MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

A High-Speed and Compact Vision System Suitable for Wearable Man-machine Interfaces

Takashi Komuro, Björn Werkmann,
Takashi Komai, Masatoshi Ishikawa
Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
Takashi_Komuro@ipc.i.u-tokyo.ac.jp

Shingo Kagami
Graduate School of Information Sciences, Tohoku University

Abstract

We developed a compact vision system for wearable interface applications using a vision chip. The vision chip is capable of real-time vision at higher frame rates than the video frame rate with a single chip. Using this system, vision-based wearable man-machine interfaces, such as a portable eye tracker and a six-degrees-of-freedom input device, can be realized at higher frame rates than ever before.

1. Introduction

High frame rate vision over the video frame rate enhances performance of machine vision. For example, the cases have been reported that a robot hand manipulates an object quickly and skillfully[1], and that a surgery robot compensates the organ motion[2], both using visual information obtained at 1000 frames/s in real time for visual feedback control. Several systems that realize such high-speed vision have also been developed[3-6].

This time we consider use of such high-speed vision for motion measurement of a human to realize a man-machine interface. Much research of noncontact man-machine interfaces using a camera has been studied. However, the frame rate was limited to the video frame rate and there was room to improve the operability. There are commercial mice that can obtain information at a sampling rate over 5000Hz, and it is said that the rate determines the following capability to fast motion. Using high-speed vision, the operability of man-machine interfaces is expected to improve.

However, in face of application to man-machine interfaces, conventional high-speed vision systems are large and expensive. Especially in case of building wearable interfaces, downsizing of the system is necessary. To achieve both high-speed image capture and high-speed image processing with a compact system, a vision chip in which an image sensor and parallel image processing circuit are integrated is effective. A vision chip can outputs necessary information extracted from images instead of the images themselves as it performs image processing in the chip. Therefore, neither high-speed data transfer nor powerful external processor is needed and it is possible to construct a small and lightweight system in its entirety.

In this paper, we describe our new vision system which was developed for wearable interfaces. We also show the results of basic experiments for evaluating usefulness of the system when it is applied to wearable interfaces.

2. Developed Vision System

A photograph of the developed vision system is shown in Fig. 1. A vision chip and a Cypress 8051-compatible microcontroller AN2131SC are equipped on a circuit board. The circuit board size is 40 mm x 26 mm, which is small enough to be used in wearable applications. A small lens is mounted on the vision chip. The output from the vision chip is transmitted to an external device via USB. The microcontroller does not perform image processing and is used mainly for control of the vision chip and communication with external devices.

Figure 1. Developed vision system.

The architecture of the vision chip is the same as the one we reported before[7], but we developed a new chip which is reduced in size for wearable applications. The size reduction is achieved mainly by use of fine process, decreasing the number of pixels, and decreasing the number of pins by parallel-serial conversion. Using 0.35µm CMOS process, the number of pixels is 48x32, and the imaging area is 1.568mm x 2.352mm (1/6 inch format)

The architecture of the chip is shown in Fig. 2. Processing circuits are provided at each pixel of the image sensor, and the chip can thus perform image processing directly in the imaging plane.

The image captured by the sensor is binarized by thresholding, then the target region is extracted by "tracking", after which the area and centroid of the target image are calculated and output. Using the centroid as position, it is possible to achieve subpixel spatial resolution.

For tracking, morphology operation is implemented in hardware. By filling the region beginning at the centroid position of the previous frame, the chip tracks the same target, continuously separated from the background and other targets. In case there are several targets, by applying the above algorithm in turn, they can be tracked individually.

The area and centroid are calculated from the 0th and 1st moments of the target image. The 0th moment (m_{00}), or summation, is calculated in the summation circuit implemented in hardware. Each of the 1st moments (m_{10} , m_{01}) is obtained by masking the target image with each bit pattern of x or y coordinate patterns, taking a summation of the masked target image, and summing the summations with bit-shift.

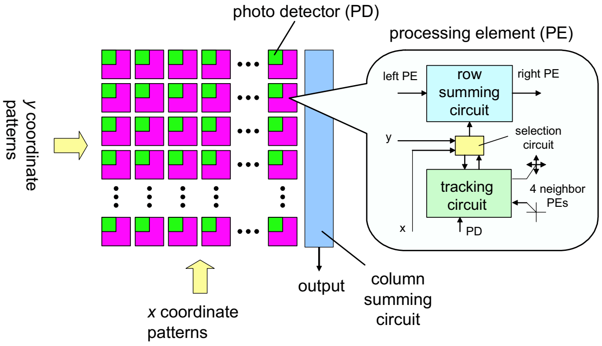


Figure 2. Architecture of the vision chip.

The processing performance of the developed system compared to a general purpose processor is shown in Table 1. By implementing specialized hardware, the system achieves equivalent processing speed to a high-end processor that is installed in a desktop PC with 3 digits lower power consumption.

Table 1. Comparison of processing performance.

	Pentium4 32bit, 2.5GHz	Developed System 1bit, 4MHz
Processing time		
Tracking (48 × 32 pixels)	80µs	0.25µs
Summation (48 × 32 pixels)	6µs	3µs
Power Consumption	115W	15mW

3. Application Examples

To examine the usefulness of the developed system when it is applied to wearable interfaces, we constructed some interface systems and carried out basic experiments.

3.1. Portable Eye Tracker

One such interface system is a portable eye tracker.

This system detects the gazing direction and blinking of a human eye by tracking the eye movement. A human eye movement involves rapid jumping motion called saccades and normal cameras cannot capture the motion. The system that can measure eye movement at the high sampling rate of 500Hz is commercially available[8]. However, the whole high frame rate images captured are transmitted to the PC via a cable and it may be difficult to apply the system for stand-alone use, or to remove the cable by wireless transmission. If saccades can be measured using a simple and low-cost system, applications to health monitoring and saccade-controlled displays[9][10] are expected.

A photograph of the system is shown in Fig. 3. For simplicity, the vision system is directly attached to a pair of safety glasses. To illuminate the user's eye, two infrared LEDs are used. A snapshot of the screen that displays the captured eye image and eye states is shown in Fig. 4.

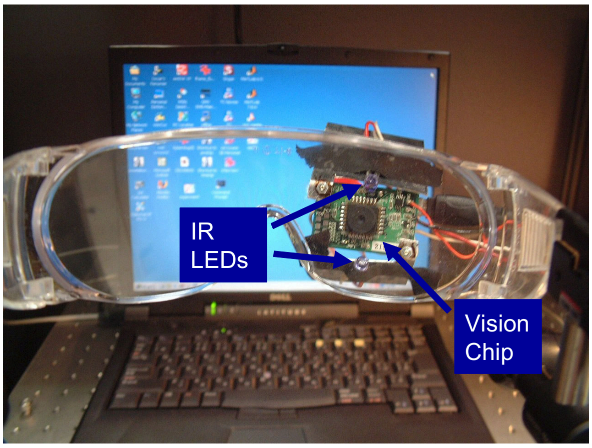
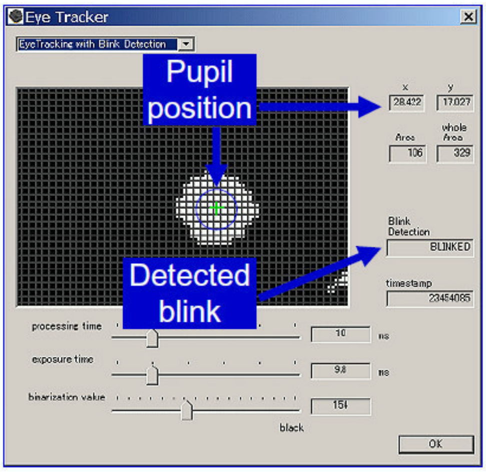


Figure 3. Eye tracking experiment.



EyeTracking with Blink Detection
x 28.422
y 17.027
Area 106
whole Area 329
Blink Detection BLINKED
timestamp 23454085
processing time 10 ms
exposure time 9.8 ms
binarization value 154
black

Figure 4. Output example of the eye tracker.

The region of the pupil is detected by image binarization, and its centroid is obtained. Taking advantage of the high frame rate of the vision chip, the system is fast enough to respond to rapid eye movements such as saccades and blinks, which cannot easily be detected using a normal camera. Fig. 5 shows the measurement result of eye movement. The frame rate was 100 frames per second. Even with the limited number pixels in the current vision chip, the positional accuracy was sufficient to use it as an interface.

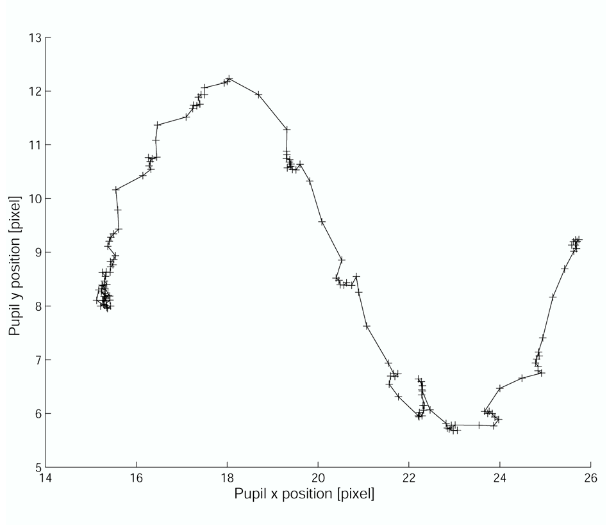


Figure 5. Measurement result of eye movement.

We also created a demonstration program to indicate the eye gazing point on the screen. The correspondence between pupil position and point indicated on the screen is obtained by calibrating the system initially and interpolating by trapezium approximation. The result of tracking a rectangle on the screen is shown in Fig. 6. Though there are some errors due to the coarse resolution of the vision chip, we demonstrated sufficient accuracy for some applications.

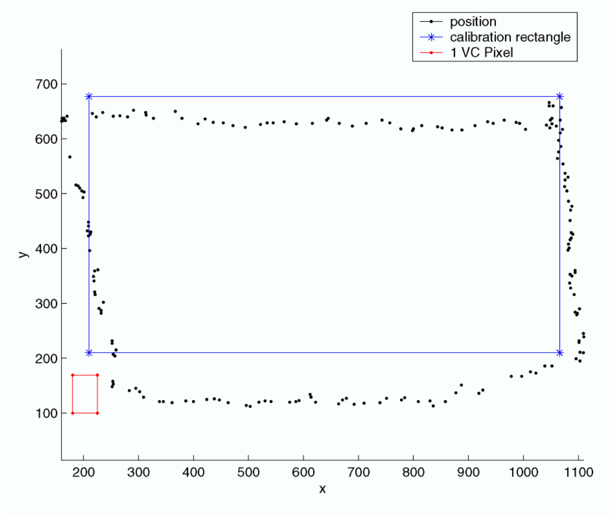


Figure 6. Result of tracing rectangle on screen.

3.2. 6DOF Input Device

Another application is to use the camera itself as an interface and its pose serves as an input. Such an equipment has already been proposed and the prototype has been developed[11]. However, it used a normal CCD camera and the sampling rate was limited to 30Hz. Our system can remove the limitation.

It is well known that the 3D position and orientation of the camera can be obtained from the movement of feature points acquired by the camera. We used markers as imaging targets and calculated the 3D position of the camera from their centroids.

Because our compact vision chip has a relatively low resolution of 48 x 32 pixels, large quantization errors will occur when using the chip as is. Therefore, we introduced a state model assuming motion of uniform translational velocity and uniform angular velocity, and a measurement model based on perspective projection; this allowed us to realize highly accurate measurement by estimating the state using an Extended Kalman Filter (EKF) and to suppress the effects of the quantization errors using dense information across the time dimension. A photograph of the experiment is shown in Fig. 7 and the obtained trajectory is shown in Fig. 8. The frame rate in the measurement was 333 frames per second. You can see in the figure that the camera position and orientation was stably obtained.

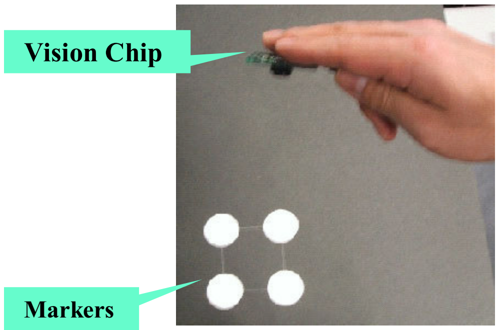


Figure 7. 6DOF input experiment.

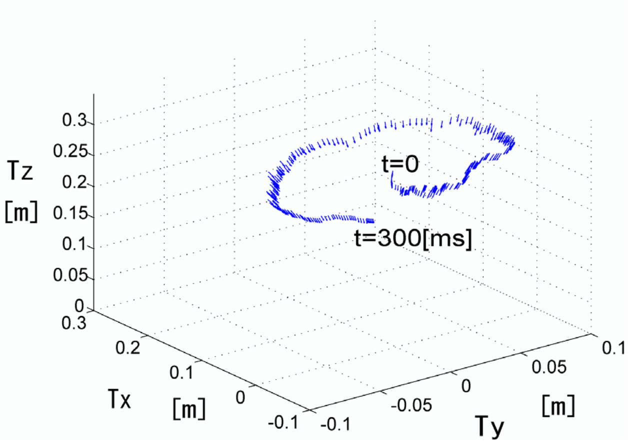


Figure 8. Result of pose detection.

4. Discussion

As described above, using a vision chip high frame rate visual information can be handled with a small and lightweight system.

However, since a vision chip has processing circuit in a pixel, the pixel area becomes large, which makes it difficult to increase the number of pixels, and fill factor, which is the area rate of the photo-sensing area in a pixel, becomes low, which results in low sensitivity.

The former problem would be solved by placing the sensor part and the processor part separately in a chip, between which images are transferred. However, to keep the performance of image processing, a processing circuit with similar size would be required and additional image transfer might increase the area and power consumption in total.

For the latter problem, use of microlens array would be effective. Using microlens array, the fill factor can be almost 100%, which is independent of the size of processing circuit. Therefore, the area efficiency is better than that of the chip having the sensor part and the processor part separately as described above.

Currently the upper limit of the frame rate of the vision chip is determined mainly by its sensitivity. If we get higher sensitivity by introducing microlens array, measurement at higher frame rates than we achieved in the experiments this time will be possible. Then, the structure of a vision chip will get more advantageous.

5. Conclusion

We have described a compact, high-speed vision system developed for wearable interfaces. We also examined some applications using the system.

This system has only simple functions compared to regular image sensors and processing systems, but it has the advantages of speed and compactness. Therefore, it is important to choose and configure its applications based on different ideas from conventional image processing systems.

For the future, we plan to advance our research by both constructing more practical applications and developing devices and systems having higher resolution and sensitivity.

References

[1] N. Furukawa, A. Namiki, T. Senoo, M. Ishikawa. Dynamic Regrasping Using a High-speed Multifingered Hand and a High-speed Vision System, *Proc. IEEE International Conference on Robotics and Automation*, pp.181-187, 2006

[2] Y. Nakamura, K. Kishi, H. Kawakami: Heartbeat Synchronization for Robotic Cardiac Surgery, *Proc. IEEE International Conference on Robotics and Automation*, pp.2014-2019, 2001

[3] S. Kagami, T. Komuro, M. Ishikawa: A High-Speed Vision System with In-Pixel Programmable ADCs and PEs for Real-Time Visual Sensing, *Proc. IEEE International Workshop on Advanced Motion Control*, pp.439-443, 2004.

[4] Y. Nakabo, M. Ishikawa, H. Toyoda, S. Mizuno: 1ms Column Parallel Vision System and Its Application of High Speed Target Tracking, *Proc. IEEE International Conference on Robotics and Automation*, pp.650-655, 2000.

[5] U. Muehlmann, M. Ribo, P. Lang, A. Pinz : A New High Speed CMOS Camera for Real-Time Tracking Applications, *Proc. IEEE International Conference on Robotics and Automation*, pp. 5195-5200, 2004.

[6] I. Ishii, K. Kato, S. Kurozumi, H. Nagai, A. Numata, K. Tajima : Development of a Mega-pixel and Milli-second Vision System using Intelligent Pixel Selection, *Proc. IEEE Technical Exhibition Based Conference on Robotics and Automation*, pp.9-10, 2004.

[7] T. Komuro., I. Ishii, M. Ishikawa, A. Yoshida: A Digital Vision Chip Specialized for High-speed Target Tracking, *IEEE transactions on Electron Devices*, Vol. 50, No. 1, pp. 191-199, 2003.

[8] Eyelink II: http://www.sr-research.com/

[9] J. Watanabe, H. Ando, T. Maeda : Gaze-triggered selective information display, *Proc. ACM SIGCHI International Conference on Advances in computer entertainment technology*, pp. 10-17, 2004.

[10] S. Nikolov, T. Newman, D. Bull, N. Canagarajah, M. Jones, Iain D. Gilchrist: Gaze-contingent display using texture mapping and OpenGL: system and applications, *Proc. Symposium on Eye Tracking Research & Applications*, pp. 11-18, 2004

[11] K. Hinckley, M. Sinclair, E. Hanson, R. Szeliski, M. Conway: The VideoMouse: A Camera-Based Muilt-Degree-of Freedom Input Device, *Proc. ACM UIST'99 Symposium on User Interface Software & Technology*, pp. 103-112, 1999