DISPARITY ESTIMATION FROM STEREO IMAGES WITH MULTILAYERED REACTION-DIFFUSION MODELS OF ACTIVATION-INHIBITION MECHANISM

Atsushi Nomura

Faculty of Education
Yamaguchi University
Yoshida 1677-1, Yamaguchi 753-8513, Japan

Makoto Ichikawa, Hidetoshi Miike *

Faculty of Engineering
Yamaguchi University
Tokiwadai 2-16-1, Ube 755-8611, Japan

ABSTRACT

The present paper proposes a method that estimates a stereo disparity map. The method consists of multilayered networks of reaction-diffusion equations having activator and inhibitor variables. A particular set of equations describes non-linear oscillators coupled with diffusion processes, and governs a disparity layer. Estimating a stereo disparity map requires the two main constraints: the smoothness constraint and the uniqueness. The activation process of the diffusion-coupled oscillators performs the filling-in process for a stereo disparity map as the smoothness constraint. The mutual inhibition mechanism among the multilayered networks retains the uniqueness constraint of the disparity at a particular pixel site. By applying the proposed equations to cross-correlation functions derived from stereo images, we obtain a self-organized stereo disparity map. Experimental results show that the proposed method is superior to a previous one, in particular, for real stereo images.

1. INTRODUCTION

Estimating a disparity map from stereo images requires two main constraints. One of them is the smoothness constraint and the other is the uniqueness one. With the smoothness constraint we assume that a stereo disparity map does not have spatial discontinuities; with the the uniqueness one we assume that a particular pixel site has only one disparity level.

Marr and Poggio proposed the above mentioned two constraints on a stereo disparity map, and also proposed a very simple computational model that iteratively estimates a stereo disparity map [1]. Their model consists of multilayered disparity networks, each of which has "cells" excitatory connected in a spatial neighborhood and inhibitory connected among the network layers. The excitatory connection works as the filling-in process as the smoothness constraint, and the inhibitory connection performs the uniqueness constraint. They applied their model to artificially generated random-dots stereo images, which was proposed by Julesz [2].

*The present study was partly supported by the Grant-in-Aid for Scientific Research of Japan.

We have proposed a model of grouping process on a visual feature such as orientation of a short line [3]. The model consists of multilayered networks of diffusion-coupled activation-inhibition mechanism [4, 5]; each of the networks governs a group of a visual feature. In our previous study, we simply applied our model of the grouping process to stereo images [6]. This trial application provides satisfactory results for random-dots stereo images having only several disparity levels.

The present paper proposes a method that estimates a disparity map from stereo images for the analysis of real stereo images. The proposed method utilizes our previously proposed model of the grouping process. For applying the model of the grouping process to the problem of stereo disparity estimation, we take account of the uniqueness constraint on stereo disparity as well as the smoothness constraint. The mutual inhibition mechanism developed here performs the uniqueness constraint and the diffusion-coupled activation process does the smoothness constraint. By utilizing the proposed model and the previous one proposed by Marr and Poggio, we analyzed random-dots stereo images having curved surface and two kinds of real stereo images. The results of the analysis show that the performance of the proposed model is well, compared to the previous one, in particular, for real stereo images.

2. PREVIOUS METHOD

Marr and Poggio proposed a model that estimates disparity from stereo images [1]. They considered multilayered networks consisting of "cells" located at particular pixel sites (x, y). The state $S^t(x, y, d)$ of a particular cell on a network layer expresses existence or non-existence of a disparity level d. Their following model is in the iterative form of S^t at the tth step,

$$S^{t+1}(x,y,d) = \sigma\left(\sum_{\Omega_+} S^t - \varepsilon \sum_{\Omega_-} S^t + S^0\right), \quad (1)$$

where S_0 denotes the initial state and corresponds to the cross-correlation function between the stereo images overlapped at the disparity level d. The function $\sigma(s)$ denotes the switching function that returns 1 if $s \geq \theta$ and 0 if $s < \theta$. The domain Ω_+ refers to the excitatory spatial neighborhood having $M \times M$ (pixel), and Ω_- refers to the inhibitory neighborhood. The summation $\sum_{\Omega_+} S^t$ performs the filling-in process for the smoothness constraint; the summation $\sum_{\Omega_-} S^t$ does the uniqueness constraint due to the inhibitory mechanism, where ε is an inhibitory constant.

3. PROPOSED METHOD

The following set of ordinary differential equations having an activator variable $u(t)$ and an inhibitor one $v(t)$ describes non-linear oscillation [4, 5].

$$\begin{aligned} u' &= f(u,v) = \frac{1}{\epsilon}\left[u(1-u)(u-a) - v\right] \\ v' &= g(u,v) = u - bv \end{aligned} \tag{2}$$

Equation 2 has two-stable equilibrium points, depending on the constants a and b. The constant ϵ is $0 < \epsilon \ll 1$. A set of solutions $(u, v) = (0, 0)$ is one of the stable equilibrium points. On the one hand, when we stimulate the variable u at the stable equilibrium point by adding $\delta u > a$, according to $f(u, v) > 0$, the variable u is rapidly increasing as time proceeds. This is the activation process. On the other hand, when we stimulate the variable u at $(u, v) = (0, 0)$ by $\delta u < a$, the variable u directory returns to the initial state of $(u, v) = (0, 0)$. Therefore, the parameter a works as a threshold value, and the set of the equations works as the switching function dividing initial states into two stable equilibrium points. [The other stable equilibrium point is $(u, v) \simeq (1, 0)$.] In addition, the large variable v inhibits the activation process. Thus, the set of Eq.(2) has the activation-inhibition mechanism.

The following equations describe diffusion processes of $u(x, y, t)$ and $v(x, y, t)$ in a 2-dimensional space.

$$\partial_t u = D_u \nabla^2 u, \quad \partial_t v = D_v \nabla^2 v \tag{3}$$

The constants D_u and D_v are diffusion coefficients on u and v, respectively. When a point in a 2-dimensional space has a large value of u, compared to the surrounding regions, the large value diffuses from the point into its neighborhood points as time proceeds.

We have proposed the next set of reaction-diffusion equations combining the activation-inhibition mechanism described in Eq.(2) with the diffusion processes of Eq.(3) [3, 6].

$$\begin{aligned} \partial_t u_n &= D_u \nabla^2 u_n + f(u_n, v_n, u_{n_{\max}}) + rC(x,y,d) \\ \partial_t v_n &= D_v \nabla^2 v_n + g(u_n, v_n) \end{aligned} \tag{4}$$

In the application of the proposed equations to stereo disparity estimation, the function $C(x, y, d)$ corresponds to cross-correlation functions derived from stereo images overlapped at the horizontal shift d (pixel). Thus, we link the stereo disparity level d with the nth network layer having the set of two variables (u_n, v_n). We denote the number of disparity levels d or the number of network layers by N_d. When a state at a pixel site in a 2-dimensional space becomes an activated state having large value of u, the diffusion process on u causes the activation processes around the neighborhood pixel sites. The combination of the diffusion process and the activation one realizes the filling-in process or the smoothness constraint in a disparity layer.

For the uniqueness constraint of a stereo disparity at a particular pixel site, the present paper proposes the next mutual inhibition mechanism among the multilayered networks.

$$a(u_{n_{\max}}) = \frac{1}{4}[1 + \tanh(u_{n_{\max}} - a_0)] \times \frac{1}{2}[1 + \tanh(|d_n|)] \tag{5}$$

As mentioned above, the constant a in Eq.(2) works as a threshold value for the activation process. Thus, we modulate a according to the maximum value of u_n of other disparity layers,

$$u_{n_{\max}} = \max_{i \in \mathcal{N}} u_i, \quad |d_n| = |d_n - \arg_d(u_{n_{\max}})|, \tag{6}$$

where d_n denotes a disparity level governed in the nth layer and $\mathcal{N} = \{0, 1, \cdots, n-1, n+1, \cdots, N_d - 1\}$. When another network layer becomes an activated state, that is, when $u_{n_{\max}}$ is large, a becomes large. Thus, the state of the current nth network layer having (u_n, v_n) is inhibited, even if u_n is rather large. However, we do not apply this mutual inhibition mechanism to the neighboring network layers having similar disparity levels, since cross-correlation functions derived from stereo images tend to have similar values among neighboring network layers in real situations. Thus, the function $\tanh(|d_n|)$ in Eq.(5) weakens the mutual inhibition mechanism, where $|d_n|$ refers to an absolute disparity difference between the layer having $u_{n_{\max}}$ and the current layer.

Figure 1 shows the flow chart of estimating a disparity map from stereo images. First, we compute cross-correlation functions $C(x, y, d)$ between two stereo images overlapped at a horizontal shift d. Then, we compute the temporal developments of the proposed model Eq.(4). Finally, we estimate a stereo disparity map $D(x, y, t)$ with the weighted sum of $u_n(x, y, t)$,

$$D(x,y,t) = \sum_n d_n \times u_n(x,y,t) / \sum_n u_n. \tag{7}$$

In computing the partial differential equations of Eq.(4), we converted the equations into the discretized form by the finite difference method. Then, we solved the set of linear equations with the Gauss-Seidel scheme. The finite differences in space and time were $\delta x = \delta y = 0.2$ and $\delta t = 10^{-3}$.

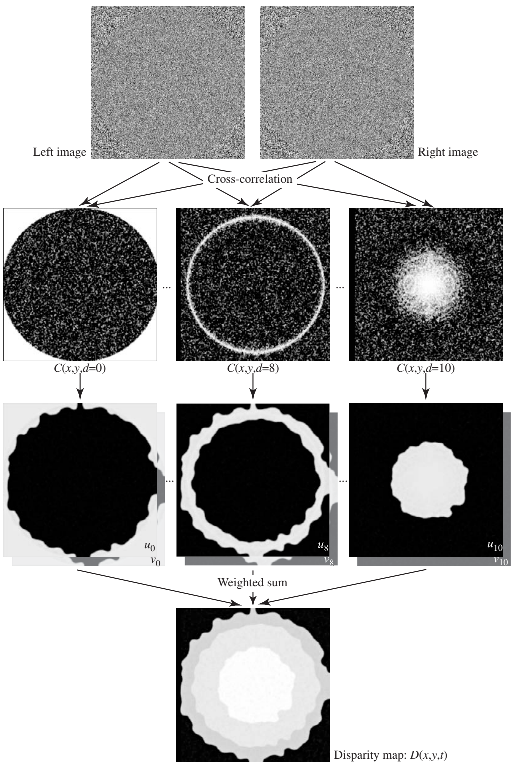

Fig. 1. Flow chart of the proposed method of disparity estimation from stereo images. A cross-correlator converts a pair of stereo images into cross-correlation functions denoted by $C(x, y, d)$. These correlation functions are provided to the proposed equations as the external stimuli in Eq.(4). As time proceeds, particular layers of $u_d(x, y, t)$ and $v_d(x, y, t)$ self-organize the domain of corresponding disparity levels. The integration (weighted sum) of the variables $u_d(x, y, t)$ provides a disparity map. The image size shown here is 300×300 (pixel). A rectangular domain for the computation of cross-correlation was 5×5 (pixel). A sphere locates in the center of the random-dots stereo images. The disparity levels was $N_d = 5$.

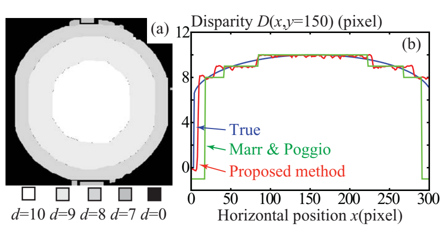

Fig. 2. Experimental results for the random-dot stereo images (see Fig.1) having curved surface of a sphere in 3-dimensional space. (a) Disparity map estimated by the previous method proposed by Marr and Poggio with the parameter values of $M = 5$ (pixel), $\theta = 5.0$ and $\varepsilon = 2.0$. (b) Comparison of 1-dimensional profiles of the stereo disparity maps estimated by the proposed model and the previous one of Marr and Poggio. Figure 1 has the disparity map estimated by the proposed method.

4. EXPERIMENTS

4.1. Random-dots stereo images

First, we applied the proposed method and the previous one proposed by Marr and Poggio [1] to the artificially generated random-dots stereo images [2] shown in Fig.1. The disparity map estimated by the proposed model was shown in the last row of Fig.1; that by the previous model was done in Fig.2(a). In addition, Fig.2(b) shows 1-dimensional profiles of the disparity maps along the center horizontal line. These results show that the two methods work well for the random-dots stereo images.

4.2. Real stereo images

We have two kinds of stereo image pairs, one of which represents four computers facing a camera as shown in Figs. 3(a) and 3(b). The computers locate at different distances from the camera. Thus, four different disparity levels must appear in a horizontal line in a disparity map. Figures 3(c) and 3(d) show two examples of cross-correlation functions derived from the

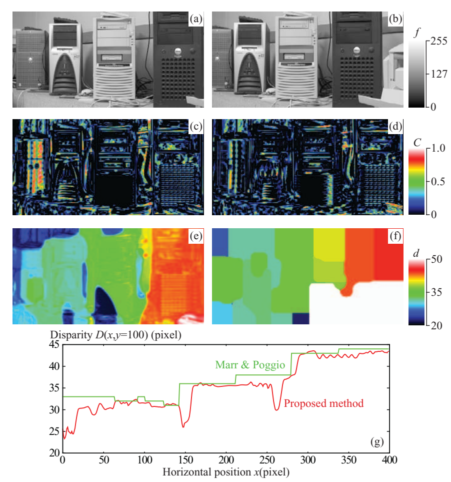

Fig. 3. Experimental results for real stereo images representing four computers. (a) Left and (b) right images. Image size is 400×200 (pixel). Brightness f was quantized into 256 levels. A rectangular domain for the computation of cross-correlation was 5×5 (pixel). The number of disparity levels is $N_d = 30$. (c) $C(x, y, d = 50)$ and (d) $C(x, y, d = 40)$. (e) Disparity map $D(x, y, t = 30)$ estimated by the proposed method with the parameter values of $D_u = 1.0$, $D_v = 2.0$, $a_0 = 0.25$, $b = 20$, $\epsilon = 1.0 \times 10^{-2}$ and $r = 2.55$. (f) Disparity map estimated by the previous model proposed by Marr and Poggio with the parameter values of $M = 15$ (pixel), $\theta = 30$ and $\varepsilon = 10$. (g) 1-dimensional profiles of the stereo disparity maps estimated by the proposed method and the previous one at $y = 100$.

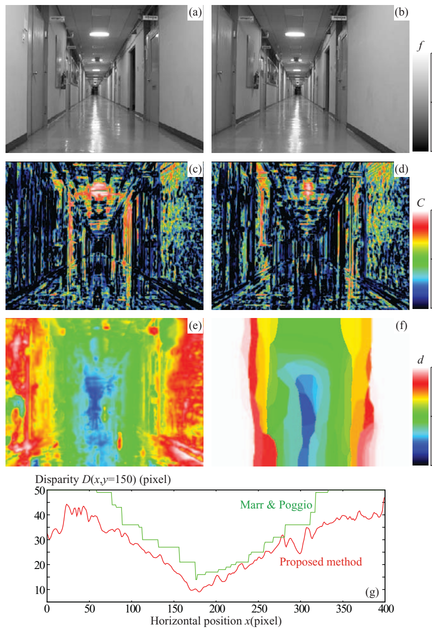

Fig. 4. Experimental results for real stereo images representing a corridor in a building. (a) Left and (b) right images. Image size is 400×300 (pixel). Brightness f was quantized into 256 levels. A rectangular domain for the computation of cross-correlation was 5×5 (pixel). The number of disparity levels is $N_d = 46$. (c) $C(x, y, d = 20)$ and (d) $C(x, y, d = 30)$. (e) Disparity map $D(x, y, t = 30)$ estimated by the proposed method with the parameter values of $D_u = 1.0$, $D_v = 2.0$, $a_0 = 0.20$, $b = 10$, $\epsilon = 3.0 \times 10^{-2}$ and $r = 2.55$. (f) Disparity map estimated by the previous model proposed by Marr and Poggio with the parameter values of $M = 19$ (pixel), $\theta = 30$ and $\varepsilon = 10$. (g) 1-dimensional profiles of the stereo disparity estimated by the proposed method and the previous one at $y = 150$.

stereo images. These functions were provided to the proposed model and the previous one of Marr and Poggio. By comparing the estimated two disparity maps shown in Figs. 3(e) and 3(f), we can recognize that both maps have well-estimated similar global structure. Figure 3(g) shows the 1-dimensional horizontal disparity profiles estimated by the two methods. The proposed method clearly reconstructed the detailed disparity structure of spatial gaps between two neighboring computes, compared to the previous method. Thus, we confirmed that the proposed method is much superior to the previous one, in particular, in detailed structure.

The other pair of real stereo images represents a corridor in a building as shown in Figs. 4(a) and 4(b). Disparity level continuously changes along the center horizontal line and the center vertical one on the image plane. Around the center of the image plane the depth becomes the largest value or the disparity becomes the smallest level. Cross-correlation functions, two samples of which are shown in Figures 4(c) and 4(d), were provided to the proposed method and the previous one. Figures 4(e) and 4(f) show the estimated disparity maps; Fig. 4(g) shows a comparison between the two maps along the center horizontal line. The disparity map estimated by the previous method has the step-wise disparity levels. In comparison to that, the proposed method successfully estimated continuous change on disparity level. The global structure was also estimated well in both of the disparity maps.

5. CONCLUSIONS

The present paper proposed a method that estimates a disparity map from stereo images. The method utilizes our previously proposed multilayered network model consisting of reaction-diffusion equations [3, 6], which has activator and inhibitor variables and describes non-linear oscillation. The combination of the activation process and the diffusion one within a network layer performs the filling-in process required for the smoothness constraint. In addition, a mutual inhibition mechanism among the multilayered networks has been developed in the present study. This mutual inhibition mechanism realizes the uniqueness constraint through a model parameter denoted by a [see Eq.(5)]. We applied the proposed method and the previous one proposed by Marr and Poggio [1] to a random-dots stereo image pair and two kinds of real ones. As the results, we have confirmed that the proposed model is much superior to the previous model, in particular, for the real stereo images.

6. REFERENCES

[1] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283–287, 1976.

[2] B. Julesz, "Binocular depth perception of computer-generated patterns," *The Bell System Tech. J.*, vol. 39, pp. 1125–1162, 1960.

[3] A. Nomura, M. Ichikawa, and H. Miike, "Realizing the grouping process with the reaction-diffusion model," *IPSJ Trans. Computer Vision and Image Media [in Japanese]*, vol. 45, pp. 26–39, June 2004.

[4] R. FitzHugh, "Impulses and physiological states in theoretical models of nerve membrane," *Biophys. J.*, vol. 1, pp. 445–455, 1961.

[5] J. Nagumo, S. Arimoto, and S. Yoshizawa, "An active pulse transmission line simulating nerve axon," *Proc. IRE*, vol. 50, pp. 2061–2070, 1962.

[6] A. Nomura, M. Ichikawa, and H. Miike, ," in *Proc. Second Iberian Conf. Patt. Recog. Image Anal. (LNCS 3522)*, 2005, vol. Part I, pp. 137–144.