhood, is planned to be improved to d for each pixel when dealing with drift and microsaccade in future research.

In this study, we assume 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 $f_t^{(ij)}$ is a Gaussian random variable with mean 0 and variance σ_o^2 , and $f_x^{(ij)}$ and $f_y^{(ij)}$ have no error.

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

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

As mentioned above, considering the neighborhood correlation of d to be recovered, in this study, to simplify modeling, we use the following equation as the depth model.

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

where $\mathbf{d}$ is a N -dimensional vector composed of $\{d^{(i)}\}$ and $\mathbf{L}$ indicates a matrix corresponding to the 2-D Laplacian operator with a free end condition. By assuming this probabilistic density, we make a recovered depth map smooth. In this study, the variance σ_d^2 is controled heuristically in consideration of smoothness of a recovered depth map. In future, we are going to examine a strategy for determination of σ_d^2 in the whole system which models all fixational movements including microsaccade and drift also. Hereafter, we use the definition $\Theta \equiv \{\sigma_o^2, \sigma_r^2\}$. In the simulation described later, random rotation is sampled according to Eq. (4) as the rotation for the initial image, but since the rotation between successive frames is estimated during depth restoration, the determined σ_r^2 and the set value are different.

3.3 Algorithm for differential-type method

By applying the MAP-EM algorithm [19], parameter $\{\mathbf{d}, \Theta\}$ can be estimated as a MAP estimator based on $p(\mathbf{d}, \Theta | \{f_t^{(ij)}\})$, which is formulated by marginalizing the joint probability $p(\{r^{(j)}\}, \mathbf{d}, \Theta | \{f_t^{(ij)}\})$ with respect to $\{r^{(j)}\}$, but the prior of Θ is formally regarded as an uniform distribution. Additionally, $\{r^{(j)}\}$ can be estimated as a MAP estimator based on $p(\{r^{(j)}\} | \{f_t^{(ij)}\}, \hat{\Theta}, \hat{\mathbf{d}})$, in which $\hat{\cdot}$ means a MAP estimator described above.

In the EM scheme, $\{\{f_t^{(ij)}\}, \{r^{(j)}\}\}$ is considered as a complete data, $\{r^{(j)}\}$ is treated as a missing data, and $\{\mathbf{d}, \Theta\}$ is treated as an unknown parameter. E step and M step are mutually repeated until they converge. At first, at the E step, calculate the conditional expectation of the log likelihood of the complete data with observing $\{f_t^{(ij)}\}$, using the current MAP estimates $\{\hat{\mathbf{d}}, \hat{\Theta}\}$. It is generally called Q function. Especially for the MAP-EM algorithm, the objective function $J(\mathbf{d}, \Theta)$ maximized at the M step is equal to the Q function augmented by the log prior densities of parameters. In the following, the values computed using $\hat{\Theta}$ are indicated as $\hat{\cdot}$ also.

Based on the densities defined at 3.2, the objective function is derived as

$$J(\mathbf{d}, \Theta) = \text{Const.} - \frac{MN}{2} \ln \sigma_o^2 - \frac{3M}{2} \ln \sigma_r^2 - \frac{1}{2\sigma_o^2} \sum_{j=1}^M \left\{ \sum_{i=1}^N f_t^{(ij)^2} + 2 \left(\sum_{i=1}^N f_t^{(ij)} \mathbf{w}^{(ij)\text{T}} \right) \hat{\mathbf{r}}_m^{(j)} + \text{tr} \left[\left(\sum_{i=1}^N \mathbf{w}^{(ij)} \mathbf{w}^{(ij)\text{T}} \right) \hat{\mathbf{R}}^{(j)} \right] \right\} - \frac{1}{2\sigma_r^2} \sum_{j=1}^M \text{tr} \hat{\mathbf{R}}^{(j)} - \frac{\mathbf{d}^T \mathbf{L} \mathbf{d}}{2\sigma_d^2}, \quad (9)$$

using the following definitions derived by formulating the posterior density $p(\{\mathbf{r}^{(j)}\} | \{f_t^{(ij)}\}, \hat{\mathbf{d}}, \hat{\Theta})$.

$$\hat{\mathbf{r}}_m^{(j)} \equiv \mathbb{E}[\mathbf{r}^{(j)} | \{f_t^{(ij)}\}, \hat{\mathbf{d}}, \hat{\Theta}] = -\frac{1}{\hat{\sigma}_o^2} \hat{\mathbf{V}}_r^{(j)} \sum_{i=1}^N f_t^{(ij)} \hat{\mathbf{w}}^{(ij)}, \quad (10)$$

$$\hat{\mathbf{V}}_r^{(j)} = \left(\frac{1}{\hat{\sigma}_o^2} \sum_{i=1}^N \hat{\mathbf{w}}^{(ij)} \hat{\mathbf{w}}^{(ij)\text{T}} + \frac{1}{\hat{\sigma}_r^2} \mathbf{I} \right)^{-1}, \quad (11)$$

$$\hat{\mathbf{R}}^{(j)} \equiv \mathbb{E}[\mathbf{r}^{(j)} \mathbf{r}^{(j)\text{T}} | \{f_t^{(ij)}\}, \hat{\mathbf{d}}, \hat{\Theta}] = \hat{\mathbf{V}}_r^{(j)} + \hat{\mathbf{r}}_m^{(j)} \hat{\mathbf{r}}_m^{(j)\text{T}}, \quad (12)$$

$$\mathbf{w}^{(ij)} = \begin{pmatrix} f_x^{(ij)} \mathbf{x}^{(i)} \mathbf{y}^{(i)} + f_y^{(ij)} (1 + \mathbf{y}^{(i)2}) \\ -f_x^{(ij)} (1 + \mathbf{x}^{(i)2}) - f_y^{(ij)} \mathbf{x}^{(i)} \mathbf{y}^{(i)} \end{pmatrix} + Z_0 d^{(i)} \begin{pmatrix} f_y^{(ij)} \\ -f_x^{(ij)} \end{pmatrix} \equiv \mathbf{w}_0^{(ij)} + Z_0 d^{(i)} \mathbf{w}_d^{(ij)}. \quad (13)$$

In the M step, $\{\mathbf{d}, \Theta\}$ are updated so that Eq. (9) is maximized. Eq. (9) can be rewritten as follows, ignoring constant values.

$$J(\mathbf{d}, \Theta) = -\frac{MN}{2} \ln \sigma_o^2 - M \ln \sigma_r^2 - \frac{1}{2\sigma_o^2} \hat{F}(\{d^{(i)}\}) - \frac{1}{2\sigma_r^2} \hat{G} - \frac{\mathbf{d}^T \mathbf{L} \mathbf{d}}{2\sigma_d^2}. \quad (14)$$

From this representation, σ_o^2 and σ_r^2 can be updated as

$$\sigma_o^2 = \frac{\hat{F}(\{d^{(i)}\})}{MN}, \quad \sigma_r^2 = \frac{\hat{G}}{2M}. \quad (15)$$

For $\mathbf{d}$, the partial derivative for each $d^{(i)}$ in the last term of Eq. (9) contains the adjacent surrounding d s. Therefore, to update $\mathbf{d}$, we need to solve the simultaneous equations. To avoid this, use the One-Step-Late (OSL) method [20]. That is, consider the surrounding d s as a constant and evaluate it with the current estimate $\hat{\mathbf{d}}$ s. This allows each $d^{(i)}$ to be updated individually as follows:

$$d^{(i)} = \frac{\hat{\sigma}_o^2 \hat{d}^{(i)}}{\sigma_d^2 Z_0^2 \sum_{j=1}^M \text{tr}(\mathbf{A}^{(ij)} \hat{\mathbf{R}}^{(j)}) + \hat{\sigma}_o^2 - \frac{\sigma_d^2 Z_0 \sum_{j=1}^M \left\{ f_t^{(ij)} \mathbf{w}_d^{(ij)\text{T}} \hat{\mathbf{r}}_m^{(j)} + \text{tr}(\mathbf{B}^{(ij)} \hat{\mathbf{R}}^{(j)}) \right\}}{\sigma_d^2 Z_0^2 \sum_{j=1}^M \text{tr}(\mathbf{A}^{(ij)} \hat{\mathbf{R}}^{(j)}) + \hat{\sigma}_o^2}}, \quad (16)$$

where σ_o^2 is also evaluated by the current estimate, and the matrices $\mathbf{A}^{(ij)}$ and $\mathbf{B}^{(ij)}$ are defined as

$$\mathbf{A}^{(ij)} \equiv \mathbf{w}_d^{(ij)} \mathbf{w}_d^{(ij)\top}, \quad (17)$$

$$\mathbf{B}^{(ij)} \equiv \left(\mathbf{w}_d^{(ij)} \mathbf{w}_0^{(ij)\top} + \mathbf{w}_0^{(ij)} \mathbf{w}_d^{(ij)\top} \right) / 2, \quad (18)$$

and $\bar{d}^{(i)}$ indicates the local mean using a four-neighboring system that does not include $d^{(i)}$.

3.4 Numerical experiments of differential-type method

In order to confirm the function of the proposed method, we conducted numerical experiments using artificial images. **Figure 2(a)** shows the original image generated using the depth map shown in **Figure 2(b)**. The image size is 128×128 pixels, which is equivalent to $-0.5 \leq x, y \leq 0.5$ measured in focal length units. In **Figure 2(b)**, the vertical axis shows the depth Z in units of focal length, and the horizontal axis shows the pixel position on the image plane.

In our model, the successive image pairs are used in turn to calculate f_t . This study ignores the drift component of fixation eye movements and assumes that there is no temporal divergence of the range of motion at each image position. Therefore, each rotation sampled as a Gaussian independent random variable is considered to be relative to the initial image shown in **Figure 2(a)**. We can think of a gradient equation that holds between the resulting image and the initial image. Additionally, in order to firstly justify our algorithm for the assumed statistical models, we computed $\{f_t\}$ using Eq. (6) with true values of $\mathbf{r}$ and $\{d\}$ and use

Figure 2.
Example of the data used in the experiments: (a) artificial image; (b) true depth map (reprinted from [13]).

them for depth recovery. The performance evaluation for f_t , f_x , and f_y actually measured from the image will be explained as the result of the real image experiment in Section 5.

We executed the proposed algorithm using $\sigma_r^2 = 10^{-4}$. Under this condition, the average size of the optical flow is about one pixel, which is sufficiently small compared to the size of the intensity pattern in the **Figure 2(a)**, which meets the conditions of the gradient method. A Gaussian random value with a zero mean and with a deviation corresponding to 1% of the mean of $\{f_t\}$ was added to the $\{f_t\}$ which completely satisfies Eq. (6) as observation noise. We evaluated the effectiveness of the smoothness constraint introduced by Eq. (8) by performing a depth recovery by varying the value of σ_d^2 . The initial values of both σ_o^2 and σ_r^2 were 1.0×10^{-2} as arbitrary values, and $\{d\}$ was assumed initially as a plane of $Z = 9.0$. Examples of the results with $M = 100$ are shown in **Figure 3**. In addition, we varied M for each σ_d^2 to see the effectiveness of using many observations caused by small camera rotations together. The RMSE of the recovered depth for the values of σ_d^2 and M is organized in the **Table 1**. From these results, we can see that the smoothness constraint is important to reduce the degrees of freedom of d . However, over-applying this constraint will increase the recovery error because the recovered depth map will be too smooth. Note that the scales of the Z axes in **Figure 3(a)-(c)** are different. We can also see that as the number of camera rotations increases, observation collection works well to improve recovery accuracy. In the future, it is desirable that σ_d^2 be estimated adaptively for each pixel or local region. We plan to treat it as a stochastic unknown variable and formulate it in the framework of variational Bayes scheme.

Figure 3. Results of recovered depth maps with $M = 100$: σ_d^2 is (a) $\sigma_o^2 \times 10^{-1}$; (b) $\sigma_o^2 \times 10^{-3}$; (c) $\sigma_o^2 \times 10^{-5}$ (reprinted from [13]).

σ_d^2/σ_o^2	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
$M = 50$	0.6983	0.3948	0.2097	0.0945	0.3699
$M = 100$	0.4741	0.3079	0.1841	0.0769	0.2124

Table 1. RMSE of recovered depth.

4. Integral-type method

4.1 Image motion blurring related to depth

From Eqs. (2) and (3), and the probabilistic characteristics of $\mathbf{r}^{(j)}$, $\mathbf{v}$ is also a 2-D Gaussian random variable with $E[\mathbf{v}] = \mathbf{0}$ and the variance–covariance matrix of

$$\mathbf{V}[\mathbf{v}] = \begin{bmatrix} x^2y^2 + (1 + x^2 + Z_0d)^2 & 2xy\left(1 + \frac{x^2 + y^2}{2} + Z_0d\right) \\ 2xy\left(1 + \frac{x^2 + y^2}{2} + Z_0d\right) & x^2y^2 + (1 + y^2 + Z_0d)^2 \end{bmatrix} \sigma_r^2. \quad (19)$$

This equation can be seen as a function of the inverse depth d , so if we know the variance–covariance matrix at each pixel position, we can calculate the depth map.

There are some schemes to obtain the variance–covariance matrix of optical flow defined by Eq. (19) locally at each image position from multiple images observed through random camera rotations imitating tremor. The simplest and most natural way is to first detect the optical flow from the image and then calculate its quadratic statistic. However, here we are considering a case where the intensity pattern is fine with respect to the temporal sampling rate and it is difficult to accurately detect the optical flow. Therefore, we adopt an integral-formed scheme in which the variance–covariance matrix is calculated as the distribution of local image blur.

We define an averaged image brightness $f_{ave}(\mathbf{x})$ as an arithmetic average of observed M images $\{f_j(\mathbf{x})\}_{j=1, \dots, M}$ with fixational eye movements. If M is asymptotically large, the following convolution holds using locally defined a two-dimensional Gaussian point spread functions $g_x(\cdot)$ and an original image brightness $f_0(\mathbf{x})$.

$$f_{ave}(\mathbf{x}) = \int_{\mathbf{x}' \in \mathcal{R}} g_x(\mathbf{x}') f_0(\mathbf{x} - \mathbf{x}') d\mathbf{x}', \quad (20)$$

where $\mathbf{x}$ indicates the image position, $\mathcal{R}$ is a local region around $\mathbf{x}$, and $g_x(\cdot)$ has a variance-covariance matrix indicated by Eq. (19). Furthermore, it is assumed that $\int g_x(\mathbf{x}') d\mathbf{x}' = 1$ is satisfied.

As explained above, we model $f_{ave}(\mathbf{x})$ as a blurred image due to fixation eye movements. The difference in the degree of blur of $f_{ave}(\mathbf{x})$ indicates the difference in depth. In other words, the closer the imaging target is to the camera, the greater this image blur.

4.2 Direct algorithms for integral-type method

In the two-step algorithm, after detecting the variance–covariance matrix of the image blur shown in Eq. (19), the maximum likelihood estimation of d using that as the observable is analytically possible [15]. However, the solution is not optimal because it observes indirect quantities. With optimality in mind, we need to use the direct method of directly estimating the depth map without determining $g_x(\cdot; d)$. This strategy usually requires a numeric search or iterative update [16]. We constructed two algorithms as a direct method, each adopting local optimization and global optimization respectively.

A. Local optimization algorithm

We assume that a depth value in a local region $\mathcal{L}$ around each $\mathbf{x}$ is constant. Based on the minimum least square criterion, the objective function can be defined with respect to the depth corresponding to each pixel.

$$J_L(d(\mathbf{x})) \equiv \int_{\mathcal{L}} (f_m(\mathbf{x} - \mathbf{x}'') - f_{conv}(\mathbf{x} - \mathbf{x}''))^2 d\mathbf{x}'', \quad (21)$$

$$f_{conv}(\mathbf{x}) \equiv \int_{\mathcal{R}} g\mathbf{x}(\mathbf{x}'; d) f_0(\mathbf{x} - \mathbf{x}') d\mathbf{x}'. \quad (22)$$

We can recover the depth individually for each pixel by minimizing this function with respect to the depth corresponding to each pixel. Therefore, simultaneous multivariate optimization is not required and one-dimensional numerical search can be adopted.

B. Global optimization algorithm

By assuming a spatially smooth depth map in the solution, we can define the following objective function based on regularization theory of Poggio et al. [21].

$$J_G(d(\mathbf{x})) = (1 - \lambda) \int (f_m(\mathbf{x}) - f_{conv}(\mathbf{x}))^2 d\mathbf{x} + \lambda \int \left\{ \left(\frac{\partial d(\mathbf{x})}{\partial x} \right)^2 + \left(\frac{\partial d(\mathbf{x})}{\partial y} \right)^2 \right\} d\mathbf{x}, \quad (23)$$

where λ is a parameter for adjusting the degree of constraint on the smoothness of the depth map, and the integration of Eq. (23) is performed for the entire image. From the variational principle, the Euler–Lagrange equation for $d(\mathbf{x})$ is derived using $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$ as follows:

$$\nabla^2 d = -\frac{1 - \lambda}{\lambda} (f_m - f_{conv}) \frac{\partial f_{conv}}{\partial d}. \quad (24)$$

For discrete computation, we can approximate the smoothness constraint in Eq. (23) using (i, j) as a description of an image position.

$$\begin{aligned} \left(\frac{\partial d(\mathbf{x})}{\partial x} \right)^2 + \left(\frac{\partial d(\mathbf{x})}{\partial y} \right)^2 &\approx \frac{1}{5} \left\{ (d_{i+1,j} - d_{i,j})^2 + (d_{i,j+1} - d_{i,j})^2 \right\} \\ &+ \frac{1}{20} \left\{ (d_{i+1,j-1} - d_{i,j})^2 + (d_{i+1,j+1} - d_{i,j})^2 \right\}. \end{aligned} \quad (25)$$

Using Eq. (25) and the discrete representation of Eq. (24), we can minimize Eq. (23) by the following iterative formulation with an iteration number of n .

$$d_{i,j}^{(n+1)} = \bar{d}_{i,j}^{(n)} + \frac{1 - \lambda}{\lambda} \left(f_{m,i,j} - f_{conv}(d_{i,j}^{(n)}) \right) \frac{\partial f_{conv}(d_{i,j}^{(n)})}{\partial d}, \quad (26)$$

$$\bar{d}_{i,j}^{(n)} = \frac{1}{5} \left(d_{i+1,j}^{(n)} + d_{i,j+1}^{(n)} + d_{i-1,j}^{(n)} + d_{i,j-1}^{(n)} \right) + \frac{1}{20} \left(d_{i+1,j+1}^{(n)} + d_{i+1,j-1}^{(n)} + d_{i-1,j-1}^{(n)} + d_{i-1,j+1}^{(n)} \right). \quad (27)$$

4.2.1 Numerical experiments of integral-type method

The proposed algorithm assumes that the definition of the motion blur image in Eq. (20) holds. To observe such ideal motion blur, it takes enough exposure time for

imaging. Here, we use artificial images to examine the characteristics of the proposed algorithm with respect to the relationship between the size of image motion and the fineness of intensity texture.

We artificially generate motion blur images. First, a true depth map is set up, and a large number of images are generated by a computer graphics technique that randomly samples r according to the Gaussian distribution in Eq. (4). By averaging these images, we can artificially generate motion blur images. Input motion blur images averaged 10,000 images to mimic analog motion blur, **Figure 4** shows an example of a reference (original) image and a true inverse depth map. The image size is 256×256 pixels, which is equivalent to $-0.5 \leq x, y \leq 0.5$ measured in focal length units.

Figure 4.

Example of the artificial data used in the experiments: (a) original image; (b) true inverse depth map used for generating the blurred image (reprinted from [16]).

Figure 5.

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) $\lambda = 0.6$ (reprinted from [16]).

Figure 6.
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) $\lambda = 0.6$ (reprinted from [16]).

Figure 7.
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) $\lambda = 0.6$ (reprinted from [16]).

Local optimization algorithms (LOA) are computationally expensive, and global optimization algorithms (GOA) are slow to converge. Therefore, we considered a hybrid algorithm that uses LOA sparsely to obtain the initial value of GOA. For the initial value of LOA, use the plane corresponding to the background of **Figure 4(b)**. To make a rough estimate, LOA uses blocks of 41×41 pixels as $\mathcal{L}$ in Eq. (21) and applies LOA once to each block. On the other hand, the size of $\mathcal{R}$ in Eq. (22) was adaptively determined according to the depth value updated by the optimization process. Therefore, $\mathcal{R}$ took a different size at each position in the image. We assumed a square area for $\mathcal{R}$. The length of that side was 10 times the larger of the two deviations of $g\mathbf{x}(\cdot; d)$ along x axis and y axis. These can be evaluated using Eq. (19).

The depth restoration simulation was executed while changing the camera rotation size σ_r . The recovered inverse depth map is shown in **Figures 5–7**. GOA uses various values of λ . The relationship between the root mean square error (RMSE) of the recovered depth map and the value of λ is also shown in **Figure 8**. From **Figure 5**, we can see that it is not suitable for depth recovery because it is difficult to accurately measure the motion blur of the image position with a small rotation of the camera. From **Figure 8(a)**, small rotations do not provide sufficient measurement information, so to reduce the RMSE of the recovered depth map, the smoothness constraint indicated by λ is strongly required. On the contrary, when the rotation is large, the point image distribution function becomes wider than the spatial change of the target shape. Therefore, the Gaussian function with the variance–covariance matrix in Eq. (19) is inappropriate, the motion blur recognized by this model is smoother than the actual blur of the image, and a depth recovery error occurs. This can be confirmed from the RMSE value in **Figure 8(c)**. Since the

Figure 8.

Relation between RMSE of recovered depth and lambda: (a) $\sigma_r = 0.06$; (b) $\sigma_r = 0.08$; (c) $\sigma_r = 0.1$ (reprinted from [16]).

smooth depth tends to be recovered from the recognized smooth motion blur from **Figure 8(c)**, it can be confirmed that the smoothness constraint of Eq. (23) is an obstacle to the reduction of RMSE.

5. Real image experiments of differential-type method

5.1 Selective use of image pairs to improve accuracy

When applying the difference-type method to an actual image and checking the actual performance, the performance was improved by selecting the image pair used for depth restoration. We have adopted a scheme that excludes image pairs that are expected to have large approximation errors in the gradient equation on a pixel-by-pixel basis. We can use the inner product of the spatial gradient vectors of consecutive image pairs to select image pairs that do not cause aliasing problems. For each pixel, the image pairs of which the sign of the inner product $\mathbf{f}_s^{(ij)T} \mathbf{f}_s^{(i,j-1)}$ is negative are discarded. It is noted that $\mathbf{f}_s^{(ij)} = [f_x^{(ij)}, f_y^{(ij)}]^T$.

In the next step, from the image pairs remained by the above decision, we additively select the suitable image pairs at each pixel by estimating the amount of the higher order terms included in the observation of f_t . f_t is exactly represented as follows:

$$f_t = -f_x v_x - f_y v_y - \frac{1}{2} \{ f_{xx} v_x^2 + f_{yy} v_y^2 + 2f_{xy} v_x v_y \} + \dots \quad (28)$$

After discarding a bad image pair, the higher-order terms can be considered small. In this case, the quadratic term in Eq. (28) can be estimated for each pixel i as follows:

$$-\frac{1}{2} \{ (f_x^{(ij)} - f_x^{(i,j-1)}) v_x^{(ij)} + (f_y^{(ij)} - f_y^{(i,j-1)}) v_y^{(ij)} \}. \quad (29)$$

We can define a measure for estimating the equation error as the ratio of this higher order term to the first order term.

$$J = \frac{|(f_x^{(ij)} - f_x^{(i,j-1)}) v_x^{(ij)} + (f_y^{(ij)} - f_y^{(i,j-1)}) v_y^{(ij)}|}{2|f_x^{(ij)} v_x^{(ij)} + f_y^{(ij)} v_y^{(ij)}|}. \quad (30)$$

This measurement depends on the direction of the optical flow but is invariant with respect to the amplitude of the optical flow. To calculate the value of J , we need to know the true value of the optical flow. By examining the details of J , even if the difference of the spatial gradient $\mathbf{f}_s^{(ij)} - \mathbf{f}_s^{(i,j-1)}$ is large, when the direction of $\mathbf{f}_s^{(ij)} - \mathbf{f}_s^{(i,j-1)}$ is perpendicular to that of optical flow, the equation error becomes small. Therefore, the value $|\mathbf{f}_s^{(ij)} - \mathbf{f}_s^{(i,j-1)}|/|\mathbf{f}_s^{(ij)}|$ can be used as the worst value. In the following, the image pairs for which $|\mathbf{f}_s^{(ij)} - \mathbf{f}_s^{(i,j-1)}|/|\mathbf{f}_s^{(ij)}|$ is less than the certain threshold value are selected at each pixel to be used for depth recovery.

5.2 Camera system implementation

We built the camera hardware system for examining the practical performance of our camera model shown in **Figure 1**. The implemented camera system is shown in **Figure 9**.

Figure 9.
Camera system implemented for tremor rotations.

The camera system can be rotated around the horizontal axis i.e. X axis and around the vertical axis, i.e. Y axis. The rotation around the optical direction, i.e. Z direction, cannot be performed, which is not needed to gain the depth information. The parameters of the system are shown as follows: focal length is 2.8 – 5.0 mm, image size is $1,200 \times 1,600$ pix., movable widths are 360 deg. for X axis and $(-10, +10)$ deg. for Y axis, and drivable minimum units are 1 pulse = 0.01 deg. for X -axis and 1 pulse = 0.00067 deg. for Y -axis.

5.3 Experimental results

In this section, we explain the results of the experiments using the real images captured by the developed camera system [22]. Our camera system has a parallel stereo function. That is, the camera can be moved laterally by the slide system. Prior to the experiment, we calibrated the camera's internal parameters, including focal length and Z_0 , using the method in [23] and stereo calculations. The image used in the experiment is grayscale, consists of 256×256 pixels, and is 8-bit digitized. An example is shown in **Figure 10(a)**. The true inverse depth of the target object is shown in **Figure 10(b)**. It was measured in parallel stereo above using a two-plane model. In this figure, the horizontal axis shows the position in the image plane, and the vertical axis shows the inverse depth in units of focal length. We captured 100 images. The maximum number of iterations of the MAP-EM

Figure 10.
 Data for experiments: (a) example of captured image, (b) true inverse depth of object (reprinted from [22]).

Figure 11.
 Profiles of cross-section of recovered inverse depth: (a) all image pairs are used (100%), (b) threshold $\times 1.5$ (94% image pairs were used), (c) threshold $\times 1.25$ (86%), (d) threshold $\times 1$ (68%), (e) threshold $\times 0.75$ (62%), (f) threshold $\times 0.5$ (62%) (reprinted from [22]).

algorithm was set to 600. Within this number of iterations, the iterations of almost all experiments converged. σ_d^2 is heuristically determined. The average value of $|f_s^{(i,j)} - f_s^{(i,j-1)}|/|f_s^{(i,j)}|$ explained in the previous section with respect to all pixels was

defined for each image pair as a standard magnification ($\times 1$) of the threshold for selecting the suitable image pairs. Namely, by decreasing the threshold magnification, we can discard more image pairs. Conversely, by increasing the magnification, many image pairs can be used for recovery. Because of the limit of pages, we only show the results with $\sigma_r^2 = 2.64 \times 10^{-2}$ by which the average of the optical flow's amplitude approximately coincides with $\lambda/4$.

Figure 11 shows the result of the recovered depth for each threshold set as a constant multiple of the reference value. We also looked at the results using all image pairs. From these results, it can be confirmed that by reducing the magnification, inappropriate image pairs can be discarded and the accuracy of depth recovery is improved. The percentage shown in the caption of the figure shows the number of image pairs used for recovery, which is determined in conjunction with the change in threshold.

6. Conclusions

In this chapter, we introduced a depth recovery algorithms that uses large number of images with small movements by using camera motion that simulates fixational eye movements, especially the tremor component. The algorithms can be divided into a differential-type and an integral-type. For the differential-type, it is desirable that the movement on the image is relatively small with respect to the texture pattern of the surface to be imaged, and conversely, for the integral-type, it is appropriate to apply it to a fine texture compared to the movement on the image. Therefore, ideally, the development of a depth recovery system in which both schemes function adaptively and selectively according to the target texture is the most important task in the future.

A detailed technical issue is to automatically determine the parameters that control the smoothness of the depth. This can be achieved by considering all unknowns as stochastic variables and formulating them in the variational Bayesian framework. As for the integration method, since the resolution of the recovered depth is low in principle, it is possible to consider a composite type in which the differential-type is applied again and refinement is performed on the result obtained by the integral-type.

So far, we have considered a method that assumes only tremor, but in the future, we are planning to study camera motion that also simulates drift and microsaccade. In the method for drift component, it is necessary to extend the method based on tremor to the online version, and then update the depth estimate while advancing the tracking of the target as time series processing. When using microsaccades, it is necessary to handle large movements between frames. Therefore, based on the correspondence of feature points, sparse but highly accurate depth restoration can be expected. Drift itself does not have much merit in its use, but it plays an important role in generating microsaccades. As described above, we believe that an interesting system can be realized by comprehensively using the three components.

On the other hand, stereoscopic vision and motion stereoscopic vision are difficult to handle objects with few textures. In [24], we proposed a stereo system that considers shading information. The projected images to both cameras are calculated by computer graphics technique while changing the depth estimation value. The depth is determined so that the generated image matches the image observed by each camera. As a result, the association between images is indirectly realized. By introducing this method, it becomes possible to handle textureless objects. We aim to develop a comprehensive depth restoration method, including the

multi-resolution processing proposed in [12]. In another scheme that deals with the textureless region in stereo vision, the region where the depth value is constant or changes smoothly, called the support region, is adaptively determined [25]. We will also consider whether the relationship between image changes due to tremor and microsaccade can be used for adaptive determination of this support region.

In recent years, many realizations of stereoscopic vision and motion stereoscopic vision by deep learning have been reported [26–28]. And the relationship with the conventional method based on mathematical formulas is often questioned. The deep learning method is hampered by the addition of a large number of images and annotations to them. Although unsupervised learning is often devised, the solution is often limited. Therefore, even if the conventional method is rather complicated and takes time, if a method capable of more precise depth recovery is constructed, it can be used for annotation calculation of deep learning. This can be understood as copying the conventional method to deep neural network (DNN). DNN takes time to learn, but has the advantage of being able to infer at high speed. In this way, it is important that both schemes develop in a two-sided relationship.

Appendix

Here, the method of calibrating the axis of rotation is explained using **Figure 12**. Let a point in 3-D space be $\mathbf{X}_1 = [X_1, Y_1, Z_1]^T$ in the coordinate system before camera rotation and $\mathbf{X}_2 = [X_2, Y_2, Z_2]^T$ in the coordinate system after rotation, and the coordinates of the corresponding points on the image be $\mathbf{x}_1 = [x_1, y_1, z_1]^T$ and $\mathbf{x}_2 = [x_2, y_2, z_2]^T$, respectively. Similarly, the optical axes before and after rotation are $\mathbf{z}_1^1 = [0, 0, 1]^T$ and $\mathbf{z}_2^2 = [0, 0, 1]^T$, respectively. If the rotation is taken around the X-axis, the rotation matrix is given by the following equation.

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}. \quad (31)$$

The translation $\mathbf{T}$ of the lens center generated by this rotation is given by the following equation in the coordinate system before rotation.

Figure 12.
 Explanation of rotation axis calibration.

$$\mathbf{T}^1 = Z_0 \mathbf{z}_2^1 - Z_0 \mathbf{z}_1^1 = Z_0(\mathbf{R} - \mathbf{I}) \mathbf{z}_1^1 \equiv Z_0 \mathbf{S} \mathbf{z}_1^1, \quad (32)$$

where $\mathbf{z}_2^1$ represents the optical axis after rotation in the coordinate system before rotation. In addition, $\mathbf{X}_1$ and $\mathbf{X}_2$ have the following relationship.

$$\mathbf{X}_2 = \mathbf{R}^T(\mathbf{X}_1 - \mathbf{T}^1) \rightarrow \mathbf{R} \mathbf{X}_2 = \mathbf{X}_1 - \mathbf{T}^1 \quad (33)$$

Furthermore, by substituting the relation of $\mathbf{x}_1 = \mathbf{X}_1^1/Z_1$, $\mathbf{x}_2 = \mathbf{X}_2^2/Z_2$ into Eq. (33), the following equation is obtained.

$$Z_2 \mathbf{R} \mathbf{x}_2 = \mathbf{X}_1 - Z_0 \mathbf{S} \mathbf{z}_1^1. \quad (34)$$

By expressing this equation in terms of components and organizing it, the following two equations are derived.

$$Z_2(y_2 \cos \theta - \sin \theta) = Y_1 + Z_0 \sin \theta, \quad (35)$$

$$Z_2(y_2 \sin \theta + \cos \theta) = Z_1 - Z_0(\cos \theta - 1). \quad (36)$$

By substituting Eq. (35) into Eq. (36), the solution of Z_0 can be derived as follows:

$$Z_0 = \frac{Z_1(y_2 \cos \theta - \sin \theta) - Y_1(y_2 \sin \theta + \cos \theta)}{\sin \theta(y_2 \sin \theta + \cos \theta) + (\cos \theta - 1)(y_2 \cos \theta - \sin \theta)}. \quad (37)$$

Author details

Norio Tagawa

Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo, Japan

*Address all correspondence to: tagawa@tmu.ac.jp

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] Martinez-Conde S, Macknik S L, and Hubel D. The role of fixational eye movements in visual perception. *Nature Reviews*. 2004;5:229–240.
- [2] Stemmler M. A single spike suffices: the simplest form of stochastic resonance in model neuron. *Network: Computations in Neural Systems*. 1996; 61:687–716.
- [3] Propokopowicz P and Cooper P. The dynamic retina. *Int. J. Computer Vision*. 1995;16:191–204.
- [4] Hongler M-O, de Meneses Y L, Beyeler A, and Jacot J. The resonant retina: Exploiting vibration noise to optimally detect edges in an image. *IEEE Trans. Pattern Anal. Machine Intell*. 2003;25:1051–1062.
- [5] Ando S, Ono N, and Kimachi A. Involuntary eye-movement vision based on three-phase correlation image sensor. In: *Proceedings on 19th Sensor Symposium*; 2002. p. 83–86.
- [6] Lazaros N, Sirakoulis G-C, and Gasteratos A. Review of stereo vision algorithm: from software to hardware. *Int. J. Optomechatronics*. 2008;5:435–462.
- [7] Wang J and Zickler T. Local detection of stereo occlusion boundaries. In: *Proceedings on CVPR*; 2019. p. 3818–3827.
- [8] Liu F, Zhou S, Wang Y, Hou G, Sun Z, and Tan T. Binocular light-field: imaging theory and occlusion-robust depth perception application. *IEEE Trans. Image Process*. 2019;29: 1628–1640.
- [9] Horn B K P and Schunk B. Determining optical flow. *Artif. Intell*. 1981;17:185–203.
- [10] Simoncelli E P. Bayesian multi-scale differential optical flow. In: *Jahne B, Haubecker H, Geibier P, editors. Handbook of Computer Vision and Applications*. Academic Press; 1999. vol. 2. p. 397–422.
- [11] Bruhn A and Weickert J. Locas/kanade meets horn/schunk: combining local and global optic flow methods. *Int. J. Computer Vision*. 2005;61:211–231.
- [12] Tagawa N, Kawaguchi J, Naganuma S, and Okubo K. Direct 3-d shape recovery from image sequence based on multi-scale bayesian network. In: *Proceedings on ICPR*; 2008. p. CD-ROM.
- [13] Tagawa N. Depth Perception model based on fixational eye movements using Bayesian statistical inference. In: *Proceedings on ICPR*; 2010. p. 1662–1665.
- [14] Tagawa N and Alexandrova T. Computational model of depth perception based on fixational eye movements. In: *Proceedings on VISAPP*; 2010, p. 328–333.
- [15] Tagawa N, Iida Y, and Okubo K. Depth perception model exploiting blurring caused by random small camera motions. In: *Proceedings on VISAPP*; 2012, p. 329–334.
- [16] Tagawa N, Koizumi S, and Okubo K. Direct depth recovery from motion blur caused by random camera rotations imitating fixational eye movements. In: *Proceedings on VISAPP*; 2013, p. 177–186.
- [17] Sorel M and Flusser J. Space-variant restoration of images degraded by camera motion blur. *IEEE Trans. Image Processing*. 2008;17:105–116.
- [18] Paramanand C and Rajagopalan A N. Depth from motion and optical blur with unscented Kalman filter. *IEEE Trans. Image Processing*. 2012;21:2798–2811.

- [19] Dempster A-P, Laird N-M, and Rubin D-B. Maximum likelihood from incomplete data. *J. Roy. Statist. Soc. B.* 1977;39:1–38.
- [20] Green P.-J. On use of the Em algorithm for penalized likelihood estimation. *J. Roy. Statist. Soc. B.* 1990; 52:443–452.
- [21] Poggio T, Torre V, and Koch C. Computational vision and regularization theory. *Nature.* 1985;317:314–319.
- [22] Tsukada S, Ho Y, Tagawa N, and Okubo K. Accuracy improvement for depth from small irregular camera motions and its performance evaluation. In: *Proceedings on ICIAR*; 2015, p. 306–315.
- [23] Zhang Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Machine Intell.* 2000;22: 1330–1334.
- [24] Wakabayashi K, Tagawa N, and Okubo K. Shape from multi-view images based on image generation consistency. In: *Proceedings on VISAPP*; 2013, p. 334–340.
- [25] Wu W, Zhu H, Yu S, and Shi J. Stereo matching with fusing adaptive support weights. *IEEE Access.* 2019;7: 61960–61974.
- [26] Laina I, Rupprecht C, Belagiannis V, Tombari F, and Navab N. Deeper depth prediction with fully convolutional residual networks. In: *Proceedings on 3D Vision*;2016, p. 239–248.
- [27] Ummenhofer B, Zhau H, Uhrig J, Mayer N, Ilg E, Dosovitskiy A, and Brox T. DeMoN: Depth and motion network for learning monocular stereo. In: *Proceedings on CVPR*;2017, p. 5038–5047.
- [28] Chen P-Y, Liu A H, Liu Y-C, Wang Y-C F. Towards scene understanding: unsupervised monocular depth estimation with semantic-aware representation. In: *Proceedings on CVPR*;2019, p. 2624–2632.