2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Neuromorphic binocular vision system for real-time disparity estimation

Kazuhiro Shimonomura, Takayuki Kushima and Tetsuya Yagi

Abstract— **We describe a binocular vision system that emulates disparity computation in the neuronal circuit of the primary visual cortex (V1). The system consists of two sets of silicon retinas and simple cell chips that correspond to the binocular vision and field programmable gate array (FPGA) circuit. This arrangement mimics the hierarchical architecture of the visual system of the brain. The silicon retina is an analog very large scale integrated (aVLSI) circuit and possesses a Laplacian-Gaussian-like spatial filter similar to the receptive field of the vertebrate retina. The simple cell chip generates a Gabor-like spatial filter similar to the orientation-selective receptive field of the simple cell in V1 by aggregating several pixels of the silicon retina. The FPGA receives the outputs from the two simple cell chips corresponding to binocular inputs from the left and right eyes and calculates tha binocular disparity in real-time based on the disparity energy model. The system provides output images tuned to five different disparities in parallel. The disparity map is obtained by comparing these disparity energy outputs. Due to the combination of the parallel and analog computation of the aVLSIs and the pixel-wise computation with hard-wired digital circuits, the present system can efficiently compute the binocular disparity using compact hardware and low power dissipation in real-time.**

I. INTRODUCTION

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 one of the important cues used for depth perception. The neurons that are tuned to specific disparities have been discovered in physiological studies on the primary visual cortex (V1) of cats and monkeys [1]. It is pointed out that the biological vision system performs such computation with considerable efficiency by using more compact hardware and lower power consumption than a conventional image processing system consisting of charge coupled device (CCD) cameras and a digital computer [2]. Therefore, the architecture and algorithm used for disparity computation in the brain are significant and interesting from an engineering view point.

The purpose of the present study is to design and implement a neuromorphic binocular vision system inspired by the architecture of V1 for computing the binocular disparity in real-time.

There have been some studies on analog very large scale integrated (aVLSI) implementations of disparity computation in neuromorphic systems [3], [4]. In the present study, a binocular disparity energy model [5] based on the hierarchical architecture of V1 is emulated with a mixed analog and digital configuration in which aVLSIs mimicking the response properties of orientation-selective simple cells and field programmable gate array (FPGA) are combined.

K.Shimonomura is with The Center for Advanced Medical Engineering and Informatics, Osaka university, 2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan. kazu@eei.eng.osaka-u.ac.jp

T.Yagi is with Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan. yagi@eei.eng.osaka-u.ac.jp

T.Kushima was with Graduate School of Engineering, Osaka University. He is now with Hitachi,Ltd., Ibaragi, Japan.

II. BINOCULAR DISPARITY ENERGY MODEL

Fig.1 depicts the geometry of binocular vision. The binocular disparity is the difference between the positions of an object image projected on the left and right retinas. Here, we define the disparity d as $x_R - x_L$. The fixation point F projects onto the points in the left and right retinas and corresponds to zero disparity. The far point A and the near point B have positive disparity and negative disparity, respectively, by definition.

Ohzawa et al. proposed the disparity energy model to explain the response properties of binocular complex cells in V1, as shown in Fig.2 [5]. Simple cells in the left and right eyes have spatial receptive fields that resemble Gabor functions. Monocular simple cells of even- and odd-type receptive fields, whose spatial phases differ by 90°, converge on the subunits of the binocular simple cell. There is a difference in d with regard to the center positions of the receptive fields of the monocular simple cells in the left and right eyes; this difference defines the preferred disparity of the cell. In this model, the response of the complex cell (Cx in the figure) is expressed by squaring and adding the two types of binocular simple cell responses (S in the figure). There are two methods to encode d within the scope of the disparity energy model [6]. One method is position-shift encoding in which d is defined by the difference in the position of the receptive field, as shown in Fig.2. The other method is phase-shift encoding that defines d based on the difference between phases of the receptive fields of the left and right eyes. Here, we are concerned only with position-shift encoding.

In this model, the simple cells have a vertically elongated receptive field to extract vertically oriented visual patterns that are suitable for computing a horizontal disparity. The even- and odd-type receptive fields are integrated, and the complex cell responds to a pattern that provides the preferred disparity, independent of the phase of the input pattern.

III. SYSTEM ARCHITECTURE

A. Overview and analog multi-chip system

Fig.3(A) shows a schematic of the hierarchical architecture of the vision system developed in the present study. The system consists of three layers of electric circuits and computes the disparity energy in real-time. Each circle represents an

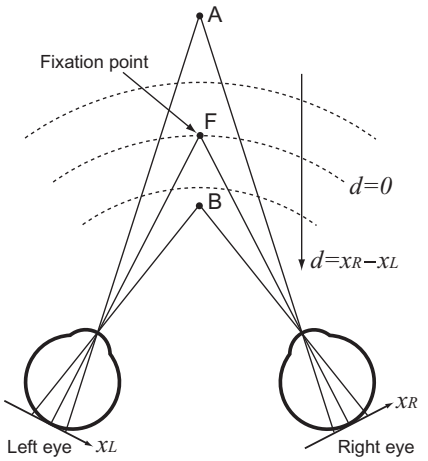

Fig. 1. Schematic of stereopsis and binocular disparity. F denotes the fixation point. Every point on the horopter has zero disparity. A and B represent far and near points, respectively. These points have positive and negative disparities, respectively, when the disparity d is defined as $x_R - x_L$.

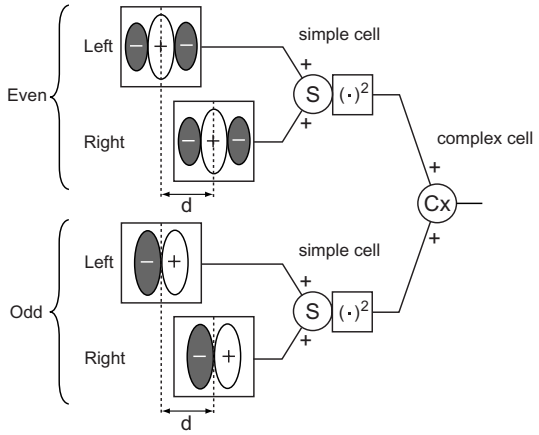

Fig. 2. Disparity energy model. The individual complex cell consists of two binocular simple cells that have different symmetries of the spatial receptive field structure. The original model has four types of spatial filters and four half-wave rectified simple cells. However, it is equivalent to the model with two simple cells shown here [6]. The difference in the retinal location of the receptive field of the simple cells, d, defines the preferred disparity of the complex cell.

individual pixel. The upper-most layers correspond to the right and left retinas and each pixel in these layers has a concentric center-surround receptive field, also referred to as a Laplacian-Gaussian-like receptive field. The second layers correspond to the monocular simple cell layers that receive inputs from the retinas via the lateral geniculate nucleus (LGN). In V1, the orientation-selective receptive field is thought to be generated by integrating the concentric center-surround receptive fields along the corresponding orientation [7]. Similarly, in the present architecture, the outputs of the pixel aligned at a specific orientation in the first layer are aggregated into a pixel circuit in the second layer. In the figure, five retinal neurons, aligned at 60°, are converged onto a simple cell. This elongated receptive field is of the even-type. The odd-type receptive fields can be obtained by taking the difference between neighboring even-type outputs tuned to the same preferred orientation, as shown in the figure. The outputs of the pixel in the simple cell layers of the left and right sides are summed in the binocular simple cell. The squared responses of the binocular simple cells containing the even- and odd-type responses of the left and right sides converge separately into a complex cell, forming the third layer.

The hierarchical architecture of the binocular vision system described above is implemented with analog and digital VLSIs. Fig.3(B) shows a block diagram of the binocular vision system. The system consists of two silicon retinas, simple cell chips, and FPGA circuits.

The silicon retina used in the present study was implemented in a previous study [8]. The architecture of this silicon retina was originally developed by Kameda and Yagi [9]. The silicon retina has 100×100 pixels, each of which consists of an active pixel sensor and an analog processing circuit. These pixels are arranged in a hexagonal array. Neighboring pixels are connected by two layers of resistive networks that model the neuronal network in the vertebrate outer retina [10]. Each pixel has a Laplacian-Gaussian-like receptive field.

The simple cell chip used in this study has a similar architecture to the one developed in a previous study [11], and the number of pixels is 100×100. The chip consists of a pixel circuit array and six shift registers. Each pixel circuit includes analog memory to store the voltage input fed from the silicon retina. The image represented by the analog voltages in the silicon retina is transferred pixel by pixel to the simple cell chip. After the image is transferred, multiple pixels aligned at a preferred orientation (90° in the present case) are aggregated to generate an orientation-selective receptive field. Fig.4 shows the output images of the simple cell chip in response to a spot of light. (A) and (B) are the even- and odd-type responses, respectively. The preferred orientation is 90° (vertical). The number of aggregated pixels is eight. The odd-type receptive fields are obtained by the off-chip subtraction of two neighboring even-type pixels placed four pixels apart and tuned to the same preferred orientation [11]. In both the even- and odd-type simple cells, vertically oriented responses are obtained with alternately arranged positive (shown as white) and negative (shown as black) subregions.

Fig.3(C) shows an overview of the present binocular vision system. The silicon retinas are installed in camera boxes placed on the left and right sides of the system. A CCTV lens (Pentax B2514D, f = 25 mm) is mounted on each camera. The distance between the two cameras is 14 cm. The three chips behind the camera boxes are simple cell chips. Two of these three chips are connected to the right and left silicon retinas. The power consumptions of the silicon retina and the simple cell chip is 36 mW and 189 mW, respectively [8], [11]. An FPGA board located on the upper right side includes analog-to-digital (AD) converters (Sony CXD1175AM), FPGA (Xilinx XCV300), SRAM, and video interface circuits.

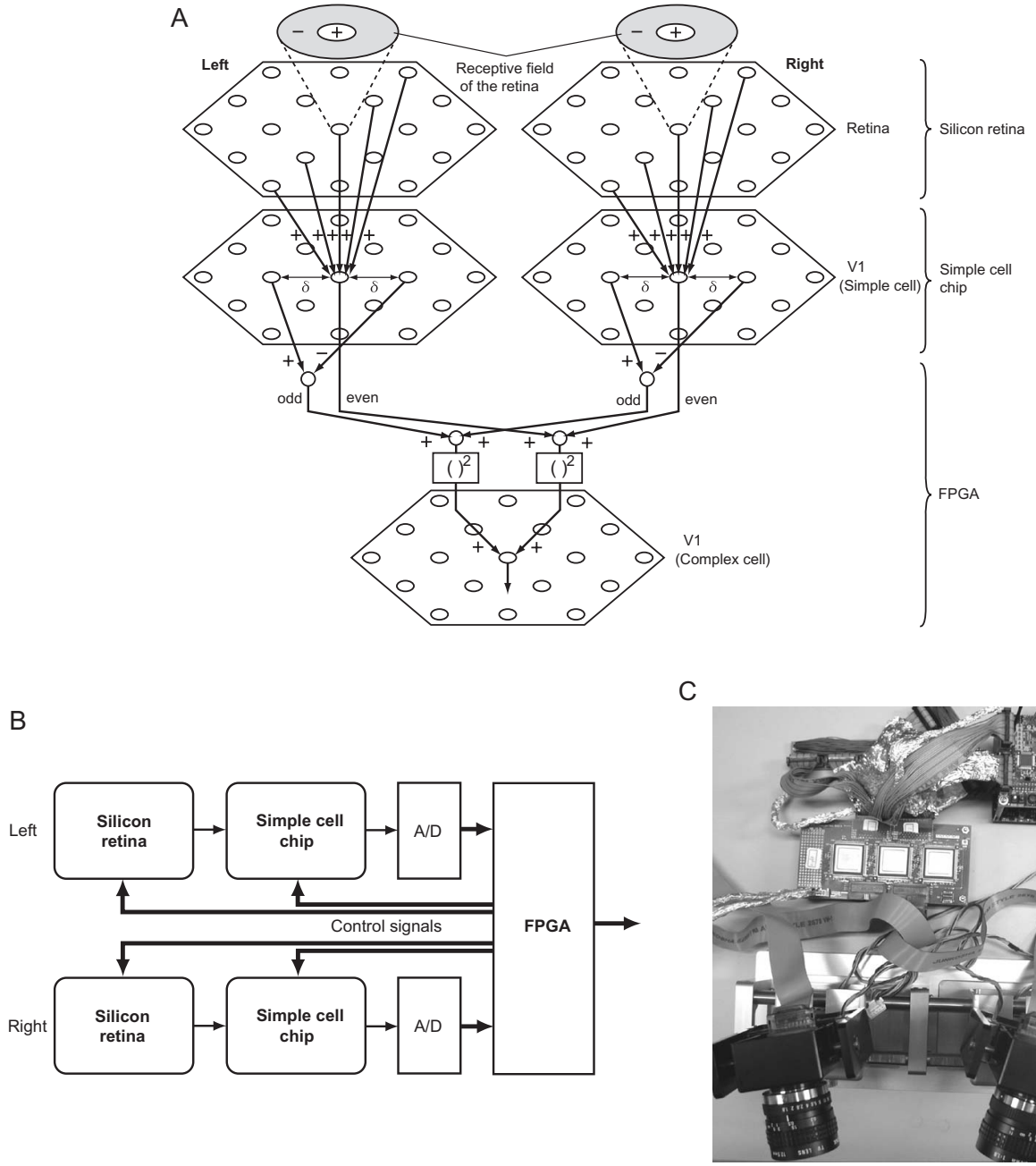

Fig. 3. System structure. (A) Basic architecture of the binocular vision system based on the disparity energy model. Each hexagon shows layers of two-dimensional arrays of neurons. Each circle in a layer represents an individual neuron. The top two layers are the retinas in the left and right eyes. Each retinal neuron has a concentric center-surround receptive field. The middle two layers show simple cell layers in the left and right eyes. Each simple cell receives the output of multiple retinal neurons and generates an orientation-selective receptive field. The receptive field structure of the neuron is of the even-type. An odd-type structure is obtained by subtraction of the outputs of two neurons separated by 2δ (pixels). (B) Block diagram of the binocular vision system. The front-end of the system is the orientation-selective multi-chip system consisting of a silicon retina and a simple cell chip. The outputs of the simple cell chips are read out with sequential scanning and converted to 8-bit digital signals and fed as input to the FPGA. The FPGA provides control signals to drive the silicon retinas, simple cell chips, and motor driver as well as to execute the disparity computation. A stepping motor is installed in each silicon retina camera to move it in the pan direction. (C) Pictures of the binocular vision system. A parallel cable from each camera is connected to the simple cell chip board, which has three simple cell chips. The cables from the simple cell chip board are connected to the FPGA board shown on the upper right side of the picture.

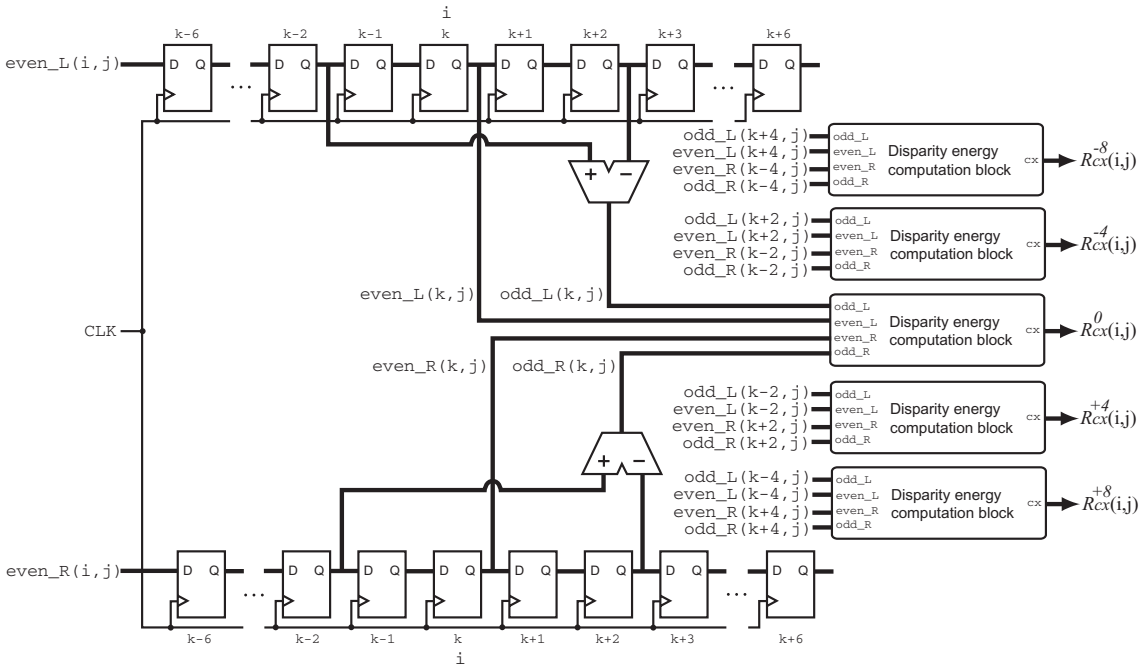

Fig. 5. Block diagram of the disparity computation circuit configured with the FPGA. The output of the simple cell chip, which is read out with sequential scanning, is sent to the shift register after being converted to an 8-bit digital signal. The odd-type response is obtained by subtracting two pixels placed four pixels apart. A pair of data at different locations of the left and right shift registers is sent to a binocular energy computation block that computes the rest of the disparity energy model after spatial filtering. Here, the difference between the addresses of the left and right shift registers defines the preferred disparity d.

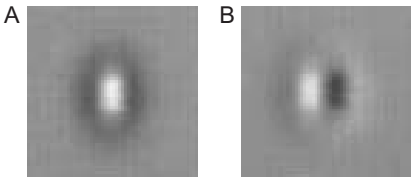

Fig. 4. Output images of the simple cell chip responding to spot illumination. (A)Even-type and (B)odd-type. The preferred orientation is 90° (vertical). The number of aggregated pixels is 8.

B. FPGA circuits for disparity computation

The outputs of the simple cell chips on the left and right sides are sent to the FPGA via the AD converters. In the FPGA, the rest of the disparity energy model is executed after performing spatial filtering with Gabor-like filters. Fig.5 shows the block diagram of the binocular computation circuit configured in the FPGA. Here, even_L and even_R are the outputs of the simple cell chip; they are read out from the chip by horizontal and vertical scanners and converted to 8-bit digital signals. These outputs are fed pixel by pixel to the shift register. Thirteen consecutive pixels (from $\mathtt{k}-6$ to $\mathtt{k}+6$) of the input data, which are aligned horizontally on the simple cell chip, are held in the shift resisters at any instant of time. The input data even_L and even_R are the even-type responses because these responses are originally the output of the simple cell chip. The odd-type responses odd_L and odd_R are obtained as the difference between two pixels horizontally separated by four pixels: $\mathtt{odd\_L(k,j)} = \{\mathtt{even\_L(k-2)} - \mathtt{even\_L(k+2)}\}/2$.

A disparity energy computation block, shown on the right side of the figure, receives the data $\mathtt{even\_L(i,j)}$, $\mathtt{even\_R(i,j)}$, $\mathtt{odd\_L(i,j)}$, and $\mathtt{odd\_R(i,j)}$ from the left and right shift registers. These data from the left and right shift resisters show differences in the addresses corresponding to the horizontal locations on the silicon retina. This difference defines the preferred disparity of the disparity energy computation block. For example, the block for $d = -8$ receives the data of $\mathtt{i} = \mathtt{k} + 4$ from the left shift register and $\mathtt{i} = \mathtt{k} - 4$ from the right one.

In the present study, we configured five energy computation blocks with different preferred disparities $d = -8, -4, 0, +4, +8$ in parallel. Each energy computation block consists of adders and squaring circuits to calculate the response of the individual complex cell as follows.

$$\mathtt{cx} = (\mathtt{even\_L} + \mathtt{even\_R})^2 + (\mathtt{odd\_L} + \mathtt{odd\_R})^2 \qquad (1)$$

This calculation is executed for the entire image by shifting the data pixel by pixel with a clock signal CLK; the complex cell response $R^d_{cx}(i,j)$ in each position is obtained in the field of view (i,j) for each preferred disparity $d \in (-8,-4,0,+4,+8)$.

In the present system, the Gabor-like spatial filtering that mimics the receptive field structure of the simple cells is executed by the analog multi-chip system. The rest of the disparity energy model, which follows the spatial filtering, is a pixel-by-pixel computation. This part was implemented with a hard-wired digital circuit by employing the FPGA.

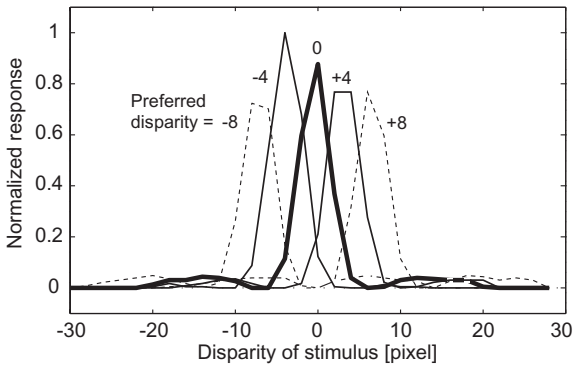

Fig. 6. Disparity tuning curve. Data are measured from the complex cells tuned to disparities of -8, -4, 0, +4, and +8.

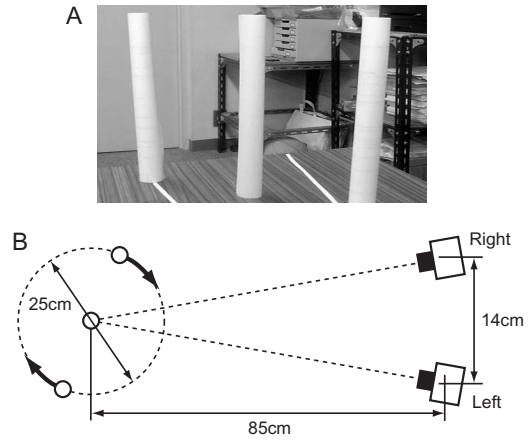

Fig. 7. Experimental setup. (A) Picture of the targets. Three white poles are set on a table with a cluttered background. (B) Arrangement of cameras and targets (view from the top).

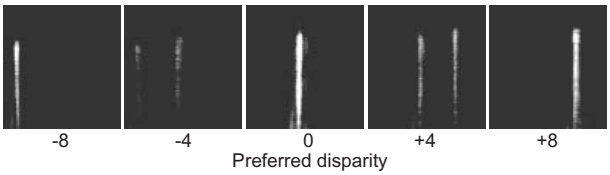

Fig. 8. Disparity energy outputs. Binocular energy outputs are obtained from the complex cell layer tuned to disparities of -8, -4, 0, +4, and +8.

The use of the FPGA allows us to implement multiple disparity energy computation blocks in parallel. Here, a CLK value of 1.4 MHz was used for the data input to the FPGA. This means that the latency for processing is only approximately 7 msec. Moreover, the frame rate of the system is 30 Hz and five disparity energy outputs with different preferred disparities can be obtained in every frame. The equivalent gate count of the FPGA circuit, including five disparity energy computation blocks, was about 16,000.

IV. REAL-TIME DISPARITY COMPUTATION

A. Disparity tuning properties

We first examined the binocular receptive field of the complex cell computed by the present system. 6 shows the disparity tuning curves obtained from the simple cell chip. In the experiment, an image of a vertical white bar on a black background was presented to each silicon retina using an LCD monitor so that the disparity of stimulus could be controlled. Data was collected from a single pixel located at the center of the visual field. The response amplitude was normalized by the maximum response. In the tuning curve of five cells with different preferred disparities, the maximum response was found for the stimulus with the preferred disparity. The width of tuning is approximately 5 pixels at half height and is almost the same for all five cells. The higher precision response of the complex cell makes it possible to compare the response amplitude of five complex cells at the same location so as to generate a disparity map, as described later.

B. Disparity map estimation

A disparity map, which is a two-dimensional map of the binocular disparity estimated for every location of the field of view, is useful for robotic vision to obtain information about the spatial structure of the external world. We computed a disparity map using the disparity energy outputs. The system provides five complex cell responses that have different preferred disparities in parallel, as described in the previous section. Therefore, the disparity map can be computed by comparing the responses of these five complex cells that have their receptive field at the same position in the field of view. We estimated the disparity $D(i,j)$ in each position as follows.

$$D(i,j) = \arg\max_{d} R^{d}_{cx}(i,j) \quad d \in (-8,-4,0,+4,+8) \tag{2}$$

This computation was implemented using the FPGA. The circuit for disparity map estimation carries out the above computation using five disparity energy outputs.

We executed a real-time disparity map estimation using the present binocular vision system in an indoor scene. Fig.7 shows the experimental setup. Three white poles were placed on a table in front of a cluttered background, as shown in (A). Cameras placed on the left and right sides of the setup were focused on the center pole placed 85 cm away from the cameras. The distance between both the cameras was 14 cm. In this setting, 1 pixel of disparity corresponds to around 3.5 cm. This resolution is generally sufficient for mobile robots to segregate multiple objects based on differences in depth.

Fig.8 shows the disparity energy output of the system at a certain moment. The poles located on the left, center, and right give strong responses with energy outputs of -8, 0, and +8, respectively. The background has little or no response because it is out of range for a calculated disparities and lens focus. These five energy outputs are computed in parallel.

Fig.9 shows the time course of the disparity map obtained from the system. The disparity map is obtained every 33 ms. In this experiment, poles on both sides rotate around the center pole. The estimated disparity is shown by the

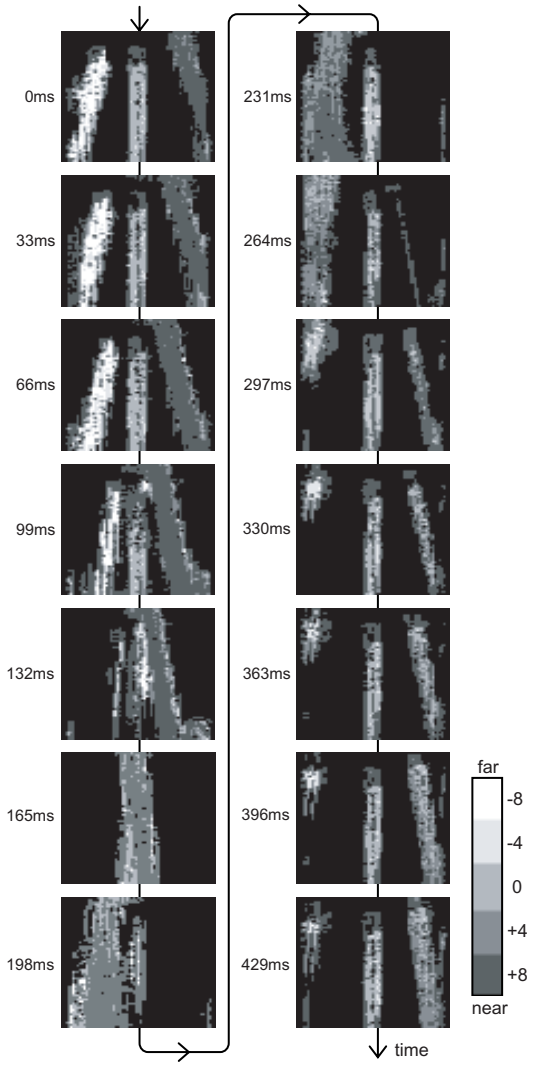

Fig. 9. Disparity map obtained in real-time (33 ms in frame period). The estimated disparity is shown by the gray scale. The brighter and darker pixels correspond to farther and nearer locations, respectively. The pole located at the center is fixed and has zero disparity. The poles on both sides rotate around the center pole.

gray scale. Brighter pixels show negative disparities corresponding to farther locations, and darker pixels show positive disparities corresponding to nearer locations. Black pixels indicate $\max R^d_{cx}(i,j) \leq threshold$ at the pixel. For these pixels, the disparity is not estimated because the reliability of the estimated disparity is lower for very small energy responses. At 0 ms, the left pole is located far away and the right one is located nearer. These poles are moving such that they rotate around the center pole; thus, the left pole is moving toward the right behind the center pole and the right one is moving toward the left in front of the center pole. From 165 ms to 264 ms, the nearest pole is moving backward and the disparity is well estimated. From 297 ms, the other pole appears on the right and is moving forward. The moving poles can be distinguished based on the estimated disparity in real-time.

V. DISCUSSION AND CONCLUSION

In the present study, we designed and constructed a binocular vision system mimicking the architecture of V1. The system consists of multiple analog VLSIs that are used to compute spatial filtering in real-time. Then, an FPGA executes pixel-wise disparity computation using simple cell chip outputs and provides a disparity map in real-time. In the present system, the spatial filtering emulating the simple cell function is executed by an analog multi-chip system, and the disparity energy is computed by a hard-wired digital circuit implemented using the FPGA. Such an analog-digital mixed architecture that combines neuromorphic sensors and conventional digital technology is thought to be a practical strategy for implementing the essence of the computation in the brain. This is because such a system can have the advantages of both neuromorphic chips and conventional digital systems, namely, power efficiency, compactness, real-time computation, and programmability.

The disparity energy computation used in this study is the first step to realizing stereoscopic vision in the brain [12]. By expanding the present system, we can develop a vision system that emulates the more advanced stages of stereoscopic processing in higher order areas of the visual cortex.

ACKNOWLEDGEMENT

The VLSI chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo. with the collaboration by Rohm Corporation and Toppan Printing Corporation.

REFERENCES

[1] G.F.Poggio, F.Gonzalez, F.Krause, "Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity," *J.Neurosci.*, vol.8, pp.4531-4550, 1988.

[2] C.Mead, "Neuromorphic electronic system," *Proc.IEEE*, vol.78, pp.1629-16336, 1990.

[3] M.Mahowald, *An analog VLSI system for stereoscopic vision*, Kluwer, Boston, MA, 1994.

[4] E.K.C.Tsang, B.E.Shi "A preference for phase-based disparity in a neuromorphic implementation of the binocular energy model," *Neural Computation*, vol.16, pp.1579-1600, 2004.

[5] I.Ohzawa, G.DeAngelis and R.Freeman, "Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, vol.249, pp.1037-1041, 1990.

[6] N.Qian, "Binocular disarity and the perception of depth," *Neuron*, vol.18, pp.359-368, 1997.

[7] D.Hubel and T.Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J.Physiology (London)*, vol.160, pp.106-154, 1962.

[8] R.Takami, K.Shimonomura, S.Kameda and T.Yagi, "A novel preprocessing vision system employing neuromorphic 100×100 pixel silicon retina," In *Proc. IEEE Intl. Symp. on Circuits and Systems*, Kobe, Japan, pp.2771-2774, 2005.

[9] S.Kameda and T.Yagi, "An analog VLSI chip emulating sustained and transient response channels of the vertebrate retina," *IEEE Trans. on Neural Networks*, vol.14, no.5, pp.1405-1412, 2003.

[10] T.Yagi, S.Ohshima and Y.Funahashi, "The role of retinal bipolar cell in early vision: an implication with analogue networks and regularization theory," *Biological Cybernetics*, vol.77, pp.163-171, 1997.

[11] K.Shimonomura and T.Yagi, "A multi-chip aVLSI system emulating orientation selectivity of primary visual cortical cell," *IEEE Trans. on Neural Networks*, vol.16, no.4, pp.972-979, 2005.

[12] P.Neri, "A stereoscopic look at visual cortex," *J.Neurophysiol.*, vol.93, pp.1823-1826, 2005.