Texture Segregation Employing Orientation-Selective Analog Multi-chip Vision System

Kazuhiro Shimonomura\textsuperscript{1} and Tetsuya Yagi\textsuperscript{1}
\textsuperscript{1}The Center for Advanced Medical Engineering and Informatics, Osaka University
\textsuperscript{1}Graduate School of Engineering, Osaka University
2-1 Yamadaoka, Suita, Osaka 565-0871, Japan
Email: \{kazu, yagi\}@ele.eng.osaka-u.ac.jp

Abstract—A high resolution neuromorphic multi-chip vision system was fabricated to emulate the orientation selective response of the simple cell in the primary visual cortex. The vision system consists of two types of analog chips: a silicon retina and a simple cell chip. The center-surround concentric receptive fields of the silicon retina are aggregated in the simple cell chip, mimicking the hierarchical architecture in the visual system of the brain. Both chips have 100\times100 pixels. Using the orientation-selective outputs obtained from the multi-chip system, a texture segregation was conducted based on a computational model inspired by neurophysiology. The texture image was filtered by the two orthogonally oriented receptive fields of the multi-chip system and the filtered images were combined to segregate the area of different texture orientation with the aid of Field Programmable Gate Array (FPGA). The present multi-chip system is useful to emulate and verify computational models for texture segregation of the cortical cells.

I. INTRODUCTION

Texture segregation is an important component to distinguish figure from ground or one object from neighboring objects. It requires a high computational cost to the conventional image processing systems which operates with a serial algorithm. In human vision, the texture segregation occurs rapidly and unconsciously even in the image with uniform local average intensity and spatial frequency.

We fabricated a multi-chip aVLSI system to emulate the orientation selective response of the simple cell in the primary visual cortex, aiming at developing a novel image processing device to be utilized for real-time image computation in engineering applications [1], [2]. The multi-chip system consists of a silicon retina and a simple cell chip, mimicking the hierarchical architecture of the visual system in the brain. A number of neuromorphic aVLSI circuits that emulate the orientation selective receptive field have already been fabricated with a monolithic [3], [4] or a multi-chip [5], [6], [7], [8], [9] configurations. Our multi-chip system used an analog image transfer technique for inter-chip communication, considering its utility for practical engineering applications, while previous neuromorphic multi-chip systems used the the address-event representation (AER) protocol [10], [11]. In the present study, we applied the multi-chip system for real-time texture segregation.

II. MODEL FOR TEXTURE SEGREGATION

Previously, a number of computational models for texture segregation were proposed using orientation-selective spatial filters which resemble to the receptive field of simple cells in the primary visual cortex (e.g., [12]). We employed a similar algorithm to segregate the texture of different orientations.

The model used in this study is shown in Fig.1. The model consists of four layers; orientation-selective layer, rectification layer, spatial pooling layer, and subtraction layer. An input image is firstly filtered by center-surround spatial filters, similar to the receptive field of lateral geniculate nuclear (LGN) neurons in mammalian visual system. The simple cells in the primary visual cortex, which receive the signal from the LGN neurons, have an elongated receptive field and respond specifically to oriented bar or edge. This orientation-selective receptive field can be produced by aggregates the responses having the center-surround receptive field [13]. We used this feed-forward model to realize the orientation-selective receptive field. The aggregation is carried out along the orientation of 0\degree (horizontal) and 90\degree (vertical) in parallel. The orientation-selective outputs are full-wave-rectified and sent to spatial pooling stage. Here, we employed a maximum operation [14]. The responses of neighboring pixels aligned perpendicular to the preferred orientation are compared and the maximum response is selected to generate a complex cell-like response.

Finally, the difference between outputs of orthogonal orientations is computed. The model discriminates the area of different texture orientation as the difference in response magnitude, in which the texture edge is enhanced.

The model employed here is suitable for hardware implementation with the silicon retina and the simple cell chip.

III. SYSTEM STRUCTURE

Fig.2 shows the block diagram to implement the model. The system consists of a silicon retina, two simple cell chips and an FPGA.

The original silicon retina was designed by Kameda and Yagi [15]. The chip was improved to 100\times100 pixels [16]. In the chip, the pixel circuits including photodiodes (PDs)
Fig. 1. Schematic of the model for texture segregation. Input image is filtered by $0^\circ$ and $90^\circ$ orientation filters consisting of feed-forward aggregation of the center-surround receptive field. This output is full-wave rectified and sent to spatial pooling stage to generate complex cell-like response. Finally, difference between $0^\circ$ and $90^\circ$ complex cell responses is calculated.

Fig. 2. Block diagram of the system. Input image is received by the silicon retina. The output of the retina is sent to two simple cell chips tuned to $0^\circ$ and $90^\circ$, respectively. The outputs of the simple cell chips are fed to FPGA through AD converter (ADC).
are arranged in a hexagonal grid. The spatial response of the silicon retina exhibits a Laplacian-Gaussian $(\nabla^2 G)$-like center-surround receptive field. There have been many chips designed to emulate the center-surround receptive field [17]. In the silicon retina used here, the fixed-pattern noise is reduced by embedded offset compensation circuits [15]. The analog voltage outputs of the silicon retina are read out using the shift resisters (VSR and HSR) in sequence and transferred to the corresponding pixel on the simple cell chip.

The simple cell chip also possesses a $100 \times 100$ pixels circuit array and six shift registers. Each pixel circuit includes an analog memory to hold the analog voltage input from the silicon retina. The image represented by analog voltages in the silicon retina is transferred to the simple cell chip pixel by pixel using the horizontal and vertical shift registers (VSI and HSI). After the retinal image is transferred to the simple cell chip, average of the outputs from multiple pixels of the silicon retina is computed with a follower aggregation circuit [18] to obtain the orientation selective outputs. The number of pixels to aggregate is defined by length of DATA signal provided to shift registers, VS and HS. Output image of the simple cell chip is then read out using horizontal and vertical shift resisters, VSO and HSO [2].

The orientation of the simple cell chip can be obtained for $0^\circ$, $60^\circ$, and $120^\circ$. The $30^\circ$, $90^\circ$, and $150^\circ$ orientation responses can be obtained as well by rotating the image at $90^\circ$ when the image is transferred from the retina to the simple cell chip. In this study, two simple cell chips tuned to $0^\circ$ and $90^\circ$ are used to compute the texture.

The analog outputs of the simple cell chips are converted to 8-bit digital signal by video A/D converter (Sony CXD1175), and fed to FPGA (Xilinx XCV300) to conduct computations following the orientation tunings. The computations, i.e. rectification, one dimensional pooling and subtraction, are carried out during the read-in to the FPGA. Therefore, texture segregation is carried out at frame rate of the system, typically 30Hz in indoor illumination. The outputs are obtained as NTSC video signal. The FPGA also provides control signals to drive the silicon retina and the simple cell chips.

Fig.3 shows the photograph of the multi-chip system in the present study. A CCTV lens (Pentax B2514D) was mounted on the silicon retina to which an simple cell chip board was connected. The power consumption of the system was about 3W including FPGA board.

IV. RESPONSE OF THE SYSTEM

The texture pattern, shown in Fig.1, was presented to the multi-chip vision system. The texture orientation of the center area is different at $90^\circ$ from that of the surround area, although local average of intensity and spatial frequency of these areas are the same. Fig.4(a) and (b) show the output images obtained from the simple cell chip. The preferred orientation was $0^\circ$ and $90^\circ$ in (a) and (b), respectively. The number of aggregated pixels of the silicon retina was eight. The filter size of the silicon retina was set so that its spatial frequency tuning matches to the input texture pattern. As shown in the Fig., the texture with the preferred orientation is enhanced and the orthogonally oriented texture to the preferred orientation is blurred. In the FPGA, the full-wave rectification and the maximum operation were carried out to generate complex cell-like responses. The kernel size to find the maximum response was 8 pixels. (c) and (d) are complex cell-like responses computed with the FPGA. The preferred orientation was $0^\circ$ and $90^\circ$ in (c) and (d), respectively. As shown in the Fig., the area in the circle responds strongly in (c), and the surround area responds strongly in (d). Output of the model is obtained as the difference between (c) and (d). Fig.5 shows diagonal section of the output voltages. The bold line is the output voltage of the system in response to the input image shown in Fig.1. The difference in texture orientation between the two regions is converted into the difference in response intensity. These two regions can be segregated by the zero level. The boundary between two areas can be detected using conventional edge detection algorithm, e.g., zero-crossing.

Gray lines in the Fig. show the outputs for the input images rotated at $15^\circ$ (b), $30^\circ$ (c) and $45^\circ$ (d). Response amplitude decreases as the texture orientation deviates from $0^\circ$ or $90^\circ$ because the difference of response amplitude between $0^\circ$ neuron and $90^\circ$ neuron becomes smaller due to the orientation tuning property of the simple cell chip. To reduce the dependence of output intensity on the orientation of the input texture, other quadrature pairs in the orientation of the simple cell chip, such as $30^\circ$ and $120^\circ$, have to be combined.

V. CONCLUSIONS

In the present study, we implemented a neuromorphic multi-chip system to conduct real-time texture segregation. The architecture of the system mimics the hierarchical structure of the visual system of the brain. The orthogonally orientated texture areas were segregated using a similar algorithm proposed in previous computational models. The texture segmentation requires a high computational cost, such as power consumption, to conventional digital image processing systems. The present multi-chip system operated with 3W, and therefore is applicable to visual navigation of autonomous machines.
VI. 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