Neural Networks 99 (2018) 42–55

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Computational study of depth completion consistent with human bi-stable perception for ambiguous figures

Eiichi Mitsukura *, Shunji Satoh

Graduate School of Information Systems, The University of Electro-communications, Tokyo, 182-8585, Japan

ARTICLE INFO

Article history:
Received 5 October 2016
Received in revised form 7 September 2017
Accepted 21 November 2017
Available online 12 December 2017

Keywords:
Vision
Depth perception
Stereopsis
Computational model
Depth filling-in
Curvature

ABSTRACT

We propose a computational model that is consistent with human perception of depth in "ambiguous regions," in which no binocular disparity exists. Results obtained from our model reveal a new characteristic of depth perception. Random dot stereograms (RDS) are often used as examples because RDS provides sufficient disparity for depth calculation. A simple question confronts us: "How can we estimate the depth of a no-texture image region, such as one on white paper?" In such ambiguous regions, mathematical solutions related to binocular disparities are not unique or indefinite. We examine a mathematical description of depth completion that is consistent with human perception of depth for ambiguous regions. Using computer simulation, we demonstrate that resultant depth-maps qualitatively reproduce human depth perception of two kinds. The resultant depth maps produced using our model depend on the initial depth in the ambiguous region. Considering this dependence from psychological viewpoints, we conjecture that humans perceive completed surfaces that are affected by prior-stimuli corresponding to the initial condition of depth. We conducted psychological experiments to verify the model prediction. An ambiguous stimulus was presented after a prior stimulus removed ambiguity. The inter-stimulus interval (ISI) was inserted between the prior stimulus and post-stimulus. Results show that correlation of perception between the prior stimulus and post-stimulus depends on the ISI duration. Correlation is positive, negative, and nearly zero in the respective cases of short (0–200 ms), medium (200–400 ms), and long ISI (>400 ms). Furthermore, based on our model, we propose a computational model that can explain the dependence.

1. Introduction

When binocular images include no visual disparity information, as shown in uniformly colored images, how does our visual system estimate the depth or surface structure of objects? Answering this question defines the main theme of this research.

Using binocular visual information, the human visual system estimates the surface structure of objects (e.g. concave, convex, flat) in addition to those objects' positional relation. Horizontal disparity embedded within right and left retinal images provides fundamental clues that support estimation of depth differences between objects. Therefore, an important task of visual systems is to calculate the horizontal disparities (signed disparities) of matching points at every location of two retinal images. Synthetic random-dot stereograms are widely used as input stimuli for stereo vision in such experiments and theoretical studies. For a synthetic random-dot stereogram, the degree of spatial disparity is defined uniquely at every spatial location of images. However, as non-textured images and periodic textures show, general images include many spatial areas for which no unique solution of disparity can ever be determined. The white paper you might now be viewing is one example of non-unique disparity. In this case, the number of the solutions of depth is infinite because the number of possible matching points is also infinite, although our visual system must determine an appropriate solution from an infinite number of solutions of depth.

Such regions are designated herein as "ambiguous regions". Fig. 1a depicts examples of images with an ambiguous region. As Fig. 1a shows, in a solid-figure stereogram, along the left and right line segments of the rectangle or trapezoid, unique solutions of horizontal disparities are determined by finding matching points (closed line in Fig. 1b). For example, the matching point of the lower left acute angle in the left image is the obtuse angle at the left-lower point of the right trapezoid. Nevertheless, no unique disparity solution exists in the black ambiguous region at any point. Periodic textured images and the half-occlusion area should also be categorized in ambiguous regions because these areas do not provide a unique solution of disparity. The analyses described in

* Correspondence to: Building W-10, Room 435, 1-5-1 Chofugaoka, Chofu, Tokyo, 182-8585, Japan.
E-mail addresses: eiichi@hi.is.uec.ac.jp (E. Mitsukura), shun@is.uec.ac.jp (S. Satoh).

https://doi.org/10.1016/j.neunet.2017.11.016
0893-6080/

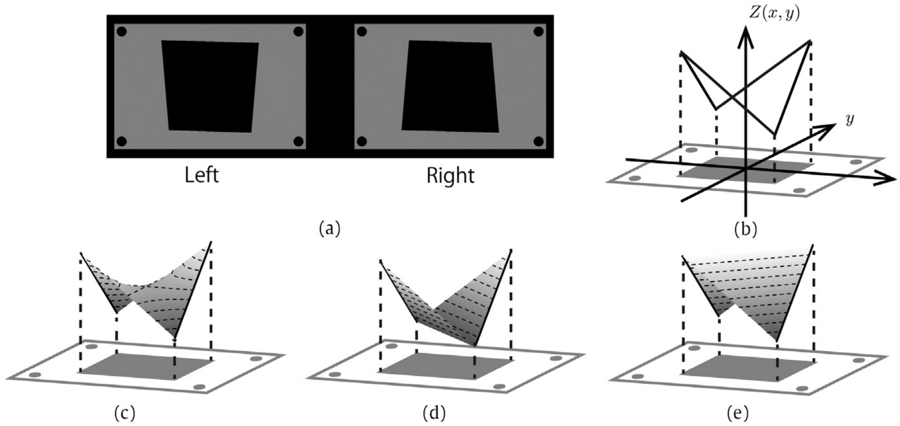

Fig. 1. a. Stereogram used for the psychological experiment explained in Section 3. Two figures show parallel view methods. b. Slanted lines of the depth value $Z(x, y)$ calculated using binocular disparity in a. $Z(x, y)$ is the depth at point (x, y). c. Example of depth propagation using a heat conduction equation: hyperbolic paraboloid (saddle; curvature of iso-depth line $\overline{\kappa} \neq 0$ and curvature of flow line $\overline{\mu} \neq 0$. Details are presented in Section 2). d and e. Human percepts: flat depth maps. All depth contours are straight ($\overline{\kappa} = 0$) and parallel ($\overline{\mu} = 0$) lines.

this report specifically examine the depth completion of regions in which no disparity information is available because of uniform luminance, and not other types of ambiguity. We do not address the half-occlusion problem. Problems related to periodic texture (periodic matching point) are also beyond the scope of this article.

Completion from the disparity or depth that is determined for the non-ambiguous region (Fig. 1b) is one means of having a unique value of disparity in the ambiguous region. Computationally, "smoothness" has been used by many models as a criterion to complete depth constrained by the determined disparity (Belhumeur, 1996; Marr & Poggio, 1976; Pollard, Mayhew, & Frisby, 1985). From a psychological perspective, Würger and Landy found that humans complete depth in ambiguous regions (Würger & Landy, 1989). Georgeson et al. investigated the fundamental algorithm of human depth completion for ambiguous regions (Georgeson, Yates, & Schofield, 2009). Based on their results, they reported that humans can implement depth completion by depth propagation from the determined region of depth into ambiguous regions.

Some visual models use the depth propagation scheme. Fig. 1c presents one example of a depth solution determined using a propagation scheme of isotropic diffusion. The isotropic diffusion constrained with a boundary condition (depth determined from Fig. 1b) generates as "smooth" a surface as possible. This "smoothness" criterion (energy function) defined by the first-order spatial derivative of the depth surface has been used for many computational models of stereopsis. For example, Nishina and Kawato (2004) propose a depth-completion model based on the heat conduction equation, which is isotropic-diffusion. In the resultant depth by isotropic-diffusion, the completed depth obligates a saddle shape. Mathematically, the saddle takes zeros of the mean curvature.

Human perception differs from the "saddle" surface shown in Fig. 1c, but humans tend to recognize a "flat" surface, as depicted in Fig. 1d and 1e (Ishikawa, 2007; Ishikawa & Geiger, 2006). Similar results were found in the case presented in Fig. 2a. From observing Figs. 1d and 2d (Figs. 1e and 2e), Ishikawa and Geiger (2006) reported that perceived depth has a common mathematical property: the Gaussian curvature is zero. No neural network model has yet reproduced human perception according to Fig. 1d and 1e and Fig. 2d and 2e. The present study specifically examines the development of a neural network that completes depth in the ambiguous region by spatial propagation so that the Gaussian curvature is zero.

This article is organized as follows. Section 2 presents our proposed model for depth completion, along with results obtained using numerical simulation with the proposed model. Section 3, with a psychological experiment, presents a new visual characteristic obtained by predictions from our model. Section 4 presents a model that is consistent with the experimentally obtained results presented in Section 3 from computational viewpoints. Section 5 includes discussion of our model from computational and physiological viewpoints. Section 6 explains our conclusions.

2. Depth completion by propagation scheme

2.1. Minimizing the first order derivatives of surfaces

The following energy function $E_{\text{smooth}}[Z]$ presents a simple evaluation of the surface "smoothness" as quantified using the first order differentiation of a depth function $Z(x, y)$.

$$E_{\text{smooth}}[Z] = \frac{1}{2} \iint_B \|\nabla Z(x, y)\|^2 \, dxdy. \tag{1}$$

In that equation, B represents an ambiguous region of depth. Applying the steepest descent method (or the Euler–Lagrange equation) to Eq. (1) to obtain an iterative update rule for $Z(x, y)$ that minimizes E_{smooth}, one obtains the diffusion equation shown below.

$$\frac{\partial}{\partial t} Z(x, y, t) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) Z(x, y, t) = \Delta Z(x, y, t). \tag{2}$$

Therein, t represents the step time during the diffusion process starting with the initial condition of $Z(x, y, 0)$. A converging $Z(x, y, t)$ is the final result of depth completion by the diffusion process. The resultant surfaces by Eq. (2) (converged Z; $\partial Z/\partial t = 0$) are saddles which are not the expected ones presented in Figs. 1c and 2c.

2.2. Minimizing Gaussian curvature

We discuss naïve deviation of the dynamics to obtain a "flat" surface by minimizing the Gaussian curvature. The following evaluation function $E_{\text{flat}}[Z]$ using Gaussian curvature K would be more suitable to obtain "flat" surfaces ($K(x, y) = 0$ at each point of B) reflecting human perception for the ambiguous region.

$$E_K[Z] = \frac{1}{2} \iint_B K(x, y)^2 dxdy. \tag{3}$$

For that equation, the following definition is used.

$$K(x, y) = \frac{Z_{xx}Z_{yy} - Z_{xy}^2}{\left(1 + Z_x^2 + Z_y^2\right)^2}. \tag{4}$$