

# Column-Parallel Vision Chip Architecture for High-Resolution Line-of-Sight Detection Including Saccade

Junichi AKITA<sup>†a)</sup>, Member, Hiroaki TAKAGI<sup>†</sup>, Keisuke DOUMAE<sup>†</sup>, Nonmembers, Akio KITAGAWA<sup>†</sup>, Masashi TODA<sup>††</sup>, Takeshi NAGASAKI<sup>††</sup>, and Toshio KAWASHIMA<sup>††</sup>, Members

**SUMMARY** Although the line-of-sight (LoS) is expected to be useful as input methodology for computer systems, the application area of the conventional LoS detection system composed of video camera and image processor is restricted in the specialized area, such as academic research, due to its large size and high cost. There is a rapid eye motion, so called 'saccade' in our eye motion, which is expected to be useful for various applications. Because of the saccade's very high speed, it is impossible to track the saccade without using high speed camera. The authors have been proposing the high speed vision chip for LoS detection including saccade based on the pixel parallel processing architecture, however, its resolution is very low for the large size of its pixel. In this paper, we propose and discuss an architecture of the vision chip for LoS detection including saccade based on column-parallel processing manner for increasing the resolution with keeping high processing speed.

**key words:** line-of-sight (LoS), saccade, vision chip, column-parallel processing

## 1. Introduction

Our eyes' line-of-sight (LoS), the direction where we are looking at, is expected to be applied for various types of human-computer interfaces. We also have a rapid eye motion, so called saccade, which is a very quick eye motion that always occurs regardless of our intention. Although the saccade is expected to be useful for new types of human-computer interfaces [1], [2], it is impossible to track it with the conventional video camera because of its high speed which is often up to 600 degrees per second. The frame rate of approximately 500 [fps] (frame per second) is required to track the saccade eye motion, and the angular accuracy in LoS detection of approximately 0.5 degrees is required for practical applications [3]. Some commercial systems for LoS detection including saccade are available, such as Eye-LinkII [3], which consists of high speed camera and PC-based image processor, however, the latency of outputting LoS detection result is quite large, for example which is up to 14 [ms] in EyeLinkII [3], for its pipe-lined image processing architecture, and in addition, it is difficult to implement a compact system for portable application.

On the other hand, there are a lot of studies on 'Vision Chip,' which is an intelligent image sensor that has photo

receptors and processing circuits on one chip [4]–[6]. Vision chip can process the acquired image information by keeping its spatial parallelism, and can achieve very fast image processing speed, as well as compact size of image processing system as an integration of photo receptors and image processors. Generally, the image processing in vision chip is performed in pixel parallel manner, and this architecture intrinsically results in low resolution because of pixel circuit size, such as  $64 \times 64$  pixels in [6]. We have already reported a vision chip architecture for LoS detection including saccade in pixel parallel architecture [7], however, the proposed vision chip has the resolution of  $16 \times 16$  pixels in 2.3 mm square chip using  $0.35 \mu\text{m}$  CMOS technology. The low resolution of vision chip is one of the most critical reasons that prevents commercial products of vision chips. Column-parallel processing architecture is one of the suitable solutions for vision chip to achieve drastically increased resolution [8], [9]. However, the column-parallel architecture has the difficulty in implementing an inter-frame image processing, because of the restriction of frame memory manipulation for image processing.

In this paper, we describe a column-parallel vision chip architecture for detecting the line-of-sight from infrared eye image, with the capability of the saccade tracking. The image processing element (PE) is placed at each column to perform the specified algorithm to detect the line-of-sight in collaboration with flag memory and dedicated working memories. We also describe the preliminary estimation for the proposed vision chip architecture in terms of both operation speed and resolution.

## 2. Algorithm for Line-of-Sight Detection

### 2.1 LoS and Eye Image

Infrared eye image is suitable to detect the line-of-sight. Figure 1 shows an example of infrared eye image. Two characteristic areas appear in the infrared eye image. One is the pupil at the center of the iris, which appears as a black area, and the other is the Purkinje's image, which appears as a white spot around or inside the pupil. The line-of-sight can be detected by the relation of these two images' positions as shown in Fig. 2.

Manuscript received February 19, 2007.

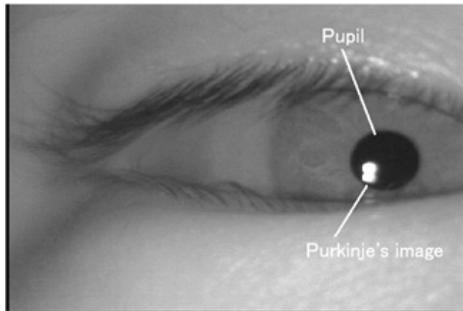
Manuscript revised May 10, 2007.

<sup>†</sup>The authors are with Kanazawa University, Kanazawa-shi, 920-1192 Japan.

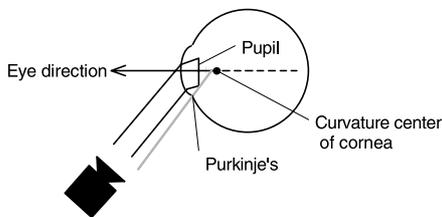
<sup>††</sup>The authors are with Future University-Hakodate, Hakodate-shi, 041-8655 Japan.

a) E-mail: akita@is.t.kanazawa-u.ac.jp

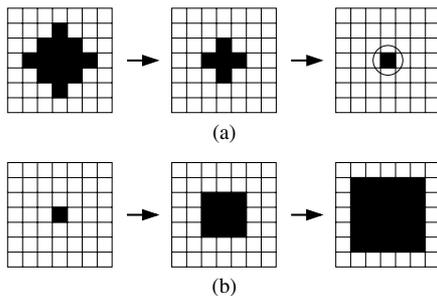
DOI: 10.1093/ietele/e90-c.10.1869



**Fig. 1** Infrared eye image.



**Fig. 2** Principle of LoS detection.



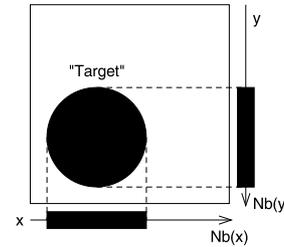
**Fig. 3** Shrink and expansion operations.

## 2.2 Image Processing for LoS Detection

The authors have indicated that the positions of both the pupil and the Purkinje's image in the infrared eye image can be detected by the following five operation procedures [7].

1. Extraction of the Purkinje's image according to pixel intensity.
2. Shrink operations and elimination detection for Purkinje's image, as shown in Fig. 3(a).
3. Extraction of the pupil according to pixel intensity and logical-OR with the Purkinje's image.
4. Expansion operations for the pupil image, as shown in Fig. 3(b).
5. Shrink operations and elimination detection for the pupil image, as shown in Fig. 3(a).

The extractions of the Purkinje's image and the pupil in steps 1. and 3. can be performed by comparison of pixel intensity with the certain reference. The shrink and expansion operations in steps 2. and 5. can be performed by cellular automaton at each pixel, which makes transition according to



**Fig. 4** Projection operation.

the values of neighboring automata. For example, the transition rule for shrink operation is implemented as the logical-AND of values of four neighboring automata and itself; in other words, it keeps '1' when four neighboring automata and itself has the values of '1.' The detection of eliminating images can be performed by counting the number of pixels of the projected images onto axes, as shown in Fig. 4. Its position can also be calculated as its  $x$ - and  $y$ -coordinates by using the priority encoders placed on both axes.

There are unnecessary areas other than the pupil and the Purkinje's image in infrared eye image, such as eyelashes and skins, and they may be extracted by intensity thresholding. The position detection performed by the shrink operations can eliminate these unnecessary areas prior to the detections of the pupil and the Purkinje's image eliminations by appropriate shrink operation rules [7].

Another possible method for detecting the positions of the pupil and the Purkinje's image is to calculate 1st momentum of these areas [6], which can calculate the positions up to sub-pixel order. However, adequate masking operations are required prior to momentum calculations in order to calculate the proper positions of more than one areas. The position detection based on shrink operations described above has the advantages of no masking operations, including mask selection, and simple circuit composed of finite state machine against that based on momentum calculation.

As described above, the required functional blocks of the vision chip for LoS detection can be summarized as follows.

- Comparators for extraction of the Purkinje's image and the pupil.
- Automaton circuit to perform shrink and expansion operations.
- Detector and coordinate generators of eliminating image.

## 3. Column-Parallel Architecture for LoS Detection

Figure 5 shows the proposed column-parallel vision chip architecture for one column to perform LoS detection algorithm described in the previous section. The circuit of one column consists of the following components.

- Photo receptor (PIX)
- Flag memory (MEM)
- Processing element (PE)

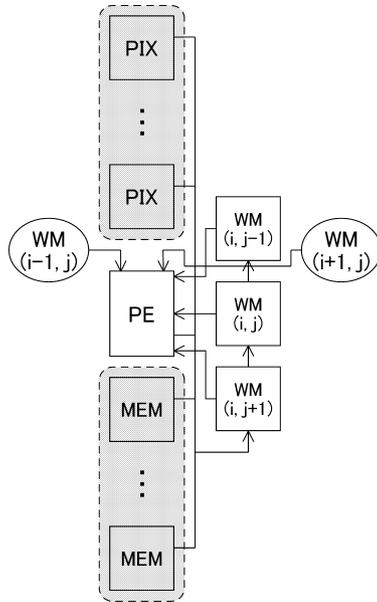


Fig. 5 Column-parallel architecture for LoS detection.

- Working memory (WM)

Photo receptor (PIX) receives the photo signal and converts it to voltage signal.

Flag memory (MEM) is a single bit memory array corresponding to PIXs, which holds the flag indicating the Purkinje’s image or the pupil in each shrink or expansion procedures.

Processing element (PE) performs the extraction of the Purkinje’s image and the pupil according to the signal from PIXs, shrink and expansion operations, and the projection operation for detecting elimination of the image.

Working memory (WM) consists of three sequential single bit memories with shift operation capability, which holds the flags at the upper ( $WM(i, j-1)$ ), the current ( $WM(i, j)$ ), and the lower ( $WM(i, j+1)$ ) pixels of corresponding row in process for each operation of shrink and expansion for one row. The PE also uses the flags of left ( $WM(i-1, j)$ ) and right ( $WM(i+1, j)$ ) of corresponding row in process, which are given from the PEs in neighboring columns.

Figure 6 shows the operation sequence, where operation means whether shrink or expansion operations according to the procedure steps. First, the operation for row  $i=2$  is performed by processing element (PE) with the flags in working memories (WMs), which hold the flags at rows  $i=1$ ,  $i=2$ , and  $i=3$ . The result of the operation is stored in flag memory at row  $i=2$ , MEM2, as shown in Fig. 6(a). Next, the operation for next row  $i=3$  is performed, which requires the flags at rows  $i=2$ ,  $i=3$ , and  $i=4$ . The contents of WMs are shifted prior to this operation, and the content of  $WM(i, j+1)$  is updated as the content of MEM4. The result of the operation is stored in flag memory at row  $i=3$ , MEM3, as shown in Fig. 6(b). The following operation for next row  $i=4$  is also performed in the same manner, as shown in Fig. 6(c).

Here, we describe the detailed configuration of each

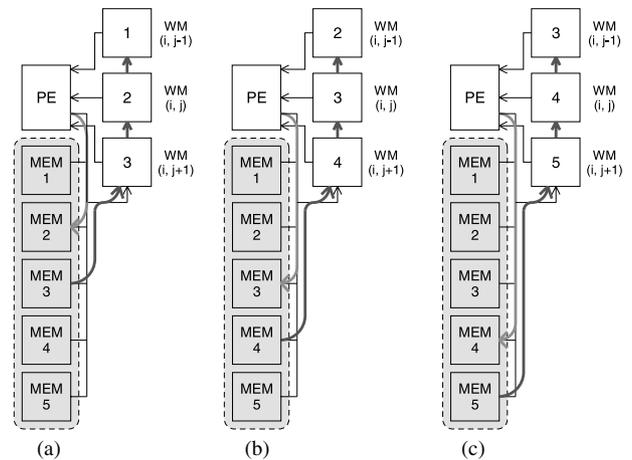


Fig. 6 Operation sequence.

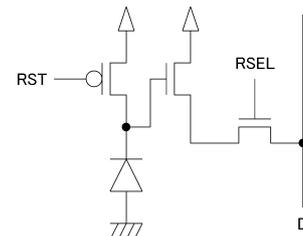


Fig. 7 Circuit of photo receptor.

functional block in the following sections.

### 3.1 Photo Receptor (PIX)

Figure 7 shows the circuit of photo receptor, PIX, which is a typical 3-Tr type CMOS image sensor pixel. The signal is output to data line, D, when the corresponding row is selected by row select signal, RSEL.

### 3.2 Flag Memory (MEM)

Flag memory, MEM, is a simple single bit memory, such as SRAM. However, it is required to have the capability of simultaneous store and read operations, so the dual port SRAM is adequate for MEM.

### 3.3 Working Memory (WM)

Working memory, WM, consists of three single bit memories, and it is required to perform shift operation in each operation step. The conventional shift register consists of three D flip-flops is adequate for WM.

### 3.4 Processing Element (PE)

Processing elements, PE, consists of following functional blocks as shown in Fig. 8.

- Flag generator (FG)
- Memory access controller (MC)

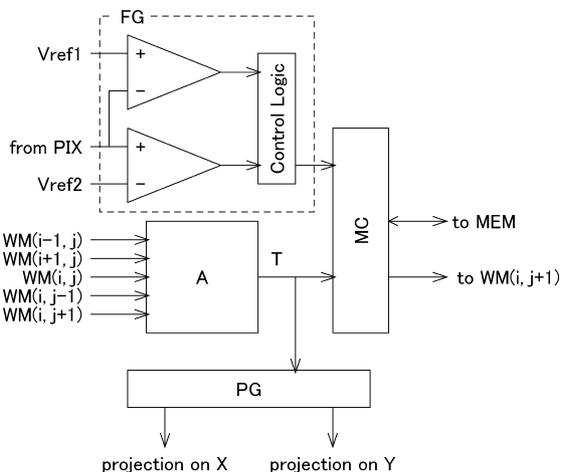


Fig. 8 Circuit diagram of processing element.

- Automaton for shrink and expansion operations (A)
- Projection generator (PG)

Flag generator, FG, generates the flags indicating the Purkinje's image in procedure step 1. in Sect. 2.2 by comparisons with the certain references,  $V_{ref1}$  and  $V_{ref2}$ , using voltage comparators. It also generates the flag indicating the pupil in procedure step 3. in Sect. 2.2 by two voltage comparators and logical OR in control logic.

Memory access controller, MC, controls the memory accesses in procedure steps of 2., 4., and 5. in Sect. 2.2; reading flag from MEM and storing it to  $WM(i, j+1)$ , storing the operation result to MEM.

Automaton, A, calculates the flags in the next operation step, T, to perform shrink or expansion operations according to the flags of neighboring pixels stored in WMs.

Projection generator, PG, generates the projected image of flags onto both  $x$ - and  $y$ -axes. The projection onto  $x$ -axis can be generated by accumulated logical-OR operations in collaboration with D flip-flop during the operation procedures for all the rows in one column. The projection onto  $y$ -axis can be generated by a chain of logical-OR gates. Figure 9 shows the circuit diagram of PGs. The wired OR circuit should be employed for realistic implementation in order to reduce the propagation delay.

### 3.5 Global Controller

Each functional block is controlled by global controller to perform the dedicated operation procedure steps for LoS detection described in Sect. 2.2.

## 4. Layout Design of Circuit Components

Here we describe the designed circuits of the components of the column-parallel implementation of LoS detector using CMOS 0.35  $\mu\text{m}$ , four layers of metal technology.

Figure 10 shows the designed layout of photo receptor, PIX. The cell size is 14.3  $\mu\text{m}$   $\times$  14.3  $\mu\text{m}$ , with fill factor of 18%. The cell size is determined to adjust the width of

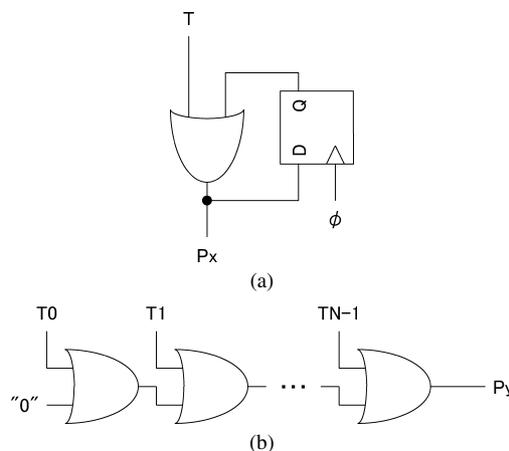


Fig. 9 Circuit for projection generation on (a)  $x$ -axis and (b)  $y$ -axis.

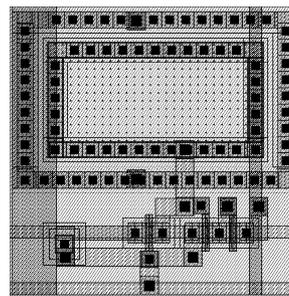


Fig. 10 Designed layout of photo receptor, PIX.

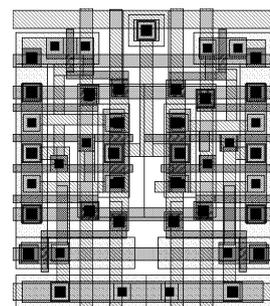


Fig. 11 Designed layout of flag memory, MEM.

processing element, PE described below for regular layout of whole vision chip. The cell size of PIX can be different from the width of PE, for example, larger pixel to increase photo sensitivity, or smaller pixel to increase resolution, with increased area of connection wires between PIXs and other components, PEs and MEMs.

Figure 11 shows the designed layout of flag memory, MEM for the unit of two bits. The cell size is 14.3  $\mu\text{m}$   $\times$  12.7  $\mu\text{m}$  for the unit of two bits. The width of MEM is also determined to adjust the width of processing element, PE.

Figure 12 shows the designed layout of processing element, PE, including three working memories, WMs. The cell size is 14.3  $\mu\text{m}$   $\times$  485.3  $\mu\text{m}$ .



**Fig. 12** Designed layout of processing element, PE, including three working memories, WMs.

## 5. Performance Estimation

Here, we describe the preliminary estimations of the proposed column-parallel vision chip architecture in terms of both operation speed and resolution.

### 5.1 Estimation on Operation Speed

The total operation time required for one image frame is the sum of the processing time of PE,  $T_p$ , and the signal transfer time from photo receptors to flag memories including flag generation,  $T_t$ . The number of PE operations in column-parallel processing architecture for whole pixel plain is equal to the number of rows in the pixel plain. Assuming that the number of both rows and columns are  $N$ , the number of PE operations for one shrink or expansion operation is equal to  $N$ . Since one PE operation is carried out in single clock cycle, the total clock cycles required for one shrink or expansion operation is also equal to the number of rows,  $N$ .

The number of shrink or expansion operations to be carried out is estimated as  $N/2$  at maximum, because both sides of the image eliminate in one shrink operation step, and after  $N/2$  shrink operations, all the flags will completely eliminate. The shrink operations are carried out for both the Purkinje's image and the pupil. On the other hand, the number of expansion operations can be ignored, since it is small enough compared with the number of shrink operations. Thus, the total maximum number of shrink or expansion operations is equal to  $N/2 \times 2 = N$ . (In the practical situations, the number of shrink operations will be smaller than  $N$ , since the Purkinje's image is much smaller than the pupil, and it requires less number of shrink operations to eliminate.) The processing time,  $T_p$ , performed by processing element, PE, for one image frame is calculated as follows.

$$T_p = t_p \cdot N^2 \quad (1)$$

Here,  $t_p$  is the clock cycle of PE operation.

The signal transfer time,  $T_t$ , can be described as follows.

$$T_t = t_t \cdot N \quad (2)$$

Here,  $t_t$  is the signal transfer time for one pixel in row including flag generation.

The total operation time for one image frame,  $T$  is described as follows.

$$T = T_p + T_t = t_p \cdot N^2 + t_t \cdot N \quad (3)$$

Assuming that the required frame rate for LoS detection including saccade as 500 [Hz] [3], [7], or  $T$  of 2 [ms],

$N$  as 480 for VGA resolution ( $640 \times 480$  pixels), and  $t_t$  of 1 [ $\mu$ s] for practical image sensors, the required operation frequency,  $f$ , for PE, WM, and MEM can be estimated as follows.

$$f = \frac{1}{t_p} = \frac{N^2}{T - t_t \cdot N} = 152 \text{ [MHz]} \quad (4)$$

This clock frequency can be achieved by using the recently available CMOS technologies.

### 5.2 Estimation on Resolution

The critical factor that decides the resolution of the vision chip with column-parallel architecture is the width of processing element (PE) and working memory (WM), since the widths of photo receptor (PIX) and flag memory (MEM) are expected to be smaller than those of PE or WM.

The width of designed circuits in one column using CMOS 0.35  $\mu$ m technology is 14.3 [ $\mu$ m]. Based on this size of PE, the width of focal plain for the resolution of QVGA ( $320 \times 240$  pixels) is estimated as 14.3 [ $\mu$ m]  $\times$  320 = 4.6 [mm], which is expected to be a practical size of image sensor. The resolution will be increased by using the more scaled CMOS process, such as VGA ( $640 \times 480$  pixels) in 4.7 [mm] using 0.18  $\mu$ m CMOS technology.

In spite of this size of width, the height of focal plain is larger than the width, twice for example, since the height of focal plain is the sum of that of PE and WMs, as well as PIXs and MEMs, that occupy the great part in the height of one column. The rectangular shape of vision chip is adequate for implementation.

Assuming that whole eye ball is projected onto focal plain composed of  $N$  pixels in one row, whose size is  $X$ , and the distance between eye and the vision chip sensor of  $L$ , the minimum expected angular error,  $\Delta\theta$ , is calculated as follows.

$$\Delta\theta = \tan^{-1} \frac{X/N}{L} \approx \frac{X/N}{L} \quad (5)$$

To achieve angular error of 0.5 [deg] for practical applications with assumption of  $X = 10$  [mm] and  $N = 480$  as VGA resolution,  $L$  should be larger than 2.6 [mm], which is realistic in practical applications. The high resolution vision chip is expected to be useful for various applications of image sensing other than LoS detection proposed in this paper.

## 6. Conclusion

In this paper, we proposed the column-parallel architecture for vision chip that has the function of LoS detection. The resolution of the vision chip is drastically increased by employing column-parallel architecture, with keeping the sufficient operation speed supporting saccade tracking.

We estimated the LoS detection at 500 [Hz] with clock frequency of 152 [MHz] for VGA resolution. We also estimated QVGA and VGA resolution using CMOS 0.35  $\mu\text{m}$  and 0.18  $\mu\text{m}$  CMOS technologies, respectively, with chip width of 4.6 [mm].

The detailed design and evaluation of LoS detection vision chip using column-parallel architecture will be reported in our future work.

## References

- [1] J. Triesch, D.H. Ballard, M.M. Hayhoe, B.T. Sullivan, "What you see is what you need," *J. Vision*, vol.3, no.1, pp.86–94, 2003.
- [2] S.S. Intille, "Change blind information display for ubiquitous computing environments," *Proc. UbiComp2002*, pp.91–106, 2002.
- [3] <http://www.monte.co.jp/srr01.htm>
- [4] T. Komuro, S. Kagami, and M. Ishikawa, "A dynamically reconfigurable simd processor for a vision chip," *IEEE J. Solid-State Circuits*, vol.39, no.1, pp.265–268, 2004.
- [5] A. Astrom, R. Forchheimer, and J.E. Eklued, "Global feature extraction operations for near-sensor image processing," *IEEE Trans. Image Process.*, vol.5, no.1, pp.102–110, 1996.
- [6] T. Komuro, et al., "A digital vision chip specialized for high-speed target tracking," *IEEE Trans. Electron Devices*, vol.50, no.1, pp.191–199, 2003.
- [7] J. Akita, H. Takagi, T. Nagasaki, M. Toda, T. Kawashima, and A. Kitagawa, "Vision chip architecture for detecting line of sight including saccade," *IEICE Trans. Electron.*, vol.E89-C, no.11, pp.1605–1611, Nov. 2006.
- [8] K. Imai, H. Taruki, T. Hamamoto, and H. Ohtake, "Image compression sensor using inter-frame and intra-frame correlation," *Proc. 2005 IEEE Workshop on CCD and Advanced Image Sensors*, pp.238–241, 2005.
- [9] K. Yamamoto, Y. Maeda, Y. Masaki, K. Kagawa, M. Nunoshita, and J. Ohta, "A cmos image sensor for id detection with high-speed readout of multiple region-of-interests," *Proc. 2005 IEEE Workshop on CCD and Advanced Image Sensors*, pp.165–168, 2005.



**Junichi Akita** was born in Nagoya, Japan in 1970. He received B.S., M.S. and Ph.D. degrees in electronics engineering from the University of Tokyo, Japan in 1993, 1995 and 1998 respectively. He joined the Department of Computer and Electrical Engineering, Kanazawa University as a research associate in 1998. He moved to the Department of Media Architecture, Future University-Hakodate as an assistant professor in 2000. He moved to the Department of Information and Systems Engineering, Kanazawa University as an assistant professor in 2004. His main research interest is in analog parallel signal processing VLSI architecture and its applications. He is also interested in electronics systems including VLSI systems in the applications of human-machine interaction and human interface. He is a member of Information Processing Society of Japan and the Institute of Image Information and Television Engineering.

His main research interest is in analog parallel signal processing VLSI architecture and its applications. He is also interested in electronics systems including VLSI systems in the applications of human-machine interaction and human interface. He is a member of Information Processing Society of Japan and the Institute of Image Information and Television Engineering.



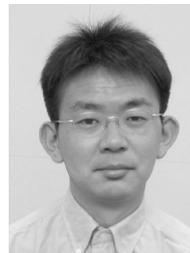
**Hiroaki Takagi** was born in Toyama, Japan in 1982. He received B.S. and M.S. degrees from the Kanazawa University, Japan in 2005 and 2007, respectively. He joined OMRON Corp. in 2007.



**Keisuke Doumae** was born in Ishikawa, Japan in 1984. He received B.S. degree from the Kanazawa University, Japan in 2007. He is currently a master course student in Kanazawa University.



**Akio Kitagawa** was born in Shiga, Japan in 1961. He received B.S. and M.S. degrees from Nagoya Institute of Technology in 1985 and 1987, respectively. From 1989, he worked at the Faculty of Engineering, Kanazawa University as a research associate. From 1995, he has been an associate professor at Kanazawa University. From 1999 to 2001, he was also an assistant professor at VLSI Design and Education Center, University of Tokyo. He is a member of Information Processing Society of Japan and the Institute of Electrical and Electronic Engineers.



**Masashi Toda** was born in Hamamatsu, Japan in 1969. He received B.S. from the University of Tokyo, Japan in 1993, and M.S. and Ph.D. in electro-informatics engineering from Hokkaido University, Japan in 1995 and 1998, respectively. From 1998 to 2001, he was a researcher in IS Labo., SECOM Co., Ltd., Japan. From 2001 to 2005, he was an assistant professor in School of Systems Information Sciences in Future University-Hakodate, Japan. Since 2005, he has been an associate professor in Future University-Hakodate. His main research interest is in sensing architecture. He is also interested in image processing technology, wearable computing, ubiquitous computing, and educational information system. He is a member of Information Processing Society of Japan.



**Takeshi Nagasaki** was born in Hokkaido, Japan in 1969. He received B.E., M.E., and Ph.D. degrees from Hokkaido University in 1992, 1994 and 1998, respectively. From 1997 to 2000, he worked on B.U.G. Inc. Since 2000 he has been with Future University-Hakodate, where he is currently an assistant professor of Media Architecture. His current research interests are computer vision and wearable system. He is member of Information Processing Society of Japan and the Institute of Electrical and

Electronic Engineers.



**Toshio Kawashima** was born in Hokkaido, Japan in 1957. He received B.E., M.E., and Ph.D. degrees from Hokkaido University in 1980, 1982 and 1990, respectively. From 1984 to 2000, he worked at the Department of Information Engineering of Hokkaido University, where he was an associate professor. In 2000 he moved to Future University-Hakodate as a professor of the Department of Media Architecture. He is engaged in research on multi-sensor system for human assistance and digital archives

for historical documents. He is a member of Information Processing Society of Japan, Association for Computing Machinery, and the Institute of Electrical and Electronic Engineers.