Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Neural Networks 17 (2004) 159–163
Neural Networks
www.elsevier.com/locate/neunet

Neural Networks Letter

A computational model of spatio-temporal dynamics in depth filling-in

Shigeaki Nishina*, Mitsuo Kawato

ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan

Received 10 June 2003; revised 17 September 2003; accepted 17 September 2003

Abstract

We present a computational model based on the heat conduction equation, which can well explain human performance of depth interpolation. The model assumes that the depth information is locally represented and spatial integration is made by iterative processing of mutual interaction of neighbors. It reconstructs a dynamically transforming surface which is in good agreement with the results of psychophysical experiments on depth perception of untextured (uniform-colored) surface moving in depth. The model can also explain a temporal-frequency property of human percept. We conclude that the local ambiguity, which is quite common in everyday visual scenes, is solved by an interpolation mechanism based on iterative local interaction of locally represented visual information.

Keywords: Binocular vision; Horizontal disparity; Depth filling-in; Lateral connection; Diffusion

1. Introduction

Retinal images are measured locally via the receptive fields of cells in the visual cortex. Some information is lost in this process, and it causes an ill-posed problem of local measurement. While we usually perceive a stable and consistent world, each fragment of an image observed by these cells is not always *correct* as a part of the percept of the whole object. The visual system must have a mechanism that integrates fragments of images into a consistent percept. Computational research has shown that many important visual tasks can be formulated as optimization problems which can be solved by relaxation methods (Geman & Geman, 1984; Grossberg & Grunewald, 2002; Grossberg & Mingolla, 1985; Hildreth, 1984; Kawato, Hayakawa, & Inui, 1993; Mumford, 1991, 1992; Poggio, Torre, & Koch, 1985; Ullman, 1979; von der Malsburg & Schneider, 1986). Although these theories assume that the percept of an area is gradually formed by iterative calculation, as the information obtained at its borders propagates over the area, it is still important to clarify whether biological systems utilize such iterative processing.

A problem caused by localized measurement also exists in depth perception based on horizontal disparities. As a typical case, the depth of a point on an untextured horizontal line is ambiguous because the correspondence between left and right images cannot be uniquely determined (Kham & Blake, 2000; Nakayama & Shimojo, 1990). It is known that many binocular neurons in the primary visual cortex show disparity selectivity (Poggio, 1995; Poggio, Gonzales, & Krause, 1988; Prince, Cumming, & Parker, 2002). When a horizontal line without any discontinuities is observed by one of such neurons, no depth will be detected by the neuron itself because the signals from the left and right eyes are completely the same. However, the whole line is usually perceived as if it is at the same depth as the endpoints of the line. It seems that the ambiguous depths of the points on the line are determined using the endpoints' disparities.

We have examined the mechanism of how this ambiguity is solved with psychophysical methods (Nishina, Okada, & Kawato, 2003). To investigate dynamical properties of the percept, we adopted a horizontal bar moving in depth as a visual stimulus and examined spatio-temporal properties of this integration process. By gradually changing the disparity at the endpoints, the whole horizontal bar is perceived as moving in depth. Although, the depth of the central part of the bar is physically ambiguous, a certain depth is perceived as a result of depth interpolation. By using a phase-matching task, we carefully measured the perceived depth of the ambiguous region while the disparity at the endpoints was continuously changing according to a sinusoidal function at frequency f. As a result, we found that the perceived depth

* Corresponding author. Tel.: +81-774-95-1091; fax: +81-774-95-1236.
E-mail address: nishina@atr.co.jp (S. Nishina).

0893-6080/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2003.09.004

of the ambiguous region slightly differed from the depth of the endpoints. In the experiments, we manipulated length of the bar, vertical position of the bar and temporal-frequency of disparity change at the endpoints as experimental variables. Then, we found that the observed temporal discrepancy of the percept was not a phasic delay but a temporal delay, which roughly depended on the cortical length of the stimulus, that is, cortical distance between the center and the endpoints of the bar. These results indicate that the depth ambiguity is solved by propagating depth information represented in the visual cortex. In the present study, we investigate whether the experimental results are reproduced with a computational theory, which describes our hypothesis. In Section 2, we present a computational model based on the heat conduction equation which solves the local ambiguity with time-consuming calculation. And then we show the psychophysical results on depth perception and verify that the experimental data are well explained by the proposed model in the following sections.

2. Computational model

The depth ambiguity of intermediate part of an untextured horizontal bar is solved by using the depth information available at the endpoints of the bar. This phenomenon, 'depth filling-in', can be considered as interpolation of locally represented depth information. Consider a heat conduction equation as a model of depth filling-in. The depth D behaves as temperature in the equation. Suppose the time parameter τ (s) and the space parameter λ (m) are constant and do not depend on the frequency of the oscillation of the stimulus

$$\tau \frac{\partial D}{\partial t} = \lambda^2 \frac{\partial^2 D}{\partial x^2}. \tag{1}$$

According to this equation, the depth diffuses over the surface of the bar. After a certain amount of time, the model reconstructs a surface, which interpolates two endpoints. To show a typical behavior of this model, consider a case of static stimulus, a horizontal bar with two endpoints at different depth. If the bar has a uniform color and no texture, the depth is ambiguous except its endpoints because of lack of left–right correspondence. A human observer, however, usually perceives a horizontal bar slant in depth, although there's no disparity information in the intermediate region. The model can exactly reproduce this phenomenon. When our model is applied to this stimulus, the depth of the intermediate area, which is initially ambiguous, gradually varies due to the depth propagation from the endpoints. Fig. 1 shows the temporal transition of the state calculated by the model. Depth of the ambiguous region is gradually interpolated and finally a slant surface is formed.

Next, to clarify the temporal characteristics of the model, let us consider the sinusoidally varying depth of the endpoints. For time periodic oscillations, we rewrite the last expression using the following transformation of parameters from the time and space constants to the velocity and angular velocity

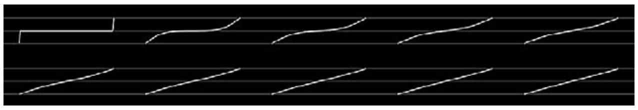

Fig. 1. Simulated time series of the interpolation of two points at different depths. Each figure shows a top view at a certain point of time during the progress of the integration. The time goes from top-left to top-right and from bottom-left to bottom-right. When two endpoints of a bar are statically presented at different depths, a slanted surface is gradually formed.

$$\frac{\lambda^2}{\tau} = \frac{v^2}{2\omega}, \tag{2}$$

where v (m/s) is a parameter corresponding to the propagation velocity, $\omega = 2\pi f$ (s^{-1}) is the angular frequency at the endpoints and f (s^{-1}) is the frequency of oscillation at endpoints. Now we can calculate the velocity

$$v = \sqrt{\frac{2\lambda^2 \omega}{\tau}}. \tag{3}$$

Using the velocity and the angular frequency, we can rewrite the heat conduction equation as

$$\frac{\partial D}{\partial t} = \frac{v^2}{2\omega} \frac{\partial^2 D}{\partial x^2}. \tag{4}$$

Under the boundary condition of sinusoidally changing depth ($\sin(\omega t)$), the analytical solution can be derived as

$$D(x,t) = c \exp\left(\frac{\omega}{v} x\right) \sin\left\{\omega t - \frac{\omega}{v} x + \theta\right\}. \tag{5}$$

It is confirmed that v is actually the propagation velocity in the above equation. We can then calculate the depth at the center of the horizontal bar ($x = 0$) and at endpoints ($x = L$)

$$\begin{aligned} D(0,t) &= 2c \sin(\omega t + \theta) \\ D(L,t) &= D(-L,t) = c \exp\left(\frac{\omega}{v} L\right) \sin\left(\omega t - \frac{\omega}{v} L + \theta\right) \\ &\quad + c \exp\left(-\frac{\omega}{v} L\right) \sin\left(\omega t + \frac{\omega}{v} L + \theta\right) \\ &= Ac \sin(\omega t + \theta - \Phi), \end{aligned} \tag{6}$$

where

$$A = \sqrt{\exp\left(2\frac{\omega}{v} L\right) + \exp\left(-2\frac{\omega}{v} L\right) + 2\left\{1 - 2\sin^2\left(\frac{\omega}{v} L\right)\right\}} \tag{7}$$

$$\Phi = \arctan\left[\frac{\left\{\exp\left(\frac{\omega}{v} L\right) - \exp\left(-\frac{\omega}{v} L\right)\right\} \sin\left(\frac{\omega}{v} L\right)}{\left\{\exp\left(\frac{\omega}{v} L\right) + \exp\left(-\frac{\omega}{v} L\right)\right\} \cos\left(\frac{\omega}{v} L\right)}\right]. \tag{8}$$

Suppose that $(\omega/v)L$ is much smaller than one, then the phase delay is well approximated as a concave quadratic function

$$A \sim 2 + O(L^2) \tag{9}$$

$$\Phi \sim \left(\frac{\omega}{v}L\right)^2. \tag{10}$$

It should also be noted that the amplitude of oscillation at the center is smaller than that of the edges, but the difference is very small (small order of the square of the bar length).

3. Psychophysical experiment

We have shown spatio-temporal properties of depth filling-in in the previous psychophysical study, and argued that the depth filling-in is a dynamic integration process, that is, propagation of depth information in the visual cortex. Detail of the experiments are described in Nishina et al. (2003). Here we introduce a part of the experiments relevant to the present study. In the experiment, we measured processing time of depth interpolation in the following way. A uniform-colored horizontal bar was presented as a stimulus on a dark background. By using LCD stereo shutter glasses, we continuously changed the disparities at the two endpoints of the bar simultaneously according to the equation

$$d = A\sin(2\pi ft), \tag{11}$$

with $A = 0.1°$ (Fig. 2). Although only the disparity of the endpoints was altered and the other part of the horizontal bar was completely fixed, the observers perceived depth motion of the whole horizontal bar. Because the bar was uniform-colored and did not have any texture on it, perceived depth change at the center of the bar is considered as a result of depth filling-in. A vertical bar was presented as a depth probe at the center of the horizontal bar (Fig. 2). The probe also oscillated in depth with the same frequency and amplitude as the horizontal bar according to the equation

$$d_p = A\sin(2\pi ft + \phi). \tag{12}$$

The initial phase difference at the beginning of a trial was randomly set between the range of $-\pi$ and π. The subjects could change ϕ, the phase difference between the endpoints of the horizontal bar and the vertical probe, by pressing keys (Fig. 3). Subjects adjusted the phase of the probe oscillation to match it to the center of the horizontal bar in perceived depth. With this phase-matching task, we can measure the perceived depth of the center of an untextured horizontal bar that moves sinusoidally in depth. A fixation point was presented near the intersection point of two bars. The fixation point was always presented with zero disparity. A chin rest was used to maintain the head position during each session. There were three experimental parameters, i.e. length of the horizontal bar (8, 10, 12, 14 or 16°), vertical position of the horizontal bar (0.5 or 3.0°) and the frequency of the oscillation (1.0 or 1.5 Hz). Five subjects participated in the experiment and each subject performed 300 trials for each vertical position and frequency condition.

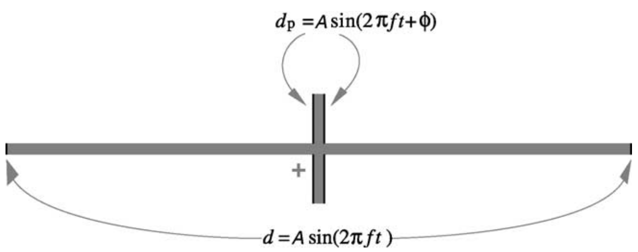

Fig. 2. The visual stimulus used in the psychophysical experiments consisted of a horizontal bar and a vertical bar. The vertical bar was used as a probe. Only the vertical edges shown as dark lines in this figure have disparity. The disparities of the edges were continuously updated according to the equations indicated with the arrows in this figure. Note that the arrows, equation, and the darker vertical edges are shown here for explanation, and were not presented during the experiment. The subjects adjusted the oscillation phase ϕ of the vertical bar so as to make it moving perceptually together with the center of the horizontal bar.

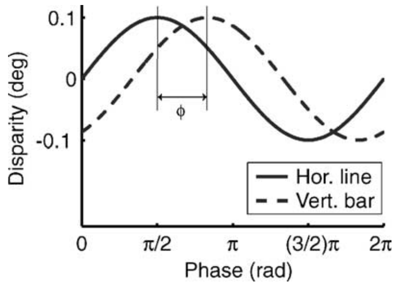

Fig. 3. Horizontal disparities of both the endpoints of the horizontal line and the vertical bar were modulated sinusoidally in the same amplitude and frequency. Only the phase between them (ϕ) were varied by the subjects.

4. Experimental results and model prediction

The results showed that in all cases, the perceived depth at the center was delayed relative to that at the endpoints (Fig. 4). The analysis of variance reveals that the delay at the center significantly increased as the bar became longer ($F_{4,8} = 5.28$, $p < 0.001$). As for the vertical position, the delay was significantly shorter when the stimulus was presented farther from the fixation point ($F_{1,4} = 15.4$, $p < 0.002$). The cortical sizes of objects of a physically equal size differ when presented at different retinal locations. The density of the receptive fields is the highest at the fovea and becomes lower toward the periphery. Accordingly, a stimulus with a fixed size covers fewer neurons in the cortex when presented peripherally than when presented foveally. If the depth information were propagated via local mutual interactions of neurons, the time for the endpoints' depth information to reach the center would depend on the number of neurons. Furthermore, the number of neurons is expected to be smaller for the periphery. The results are

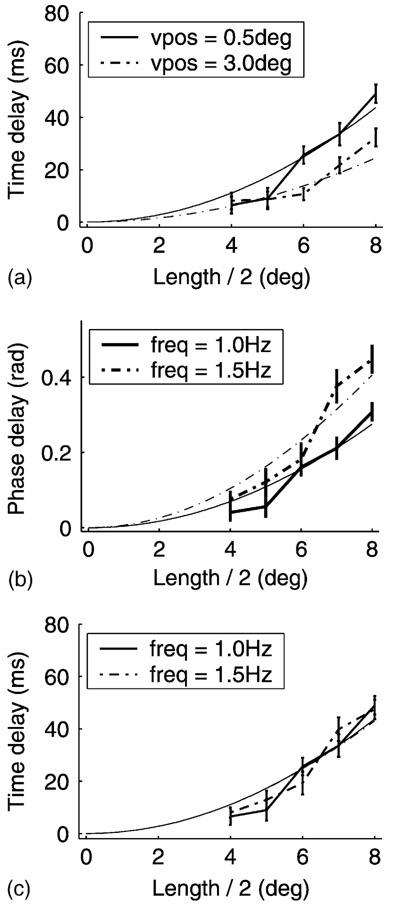

Fig. 4. (a) Effect of the line length from the results of the psychophysical experiment is shown for each vertical position (vpos). (b, c) Effect of frequency. The same data are plotted as phase delay (b) and time delay (c). The error bars show the standard errors of the means (SEM). SEM is an index of the uncertainty in the average of the measurement and calculated as SEM $= \sqrt{\sigma/n}$. The thin lines show quadratic functions fitted to the data (see text).

qualitatively consistent with this interpretation. The effect of frequency was also important. We compared the results under two oscillating frequencies conditions. The effect of oscillation frequency was significant for the phase difference ($F_{1,2} = 108.8$, $p < 0.01$) but insignificant for the time difference ($F_{1,2} = 1.34$, $p = 0.367$). The results showed that the temporal difference was almost constant when it was considered to be a time delay. This indicates that the phase difference can be treated as a time delay.

The curved lines plotted in the graphs are predictions made by the computational model of the depth propagation proposed in Section 2. As for the length values between four and eight, the experimental data were very well predicted by the model. The fitting parameters were determined by only using the results of 1.0 Hz frequency condition. That is, the prediction curve plotted for 1.5 Hz frequency was calculated using the parameters obtained by fitting the data of 1.0 Hz condition. The frequency property of the model is very consistent with the experimental data. The model prediction is invariant for frequency when the delay is considered temporal, and not invariant when considered phasic. This behavior is also in agreement with the experimental data. The propagation speed predicted by fitting the model to the experimental data was $v = 95.3$ deg/s when the vertical position was 0.5° and $v = 116.7$ deg/s when the vertical position was 3.0°. Bringuier, Chavane, Glaeser and Frégnac (1999) measured propagation speed of visual signals in cat area 17. The typical propagation speed reported in their study is 0.1 m/s, which is approximately 100 deg/s under an average cortical magnification factor of 1 mm in the cortex for 1° in visual angle (conversion ratio used in Bringuier et al. (1999)). Paradiso and Nakayama (1991) and Rossi and Paradiso (1996) reported estimated propagation speeds in terms of brightness filling-in with psychophysical experiments. They performed brightness masking experiment (Paradiso and Nakayama) or temporal brightness induction experiment (Rossi and Paradiso) and estimated 110–150 or 140–180 deg/s. Our estimation is approximately comparable to theirs.

5. Conclusion

In this study, we showed the depth interpolation is well explained by a computational model based on the heat conduction equation, by combining the model and a psychophysical experiment which had been designed carefully to verify the model. The temporal property of human depth interpolation measured with a psychophysical method was highly compatible with the prediction of the diffusion model based on local connection of locally represented depth information and iterative processing. These results strongly supports a propagation mechanism based on a kind of neural spreading. Local representations, local interactions and iterative calculations appear to form a fundamental mechanism of visual information processing in the brain.

Acknowledgements

This research was conducted as part of ‘Research on Human Communication’ with funding from the Telecommunications Advancement Organization of Japan.

References

Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, *283*(5402), 695–699.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741.

Grossberg, S., & Grunewald, A. (2002). Temporal dynamics of binocular disparity processing with corticogeniculate interactions. *Neural Networks*, *15*(2), 181–200.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological Review*, *92*(2), 173–211.

Hildreth, E. C. (1984). *The measurement of visual motion*. Cambridge: MIT Press.

Kawato, M., Hayakawa, H., & Inui, T. (1993). A forward-inverse optics model of reciprocal connections between visual areas. *Network: Computation in Neural Systems*, *4*, 415–422.

Kham, K., & Blake, R. (2000). Depth capture by kinetic depth and by stereopsis. *Perception*, *29*(2), 211–220.

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, *65*(2), 135–145.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251.

Nakayama, K., & Shimojo, S. (1990). Toward a neural understanding of surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, *LV*, 911–924.

Nishina, S., Okada, M., & Kawato, M. (2003). Spatio-temporal dynamics of depth propagation on uniform region. *Vision Research*, *43*(24), 2493–2503.

Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, *31*(7/8), 1221–1236.

Poggio, G. F. (1995). Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex*, *5*(3), 193–204.

Poggio, G. F., Gonzales, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *Journal of Neuroscience*, *8*(12), 4531–4550.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*(6035), 314–319.

Prince, S. J., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology*, *87*(1), 209–221.

Rossi, A. F., & Paradiso, M. A. (1996). Temporal limits of brightness induction and mechanisms of brightness perception. *Vision Research*, *36*(10), 1391–1398.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.

von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, *54*(1), 29–40.