A Digital Vision Chip Specialized for High-Speed Target Tracking

Takashi Komuro, Idaku Ishii, Masatoshi Ishikawa, and Atsushi Yoshida

***Abstract*—This paper describes a new vision chip architecture for high-speed target tracking. The processing speed and the number of pixels are improved by hardware implementation of a special algorithm which utilizes a property of high-speed vision and introduction of bit-serial and cumulative summation circuits. As a result, 18 objects in a 128 × 128 image can be tracked in 1 ms. Based on the architecture, a prototype chip has been developed; 64 × 64 pixels are integrated in 7 mm square chip and the power consumption for obtaining the centroid of an object per every 1 ms is 112 mW. Some experiments are performed on the evaluation board which is developed for evaluation under the condition of actual operation. High-speed target tracking including multitarget tracking with collision and separation has successfully been achieved.**

***Index Terms*—Target tracking, vision chip.**

I. Introduction

A VISION chip is an image sensor, each pixel of which has a processing element (PE). Much research on vision chips has been performed with the approach of mimicking biological visual systems. Using a vision chip, high-speed image processing over 1000 frames/s is realized as the cost of scanning and transmitting a large amount of data is unnecessary, which was the bottleneck of the whole system. This frame rate is much higher than maximum response frequency of most of mechanical systems and it is possible to control the dynamics of the system directly. Therefore, the vision chip can be applied to intelligent robots, factory automation, transportation systems, and medical/biological systems. For such industrial purposes, more processing capability is required than that of the biomimicking systems and a general-purpose vision chip with digital processing elements is effective for that purpose. It can process various algorithms by changing the program and can be applied for a wide range of applications. However, its generality makes the area of the circuits large and it is difficult to integrate many pixels on a chip. If the task is fixed, a specialized architecture is more effective in the point of compactness.

In this paper, we propose a new vision chip architecture specialized for high-speed target tracking and describe a prototype chip.

Manuscript received April 18, 2002; revised July 12, 2002. The review of this paper was arranged by Editor J. Hynecek.

T. Komuro and M. Ishikawa are with the Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan (e-mail: kom@k2.t.u-tokyo.ac.jp).

I. Ishii is with the Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan.

A. Yoshida is with Products Development Department, Nippon Precision Circuits Inc., Tokyo 135-8430, Japan.

Digital Object Identifier 10.1109/TED.2002.807255

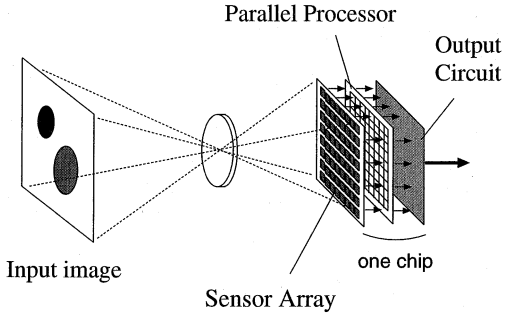


Fig. 1. Model of the digital vision chip.

II. Background

A. Digital Vision Chip

Since the proposal of a silicon retina by Mead *et al.* [1], many kinds of vision chips has been developed. Most of them use analog circuits in the processing element such as resister networks. Analog circuits are generally smaller than digital but have disadvantages in variety of functions, programmability, precision, noise tolerance, speed and stability.

On the other hand, Ishikawa *et al.* proposed the concept and an architecture of a programmable vision chip with general-purpose PEs using digital circuits [2]. In this architecture, each PE has a bit serial arithmetic logic unit (ALU) and 24-b local memory, can communicate directly with four neighboring PEs, and is controlled by programs in an single-instruction multi-data (SIMD)-type parallel processing fashion. They developed a scaled-up model of the vision chip, which consists of 256 discrete photodetectors (PDs) and 64 gate-array LSIs, each of which includes eight PEs. The PDs and the PEs are connected pixel by pixel. Using the system, various demonstration systems were developed [3]–[5]. These systems realize high-speed visual feedback at a sampling period of 1 ms. To integrate all the PDs and PEs into a one-chip VLSI, Komuro *et al.* have designed a simpler architecture [6] and developed a 64 × 64 pixels prototype chip [7]. In this architecture, all the I/Os such as PD input and neighbor input/output are mapped to the local memory. As a result, all the operations are performed by accessing the local memory and the instruction set is simple.

Such a vision chip can be defined as a "digital vision chip." A model of a digital vision chip is shown in Fig. 1. Photosensors and parallel processing elements consisting of digital circuits are integrated on a single chip, thereby realizing high-speed visual processing without the I/O bottleneck.

Other research on digital vision chips has been performed at some laboratories worldwide. In France, a digital vision chip

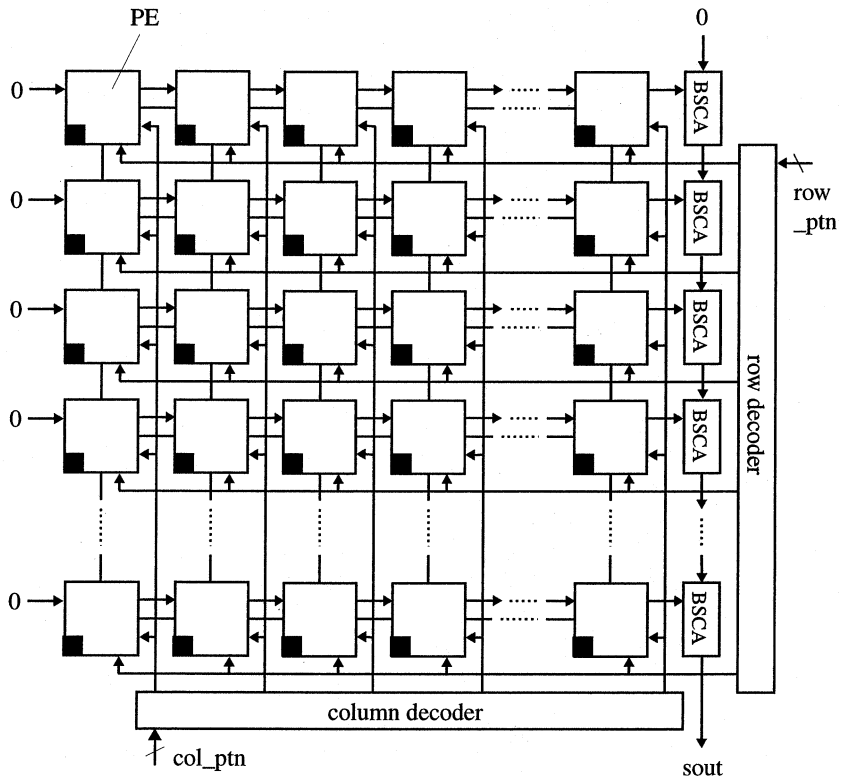


Fig. 2. Structure of the whole chip.

called the programmable artificial retina was proposed, and the researchers successfully developed chips of 65 × 76 pixel [8] resolution. In Sweden, a digital vision chip based on the their near-sensor image processing concept was proposed and a 32 × 32 pixel chip was developed [9].

B. Need for Higher Resolution

General-purpose vision chips like those referenced above introduce processor-like PEs consisting of an ALU and memory. As a result, the size of the PE becomes large, making it difficult to integrate as many pixels on a chip as a normal image sensor.

However, the need for higher resolution grows as research of vision chip application proceeds. Higher resolution sensors are particularly desirable in applications calling for monitoring a wide range of view.

To realize a higher resolution vision chip, one solution is to reduce the memory size or functions of each PE. Such an approach however involves a tradeoff between the number of pixels and PE performance. Owing to this tradeoff, the system may become useless in an actual application. Therefore, the design of a digital vision chip with both practicability and high resolution is highly desirable.

C. Global Feature Extraction

One of the important issues of the vision chip we have to consider is output. Most digital vision chips have a structure of a mesh-connected massively parallel processor and it is possible to process transformation from a two-dimensional (2-D) pattern to a 2-D pattern at high speed. However, the result data are still large and, if we try to output the whole pattern, we again face the I/O bottleneck problem which we solved for the input.

In a real application, the final output needed is not a pattern but some feature values extracted from the pattern. For example, visual feedback control of the robot gets the position of the target from the sensor and gives feedback to the actuator. In this case, pattern information is only an intermediate expression and the output of the vision chip can be scalar information.

To obtain the feature values of the whole image, global operation such as calculating summation of the array is needed. This operation takes a relatively long time if we use only communication with neighboring PEs, which is used in most digital vision chips. Alternatively, we can use a global bus connected to every PE in an array or in a row but we still need to scan out the data which also consumes much time. Special network topology such as the hypercube, pyramid, or tree can be effective in increasing speed, but has the drawbacks of increased area requirements and lack of homogeniety, which is an important factor for modularity, expandability and reusability in designing a vision chip. It is desirable to realize high speed global operation using only a 2-D mesh network.

III. Chip Architecture

In the proposed architecture, the purpose of the chip is limited to target tracking, which still has a wide application in various fields. As a result, it cost only the minimum size of circuits for the purpose. There have been some vision chips for target tracking [10], [11] but they were all analog ones. We have designed the PE in digital circuits for higher performance.

The block diagram of the whole chip is shown in Fig. 2. Homogeneous-structured PE array is arranged in parallel. Bit-serial cumulative adders (BSCAs) shown in Fig. 3 are attached to each of the left-end PEs and are connected serially in a row. The

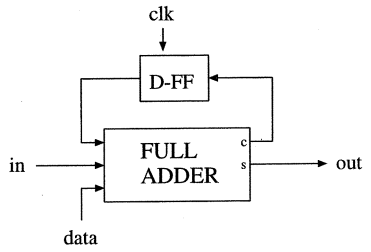


Fig. 3. BSCA.

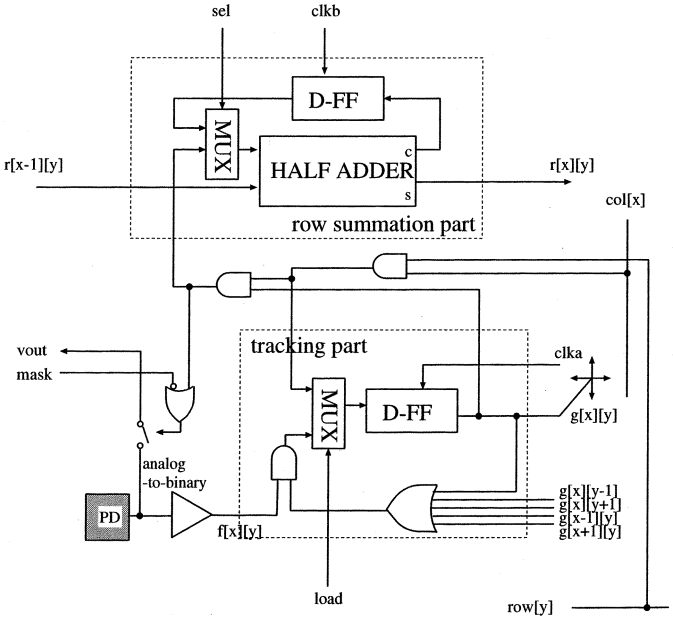


Fig. 4. Structure of the PE.

BSCA array calculates summation of the whole array from row summations of each row. Row-common buses are arranged in each row and column-common buses in each column. The signals for the buses are generated by a decoder from instructions.

The block diagram of the PE is shown in Fig. 4. The PE is roughly divided into a tracking part and a row summation part. The tracking part consists of circuits in which the self-windowing algorithm described later is hardware-implemented. The output of the tracking part is connected to the tracking part of four-neighbor PEs and to the row summation part via the mask part which consists of an AND gate with inputs of the row-common bus and the column-common bus.

Like the column summation, the row summation part consists of a BSCA, but the full-adder is replaced by a half-adder to reduce the number of transistors. By connecting this module serially in a column, the summations of each column can be calculated.

Each PE has a PD. The PD output is connected to a binarization circuit and its output is connected to the tracking part. The PD output is also connected to the monitor output part.

IV. Basic Functions

The functions of the vision chip are described below. They are not synchronized with each other and each works at a different rate respectively. Therefore, we can always configure the proper rates to optimize the performance for the target application.

A. PD Input

The image signal captured by a photo sensor is transformed to a voltage signal and then is binarized and transmitted to the PE. The voltage signal is also transferred to the monitor output part. The data is updated by sending a reset signal to the PD.

B. Tracking

Target tracking using the self-windowing algorithm [12] is handled by hardware. Self-windowing is an algorithm to search only inside the object and its one-pixel neighbors. The algorithm uses the property of high speed vision that the change between the images of two successive frames is very small. Details of this algorithm are shown below.

We define the input image at time k as $f_k(x, y)$, the tracking image as $g_k(x, y)$, and the window image as $W_k(x, y)$:

1) Initial determination of the object

$$g_0(x, y) = \text{(the image of the object at the time 0)}.$$

2) Making the window

$$W_{k+1}(x, y) = g_k(x, y) \cup g_k(x+1, y) \cup g_k(x-1, y) \cup g_k(x, y+1) \cup g_k(x, y-1).$$

3) Extracting the target

$$g_{k+1}(x, y) = W_{k+1}(x, y) \cap f_{k+1}(x, y).$$

Repeat 2) and 3) with incrementing k

We show the illustration of the self-windowing algorithm in Fig. 5.

By implementing the self-windowing algorithm into hardware, the circuits are reduced and accelerated.

To run the self-windowing algorithm, an initial image, which is usually a point image, needs to be loaded. With selecting the pattern of row and column buses as mentioned later and setting the signal *load* to 1, the initial image is loaded to the register at the moment at the positive edge of the clock *clka.* Then, with setting *load* to 0, the tracking image in the register is updated according to the self-windowing algorithm at the every positive edge of *clka*.

C. Row/Column Selection

Via common buses in rows and in columns, coordinate-dependent values are transmitted to the PEs. These values are used not only as a mask (AND operation) to calculate moments but also as an initial window and for pixel selection of monitoring. The patterns given to the buses are generated by a decoder. Sample patterns using 8×8 pixels are shown in Fig. 6. Bit patterns of coordinate values are shown in (a) and are used for calculation of moments. Patterns shown in (b) are used for pixel selection of the monitor output and for giving a point as an initial window. Patterns shown in (c) are used for area division.

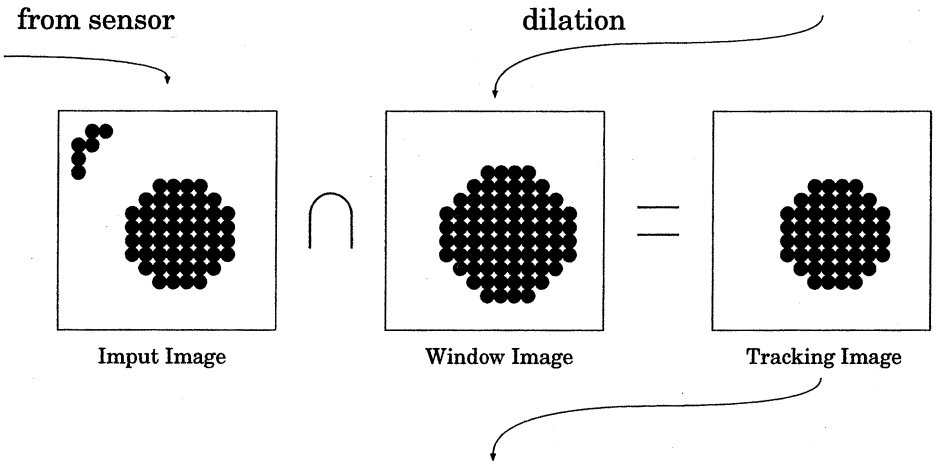


Fig. 5. Self-windowing algorithm.

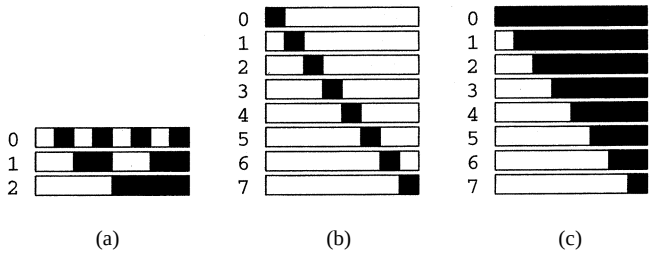


Fig. 6. Sample patterns of row/column selection. (a) Coordinate. (b) Pixel Selection. (c) Area division.

D. Summation Calculation

As global information, summation of the tracking image is calculated and output in a bit-serial manner. The process is described below. The clock cycle needed for output is $\log_2$ [maximum of output] (the decimal is rounded off).

1) First, to reset the register in the column summation BSCA of every row, $2n$ pulses are given to the clock *clkb* with the status of the signal *sel* being 0. After the first calculation, resetting is not needed as the previous calculation clears the registers.
2) With setting *sel* to 1, at the positive edge of *clkb*, the tracking image is stored to the register in the row summation part of every PE. The row summation part calculates row summation and the LSB is output from the right end to column summation BSCA. Given the row summation, the column summation BSCAs calculate the 2D array summation and the LSB is output to the line *sout.*
3) Next, with setting *sel* to 0, at the positive edge of *clkb*, the next bit of row summation is calculated and is output from the right end to the column summation BSCA; the next bit of array summation is calculated and is output to *sout.* This is repeated until all bits have been output.

A sample calculation is shown in Fig. 7.

With masking the tracking image using the function of row/column selection, any $\Sigma g(x, y)F(x)G(y)$ type of linear operation such as moments can be calculated.

E. Monitor Output

The input image can be monitored using a grayscale analog signal. Moreover, the function of cropping the tracking image can be provided.

V. Algorithms

In this section, the sample algorithms for the proposed vision chip are described. Here we define $N \times N(N = 2^n)$ as the number of pixels and $g(x, y)$ as the tracking image, and 1 cycle means the 1 clock cycle of *clkb.*

A. Centroid Extraction

Extracting the centroid of an object is an indispensable task for target tracking. The centroid $(\overline{x}, \overline{y})$ is calculated as below from the 0th moment m_{00} and the 1st moments m_{10}, m_{01} of a tracking image. The area information is also obtained from m_{00} as follows:

$$\overline{x} = m_{10}/m_{00} \tag{1}$$

$$\overline{y} = m_{01}/m_{00}. \tag{2}$$

Using the summation calculation function, the 0th moment of a tracking image

$$m_{00} = \sum_{x=1}^{N}\sum_{y=1}^{N} g(x, y) \tag{3}$$

can be calculated in $(2n + 1)$ cycles. The 1st moments can be calculated as shown below.

Here we define $x_n x_{n-1} \cdots x_1$ as x's binary expression:

$$\begin{aligned} x &= x_n 2^{n-1} + x_{n-1}2^{n-2} + \cdots + x_1 2^0 \\ &= \sum_{k=1}^{n} x_k 2^{k-1} \end{aligned} \tag{4}$$

$$\begin{aligned} m_{10} &= \sum_{x=1}^{N}\sum_{y=1}^{N} x g(x, y) \\ &= \sum_{x=1}^{N}\sum_{y=1}^{N}\sum_{k=1}^{n} x_k 2^{k-1} g(x, y) \\ &= \sum_{k=1}^{n} 2^{k-1} \sum_{x=1}^{N}\sum_{y=1}^{N} x_k g(x, y) \\ &\equiv \sum_{k=1}^{n} 2^{k-1} s_k. \end{aligned} \tag{5}$$

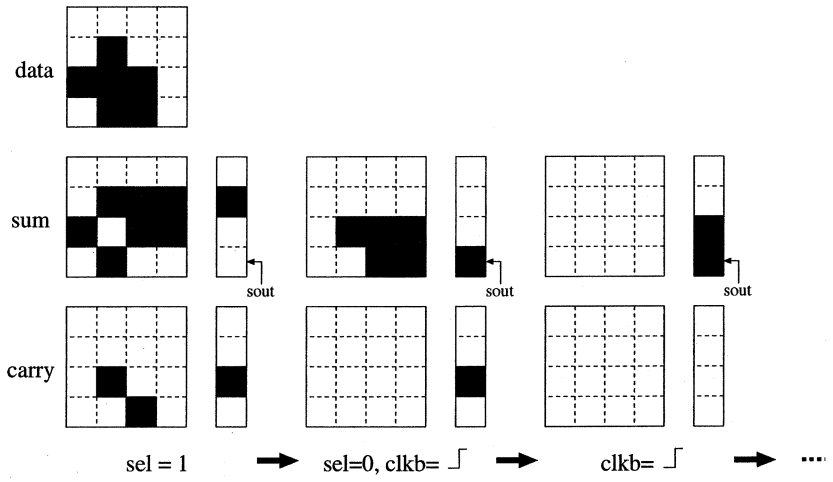


Fig. 7. Sample calculation (4 × 4).

TABLE I
CYCLES AND PROCESSING TIME FOR MOMENTS EXTRACTION

Algorithm	cycles(time)
0th moment (area) m_{00} (N=64)	13(1.6μs)
0th moment (area) m_{00} (N=128)	15(3.8μs)
0th moment (area) m_{00} (N=256)	17(8.7μs)
1st moment m_{10}, m_{01} (N=64)	72(9.2μs)
1st moment m_{10}, m_{01} (N=128)	98(25μs)
1st moment m_{10}, m_{10} (N=256)	128(66μs)
Centroid $(\bar{x}, \bar{y})$ (N=64)	157(20μs)
Centroid $(\bar{x}, \bar{y})$ (N=128)	211(54μs)
Centroid $(\bar{x}, \bar{y})$ (N=256)	273(140μs)

Using bit patterns of x by the column selection bus as masks, s_k can be calculated in $2n$ cycles. Therefore, m_{10} can be calculated by repeating the calculation of s_k from $k = 1$ to n and adding them with bit-shifting outside the chip, the total cycle number is $2n^2$. The value m_{01} can be calculated similarly.

Therefore, the centroid extraction costs $(4n^2 + 2n + 1)$ cycles using this vision chip. Cycles and processing time for moments extraction is shown in Table I. The clock period of *clkb* is supposed to be $2.0N$ ns. The processing outside the chip is not counted. From this table it can be seen that the centroid of an object is calculated at the order of μs and this vision chip has adequate performance for its target applications.

Moreover higher moments can be calculated by applying proper bit masks. These moments are applicable to, for example, model matching.

In this design, we use a half-adder in the row summation part to save area but, by replacing it by a full-adder, it is possible to calculate moments more efficiently.

B. Multitarget Tracking

To successfully track more than one object, it is usually necessary to store all of the tracking images in the chip's local memory. The vision chip described in this paper dedicates only one bit register to local memory and all data needs to be processed within this limit.

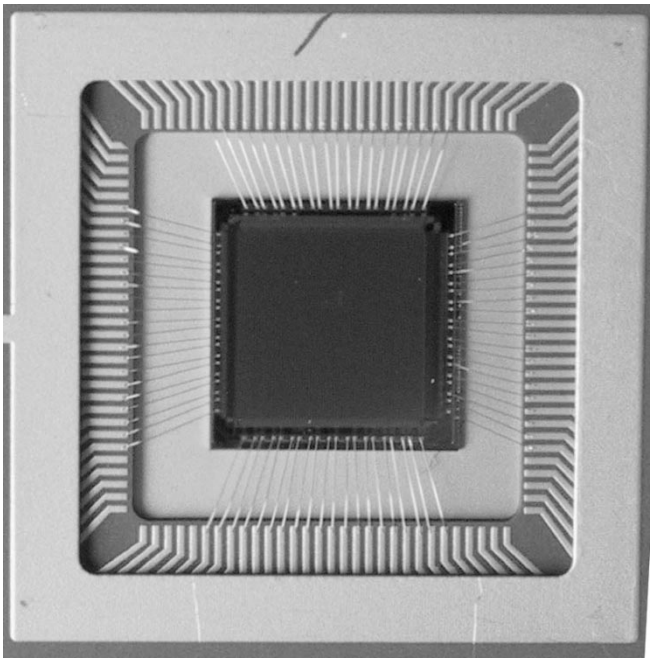


Fig. 8. Photo of the second prototype chip.

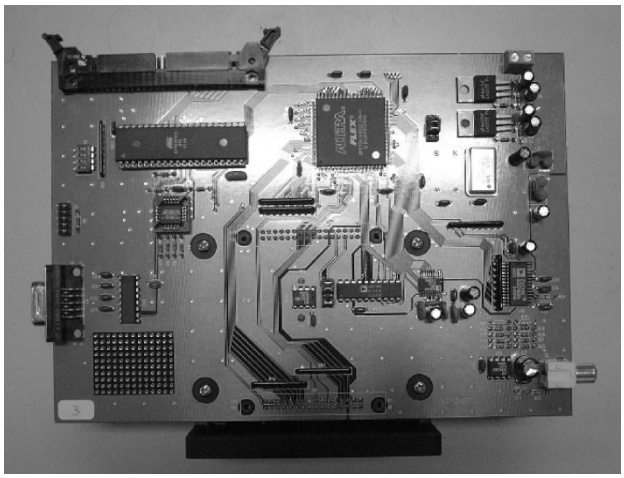
However, by storing only the centroids of objects outside the chip, multitarget tracking is realized without the limit of local memory. In this case, to restore a tracking image from the stored centroid, a point image of the centroid is used as an initial image and the self-windowing algorithm is run in several loops, thereby filling the inside of the object.

In this vision chip the filling can be realized at high speed by setting the rate of *clka* much higher than the frame rate and the time needed for restoring from centroid to image can be ignored. Therefore tracking of m objects is realized in m times processing time. Using the number of Table I, in the case of $N = 128$ 18 objects can be tracked in a 1-ms frame rate.

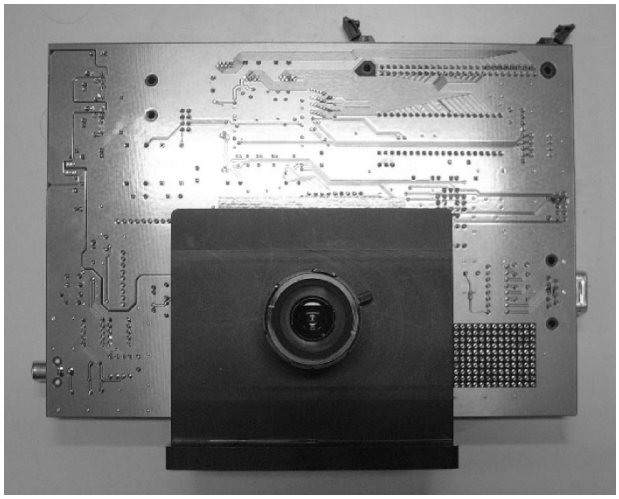
When tracking more than one object, the problem arises of collectly handling collision of objects and separation of an object. Collision and separation can be recognized by examining

TABLE II
SPECIFICATIONS OF THE SECOND PROTOTYPE CHIP

Process	0.5μm CMOS TLM
Chip area	7mm×7mm
Number of pixels	64× 64
Pixel area	80μm×80μm
Photodiode area	30μm×40μm
Max Clock Freq.	$clka > 10$MHz
	$clkb = 10$MHz
Power Consumption	74mW(1 centroid/1ms)



(a)



(b)

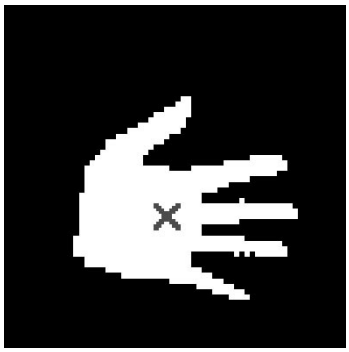
Fig. 9. Evaluation board. (a) Front. (b) Back.

the difference between the area of the whole image and the total area summation of all objects as below. S_{whole} is the area of the whole image and S_i is the area of each object.

$$S_{\text{whole}} < \sum_{i=1}^{m} S_i \quad \text{(collision)} \tag{6}$$

$$S_{\text{whole}} > \sum_{i=1}^{m} S_i \quad \text{(separation)}. \tag{7}$$

After recognizing separation of an object, the centroid of the new object can be calculated as below. $(\overline{x}_{\text{whole}}, \overline{y}_{\text{whole}})$ is the centroid of the whole image and $(\overline{x}_i, \overline{y}_i)$ is the centroid of each object.



(a)

(b)

(c)

Fig. 10. Sample output. (a) Hand. (b) Face. (c) Face analog.

$$\overline{x}_{m+1} = \frac{\overline{x}_{\text{whole}} S_{\text{whole}} - \sum_{i=1}^{m} \overline{x}_i S_i}{S_{\text{whole}} - \sum_{i=1}^{m} S_i} \tag{8}$$

$$\overline{y}_{m+1} = \frac{\overline{y}_{whole} S_{\text{whole}} - \sum_{i=1}^{m} \overline{y}_i S_i}{S_{\text{whole}} - \sum_{i=1}^{m} S_i}. \tag{9}$$

VI. VLSI IMPLEMENTATION

Based on the architecture just described, schematic circuits have been designed and the functions have been verified by simulation. The number of transistors per PE is 84, which is about a fifth of that of the general purpose vision chip described in the

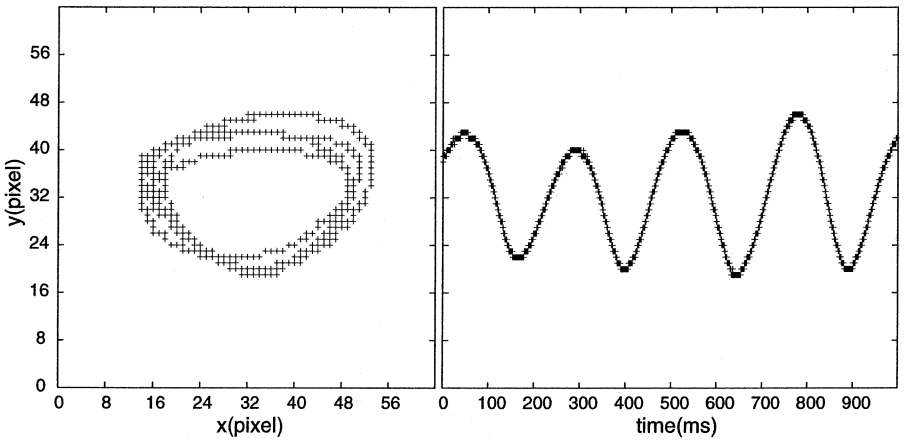


Fig. 11. Tracking result.

Section II [7]. The number of transistors can be reduced more by introducing dynamic circuits.

The first prototype chip using 0.6-μm CMOS process has been developed, with 64 $\times$ 64 pixels integrated on a 9.75 mm $\times$ 9.80 mm chip [13]. Then the second prototype chip using a 0.5-μm CMOS process has been developed. A photo of the chip is shown in Fig. 8. Then, 64 $\times$ 64 pixels are integrated on a 7 mm $\times$ 7 mm chip. Using a recent process it is expected that more than 128 $\times$ 128 pixels can be integrated into one chip. In this prototype, a latch to hold the binarized data of PD input, circuits for black/white inversion, and a pipeline register before the column summation BSCA to double the clock frequency are added.

The speed of the circuits depends on the clock frequency of *clka* and *clkb*. The maximum frequency of *clka* and *clkb* tested is 10 MHz. It is the upper limit for *clkb* but the chip will work if much higher frequency is given to *clka*. Moreover the integration time of the PD decides the frame rate. The chip can get an image with the frame rate of 1ms under the usual environment the without a strong lighting.

The power consumption of the chip is 112 mW when the chip calculates the centroid of an object at 1ms frame rate with the 4-MHz frequency of *clkb*. It is small enough for normal applications. In the future, when the number of pixels becomes larger and the clock frequency becomes higher, power consumption will become a serious problem. The tracking part does not consume so much power per clock as the input values of the logic gates do not turn often. Additionally, only a few clocks per frame need to be given if the chip tracks only one object. Even if the chip tracks more than one object, the number of clocks needed to be given per frame is as few as the product of the maximum radius of the objects and the number of the objects. On the other hand, the row summation part consumes much power per clock as the input values of the logic gates turn many times (N times at the worst case). As the number of pixels increases, the total number of gates increases on the order of N^2 and the number of clocks needed to be given for obtaining the moments of an object increases at the order of $(\log N)^2$. In total, the number of pixels affects the power consumption at the order of $N^3(\log N)^2$. Moreover, to track more than one objects, the number of clocks needed to be given increases in proportion to the number of objects. To reduce the power consumption, some mechanism is needed to suspend the gates until the proper signal comes.

The specifications of the chip are given in Table II.

VII. Evaluation Board

To evaluate the chip under the condition of actual operation, the evaluation board has been developed. The photos of the evaluation board is shown in Fig. 9. The board consists of the vision chip, a one chip micro controller, an FPGA, an analog-to-digital converter (ADC), and a digital-to-analog converter (DAC) on an A5 size PCB with optical lens module. The lens taken has a 1/2-in image format, 6.0-mm focal length, and F1.4–F16 aperture. The system clock frequency of the board is 40 MHz. The board uses a +5-V dc power supply and consumes 500 mA of current. The board can be controlled by a PC via serial port, or works alone. The monitor image is output to the video port as NTSC signals with the centroid of the tracking image drawn with a marker. The configuration such as the frame rate, the threshold voltage of binarization circuits can be set by DIP switches, or via serial port. A GUI environment on Microsoft Windows with a user programming interface has been developed for easy control of the board.

Using the evaluation board, the experiment of capturing and processing images has been performed. The output examples are shown in Fig. 10. The left image is the tracking image and the centroid of a human hand irradiated by a small 40-W electric bulb. The distance between the light and hand is about 50 cm and the distance between the camera and hand is also about 50 cm. The frame rate is 1.00 ms and the PD integration time is 0.87 ms. The middle image is the tracking image and the centroid of a human face. The condition is the same as that in the left image. The right image is the monitor output image of the same human face as the middle image.

The tracking experiment has been performed to prove the ability of high-speed target tracking. A white styrene foam ball moved rapidly and irregularly by hand is the target. The series of target centroids is obtained at the frame rate of 1.00 ms. The result is shown in Fig. 11. It is confirmed that the chip can track a rapidly moving object.

Multitarget tracking experiment has also been performed. Two black targets are drawn on the transparent sheet against the background of red LED lighting. They rotate around the center of the sheet. The algorithm was implemented to the micro controller using the user programming interface of the evaluation board. The frame rate is about 2.5 ms. The result is

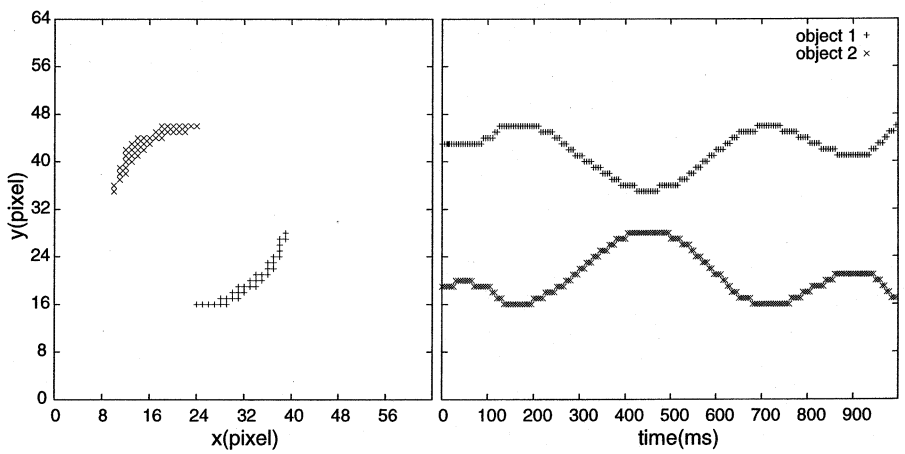
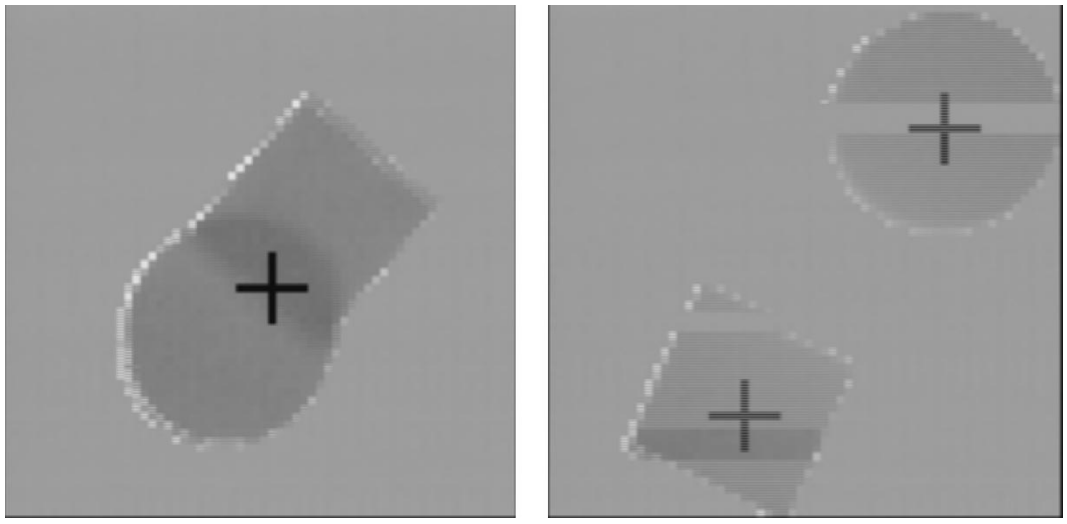


Fig. 12. Multitarget tracking result.



(a) (b)

Fig. 13. Multitarget tracking with (a) collision and (b) separation.

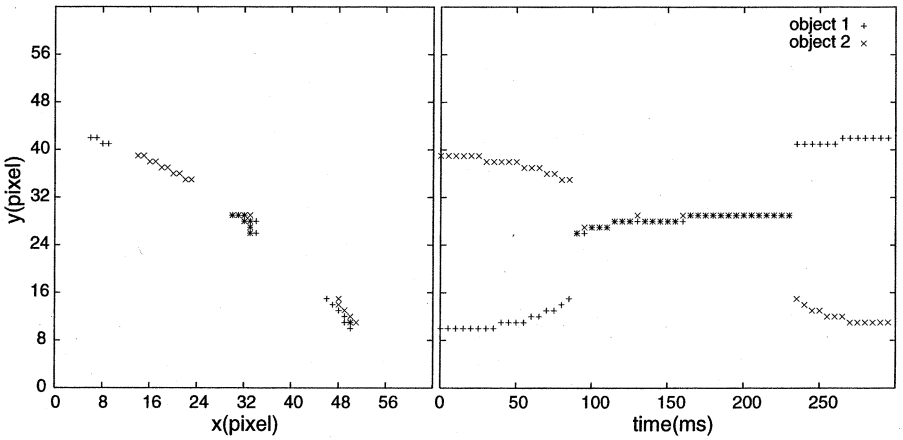


Fig. 14. Collision and separation result.

shown in Fig. 12. The result shows that the chip achieved the tracking of two objects successfully.

Multitarget tracking with collision and separation has also been achieved. Fig. 13 shows the video output and Fig. 14 shows the graph.

VIII. Conclusion

In this paper, a new vision chip promising higher resolution even with a digital approach has been proposed, and a prototype chip and an evaluation board have been shown. The pixel area of

the chip is 80 μm $\times$ 80 μm using a 0.5-μm CMOS process, and the power consumption is 112 mW. The experimental results have shown that the chip can track successfully one or more objects at high speed.

The high-resolution digital vision chip as described in this paper are expected to find widespread use in fields where precise control is needed such as for micro machining, semiconductor inspection and biological/medical systems. The chip also offer distinct advantages to the fields of robotics, factory automation, and so on with the capability of more stable control and a wider field of view, which removes the requirement of costly mechanical tracking.

Moreover, by developing algorithms, this paper has shown that even simple circuits can process high level tasks including multitarget tracking. This will broaden the application range of vision chips.

We will proceed to develop prototype chips aiming toward a more practical chip with a larger number of pixels.

REFERENCES

[1] C. Mead and M. Mahowald, "A silicon model of early visual processing," *Neural Network*, vol. 1, pp. 91–97, 1988.

[2] M. Ishikawa, A. Morita, and N. Takayanagi, "High speed vision system using massively parallel processing," in *Proc. 1992 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS'92)*, 1992, pp. 373–377.

[3] Y. Nakabo and M. Ishikawa, "Visual impedance using 1 ms visual feedback system," in *Proc. IEEE Int. Conf. Robotics and Automation*, 1998, pp. 2333–2338.

[4] T. Owaki, Y. Nakabo, A. Namiki, I. Ishii, and M. Ishikawa, "Real-time system for virtually touching objects in the real world using modality transformation from images to haptic information," *Syst. Comput. Japan*, vol. 30, no. 9, pp. 17–24, 1999.

[5] A. Namiki and M. Ishikawa, "Optimal grasping using visual and tactile feedback," in *Proc. IEEE Int. Conf. Multisensor Fusion and Integration for Intelligent Systems*, 1996, pp. 589–596.

[6] T. Komuro, I. Ishii, and M. Ishikawa, "Vision chip architecture using general-purpose processing elements for 1 ms vision system," in *Proc. 4th IEEE Int. Workshop on Computer Architecture for Machine Perception (CAMP'97)*, 1997, pp. 276–279.

[7] M. Ishikawa and T. Komuro, "Digital vision chips and high-speed vision systems (invited)," in *Dig. Tech. Papers 2001 Symp. VLSI Circuits*, 2001, pp. 1–4.

[8] T. Bernard, B. Zavidovique, and F. Devos, "A programmable artificial retina," *IEEE J. Solid State Circuits*, vol. 28, pp. 789–797, July 1993.

[9] J. Eklund, C. Svensson, and A. Åström, "VLSI implementation of a focal plane image processor—A realization of the near-sensor image processing concept," *IEEE Trans. VLSI Syst.*, vol. 4, pp. 322–335, Sept. 1996.

[10] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *IEEE J. Solid State Circuits*, vol. 33, pp. 1199–1207, Aug. 1998.

[11] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang, "A Foveated silicon retina for two-dimensional tracking," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 504–517, June 2000.

[12] I. Ishii, Y. Nakabo, and M. Ishikawa, "Target tracking algorithm for 1 ms visual feedback system using massively parallel processing vision," in *Proc. IEEE Int. Conf. on Robotics and Automation*, 1996, pp. 2309–2314.

[13] T. Komuro, I. Ishii, M. Ishikawa, and A. Yoshida, "High speed target tracking vision chip," in *Proc. 5th IEEE Int. Workshop on Computer Architecture for Machine Perception (CAMP2000)*, 2000, pp. 49–56.

Takashi Komuro received the B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1996, 1998, and 2001, respectively.

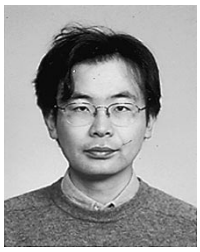
At present he is a Research Associate of information physics and computing at the University of Tokyo. His current research interests include massively parallel processing vision chip architecture and VLSI implementation.

Idaku Ishii received the B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1994, 1996, and 2000, respectively.

At present he is an Associate Professor of electrical and electronic engineering at Tokyo University of Agriculture and Technology. His current research interests include massively parallel processing vision chip, sensory-motor fusion system, multimodal human interface, and sensory information parallel processing theory.

Masatoshi Ishikawa received the B.E., M.E., and Dr.Eng. degrees in mathematical engineering and information physics from University of Tokyo, Tokyo, Japan, in 1977, 1979, and 1988, respectively.

From 1979 to 1989, he was a Senior Researcher at Industrial Products Research Institute, Tsukuba, Japan. At present, he is a Professor of information physics and computing at the University of Tokyo. His current research interests include circuit model of neural networks, optical computing, parallel processing, smart sensors, tactile sensors, and sensor data fusion.



Atsushi Yoshida received the B.E. degree in physical electronics from Tokyo Institute of Technology, Tokyo, Japan, in 1986.

From 1986 to 1998 he was a researcher at Seikosha Co., Ltd. (now Seiko Precision Inc.). At present, he is a Researcher at Nippon Precision Circuits Inc., Tokyo. His current research interests include development of sensor-related ICs.