Available online at www.sciencedirect.com

ScienceDirect

Neural Networks 21 (2008) 331–340

Neural Networks

www.elsevier.com/locate/neunet

2008 Special Issue

Binocular robot vision emulating disparity computation in the primary visual cortex☆

Kazuhiro Shimonomura[a], Takayuki Kushima[b], Tetsuya Yagi[b,*]

[a] The Center for Advanced Medical Engineering and Informatics, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
[b] Department of Electrical, Electronic and Information Engineering, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

Received 3 August 2007; received in revised form 1 December 2007; accepted 11 December 2007

Abstract

We designed a VLSI binocular vision system that emulates the disparity computation in the primary visual cortex (V1). The system consists of two silicon retinas, orientation chips, and field programmable gate array (FPGA), mimicking a hierarchical architecture of visual information processing in the disparity energy model. The silicon retinas emulate a Laplacian–Gaussian-like receptive field of the vertebrate retina. The orientation chips generate an orientation-selective receptive field by aggregating multiple pixels of the silicon retina, mimicking the Hubel–Wiesel-type feed-forward model in order to emulate a Gabor-like receptive field of simple cells. The FPGA receives outputs from the orientation chips corresponding to the left and right eyes and calculates the responses of the complex cells based on the disparity energy model. The system can provide the responses of complex cells tuned to five different disparities and a disparity map obtained by comparing these energy outputs. Owing to the combination of spatial filtering by analog parallel circuits and pixel-wise computation by hard-wired digital circuits, the present system can execute the disparity computation in real time using compact hardware.

Keywords: Neuromorphic engineering; Analog VLSI; Multi-chip; FPGA; Disparity energy model

1. Introduction

The brain computes images using quite different algorithms and architectures from those used in the conventional digital image processing systems. By using this unique algorithm and architecture, the visual system of the brain can perceive external scenes in real time with extremely low power consumption (Mead, 1990). In mammals, the visual information is transmitted mainly to the primary visual cortex (V1) in parallel via the lateral geniculate nucleus (LGN) after the information is pre-processed by the retina. Neurons in these different stages of processing are tuned to some specific cues of the input image. Bipolar cells in the retina respond maximally to a small spot of light (Kaneko, 1970). Yet simple cells in V1 respond maximally to a slit of light with a specific orientation (Hubel & Wiesel, 1962). These functional and structural characteristics, i.e., the hierarchical arrangement of neuronal networks and visual cue-tuned receptive fields, are considered to be highly relevant to the extremely efficient computations in the brain. Therefore, from an engineering perspective, it is interesting to design an image processing system that replicates the fundamental computations of V1 circuits that have been revealed in earlier physiological studies.

Owing to their computational advantages, e.g., low power dissipation and compact hardware, analog very large scale integrated (aVLSI) vision chips, namely neuromorphic vision chips, are considered to be a particularly good candidate for replicating the fundamental architecture of neuronal circuits (Indiveri & Douglas, 2000). Recently, neuromorphic vision chips have been fabricated to emulate the responses of V1 simple cells (Choi, Merolla, Arthur, Boahen, & Shi, 2005; Liu et al., 2001; Venier, Mortara, Arreguit, & Vittoz, 1997). These chips effectively replicate the receptive field of simple cells, which requires a high cost of computation to be implemented

☆ An abbreviated version of some portions of this article appeared in Yagi and Shimonomura (2007) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

* Corresponding author. Tel.: +81 6 6879 7785; fax: +81 6 6879 7784.
E-mail address: yagi@eei.eng.osaka-u.ac.jp (T. Yagi).

0893-6080/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.12.033

with conventional digital machines, in real time with low power dissipation. Such aVLSI chips, however, are not as versatile as the conventional machine vision systems, since these chips compute images with the physical properties of electronic circuits built in the chip and it is difficult to implement highly programmable architecture. Owing to this restriction, it is not feasible to replicate responses and functions of higher-order neurons in V1 circuits, particularly those having non-linear response properties, only using the aVLSI technology.

Tsang and Shi have recently reported a mixed analog–digital neuromorphic vision system to compute the responses of single binocular complex cells based on the disparity energy model (Ohzawa, DeAngelis, & Freeman, 1990) using Gabor-like chips and a microprocessor (Tsang & Shi, 2004). In the previous study, we have fabricated an aVLSI multi-chip system that can mimic the neural image of the V1 simple cell layer (Shimonomura & Yagi, 2005). This system consists of a silicon retina that computes the Laplacian–Gaussian-like spatial filtering, and the orientation chips that aggregate the outputs of the silicon retina. Here, we extended this approach to design a mixed analog–digital neuromorphic vision system consisting of silicon simple cells and field programmable gate array (FPGA) circuits (Yagi & Shimonomura, 2007). The present system calculates the response of the binocular complex cell over all the pixels of the input image based on the disparity energy model in real time and thus emulates the neural image of the binocular complex cell network in V1.

2. Disparity energy model

Binocular disparity is the difference between the positions of the retinal images of an object projected on the left and right eyes, and it is known as one of the important cues used for depth perception. The neurons that are tuned to specific disparities have been found in V1. A disparity energy model explains the response properties of the binocular complex cells in V1 (Ohzawa et al., 1990). This model predicts well the shape of the binocular receptive fields of the complex cell in cats (Ohzawa, DeAngelis, & Freeman, 1997). The validity of the disparity energy model has been confirmed in monkeys (Cumming & Parker, 1997). Fig. 1 shows the schematic of the disparity energy model. It consists of two stages, namely, the simple cell stage and the complex cell stage. Firstly, the input images to the left and right eyes are processed at the simple cell stage. Each simple cell in the left and right eyes has spatial receptive fields that resemble Gabor filters (Jones & Palmer, 1987). The responses of these monocular simple cells can be expressed as

$$R_{\mathrm{even}}(x) = \int g_{\mathrm{even}}(w-x) I(w) \mathrm{d}w \tag{1}$$

$$R_{\mathrm{odd}}(x) = \int g_{\mathrm{odd}}(w-x) I(w) \mathrm{d}w \tag{2}$$

where

$$g_{\mathrm{even}}(w) = \mathrm{e}^{-k_1 w^2} \cos(k_2 w) \tag{3}$$

$$g_{\mathrm{odd}}(w) = \mathrm{e}^{-k_1 w^2} \sin(k_2 w). \tag{4}$$

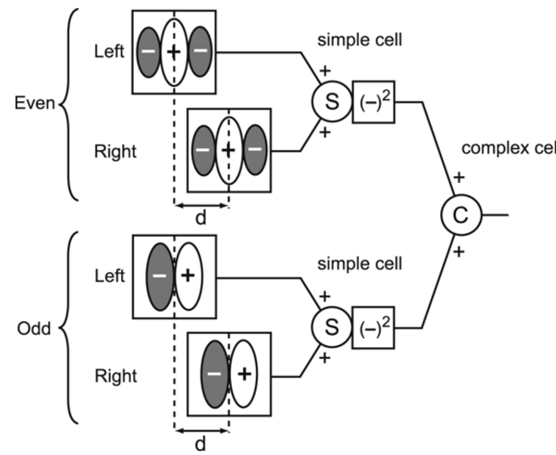

Fig. 1. Disparity energy model. Individual complex cell (C) consists of two binocular simple cells (S), which have different symmetries of spatial receptive field structures. The original model has four types of spatial filters and four half-wave rectified simple cells. However, it is equivalent to the model with two simple cells shown here. The difference in the retinal location between the left and right receptive fields of the simple cell, d, defines the preferred disparity of the resulting complex cell.

Here, I denotes the input image and x and w denote the position of the simple cell and the pixel of the input image, respectively. In this model, two types of Gabor filters, even-type g_{even} and odd-type g_{odd}, are used as the model of the receptive fields of the simple cells. Their spatial phases differ by 90°. These receptive fields of the simple cells are oriented along a specific direction. All the simple cells in the model used here have a vertically elongated receptive field to extract vertically oriented visual patterns that are suitable for computing a horizontal disparity. In this study, the responses of the monocular simple cells with the same type of receptive field in the left and right are summed up and squared. In the original model proposed by Ohzawa et al., each simple cell should be split into two with inverted receptive field profiles such that they can carry the positive and negative parts of the firing, respectively because the cortical cells cannot fire negatively. The resulting complex cell model will contain four half-wave rectified simple cells. It is, however, equivalent to the model shown here with two simple cells (Qian, 1997).

The response of the complex cell, C in the figure, is obtained by squaring and adding the two types of the responses of the binocular simple cell S. The response of the complex cell with disparity d can be expressed as

$$R_{cx}^{d}(x_L, x_R) = \left\{ R_{\mathrm{even}}^{L}\left(x_L + \frac{\mathrm{d}}{2}\right) + R_{\mathrm{even}}^{R}\left(x_R - \frac{\mathrm{d}}{2}\right) \right\}^2 + \left\{ R_{\mathrm{odd}}^{L}\left(x_L + \frac{\mathrm{d}}{2}\right) + R_{\mathrm{odd}}^{R}\left(x_R - \frac{\mathrm{d}}{2}\right) \right\}^2. \tag{5}$$

Here, x_L and x_R are the position of the stimuli in the left and right retinas, respectively. d is the difference in the center position of the receptive field between the left and right monocular simple cells, and it is referred to as the preferred disparity in which the model exhibits the maximum response.