ACCV2002: The 5th Asian Conference on Computer Vision, 23–25 January 2002, Melbourne, Australia.

High-Speed Image Processing System and It's Application

Kohtaro Ohba[1], Jesus Carlos Pedraza Ortega[2], Kazuo Tanie[1], Masataka Tsuji[3] and Shigeru Yamada[3]
{k.ohba,carlos-pedraza,tanie.k}@aist.go.jp, {yamada,tsuji}@photron.co.jp
[1] National Institute of Advanced Industrial Science and Technology, 1-2 Namiki, Tsukuba, 305-8564, JAPAN
[2] Tsukuba University, 1-1-1 Tennodai, Tsukuba, 305-8577, JAPAN
[3] Photron Co., 1-9-8 Shibuya, Shibuya-ku, Tokyo, JAPAN

Abstract— **In this paper, technique and criteria to realize the high-speed image processing system, which could capture and process with more than video-rate; $30frame/sec$, and it's applications are discussed. Then, as an example, an observational system for the tele-micro-operation has been proposed with a dynamic focusing system and a high-speed image processing system using the "depth from focus" criteria.**

In our past work [14], we had proposed the system which could obtain the "all-in-focus image" and the "depth" of object, simultaneously, in $2sec$ with a vision chip. To realize the real-time micro-operation, at least, $30frame/sec$ (60 times faster than the past) system should be required.

At first, this paper briefly summarizes the criteria and problems to realize the high-speed vision systems. Then, as an example for the high-speed vision, we propose a system to observe the micro environments to achieve the all-in-focus image and the 3D micro environments' reconstruction, simultaneously. After discussing the problem in our past system, a new video-rate system is proposed with the high-speed video camera and FPGA hardware. To apply this system in the real microscope, new criteria to integrate the all-in-focus image is proposed. Finally, the micro observation shows the validity of this system.

I. Introduction

The high-speed image capturing system had been utilized in several field to analyze the physical phenomena, such as human motion, water flow, air flow, and etc. In these analysis, the real-time processing is not necessary. After recording the high-speed image sequence, you can observe and analyze it in off-line. Actually, high-speed video camera, which is capable to capture the image more than $4,500fps$, is applied to analyze the cavitation in high-speed water flow [1].

Recently, there seems to be several applications for the high-speed vision system, which is capable to capture and process the image to extract particular information at more than video-rate, such as for the high-speed tracking system, the intelligent transportation systems (ITS), and etc. These applications could be categorized into two types;

- extract low data at high-speed,
- or integrate several images.

The first system needs only small data with high-speed, such as position, comparing to the original image itself. For example, human detecting system for ITS needs the human's position at each time. Secondly, several technique to get more functional image from multiple images are proposed [7], i.e. "integration". To integrate multiple image in real-time, high-speed vision system should be required.

To overcome the video-rate on vision, systems are proposed by many researchers all over the world for particular applications. In these days, the C-MOS technology has been developed, which could include the analog circuit into the sensor; i.e. vision chip [2] [3] [4] [5].

In this paper, at first, importance on high-speed vision is discussed, which includes the goodness and drawback on vision chip. Then, for an example of the high-speed vision application, micro observation system, which could give us the all-in-focus image and depth map, simultaneously, is proposed. Finally, the high-speed vision system with high-speed camera and FPGA technology, which could capture and process with $240fps$, shows the validity in the micro-operation usage.

II. What we needs for High-Speed Vision?

Generally speaking, the video-rate is good enough for human vision, but, sometimes, not enough as a sensor system. As the matter of fact, most of the robots are controlled with the time interval; $1msec$, beside the vision is based on the video-rate; $33msec$.

In the usual vision system, it obtains some particular information from the environments on the sensor, sends the information from the sensor to the processor, and processes the informations. To increase or integrate the information on vision system, we have to increase;

- dynamic range of sensor,
- data traffic from sensor to processor,
- and the performance of processor,

as shown in Fig.1.

Here in this paper, we would like to concentrate the data traffic and the performance of processor, because the dynamic range of sensor was discussed in the electric society to apply the particular usage. Fig. 2 shows the several architectures for the vision system. Fig. 2(a) shows the usual vision system, which include the single ADC to capture the image data into PC. Fig. 2(b) and (c) show the column and pixel parallel architecture to capture and process the multiple pixel image data at once. To construct

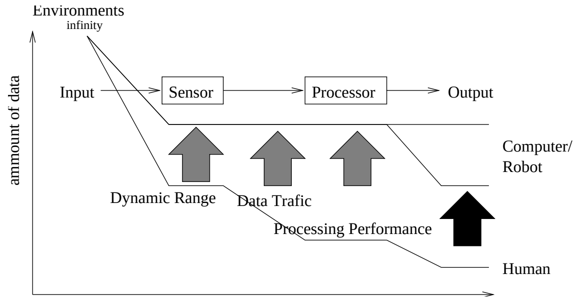


Fig. 1. Data Traffic in Vision System.

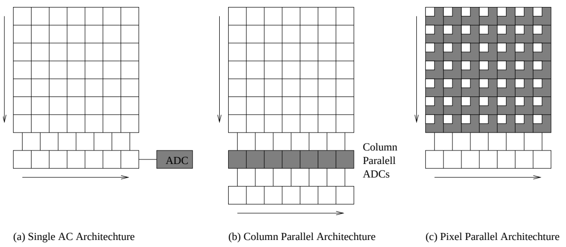


Fig. 2. Vision Chip Architectures.

these architecture, we should design and make special vision chip for particular purpose.

To increase the data traffic and processing performance more than frame-rate, several vision chip were proposed all over the world for particular use [3][5]. For example, the high-speed image processing system was proposed by Ishikawa for the robotics usage [2]. The device is constricted as the vision chip, in which each pixel is including the processor. Actually, the MAPP2200, which we used in our past work, is one of the vision chip with the column parallel architecture. The goodness on the vision chip might be

- low price as a device,
- high performance on particular process,
- and high data traffic rate from sensor to processor.

The biggest drawback might be the manufacturing cost, more than one million dollars to develop the device. Then, we should discuss a lot about the market.

III. High-Speed Vision System for Micro-environments

In this section, we would like to discuss the motivation to develop the high-speed vision system to realize the real-time micro-operation in the micro environments.

In the process of micro-operation, the observational sensor system in the micro environments to actuate objects in the micro world is becoming necessary in many fields, such as manufacturing and bio-medical usages. Somehow, the "optical scale factors" on this micro observation, i.e. the small depth of a focus on the microscope, could not allow us to feel the micro environments, intuitively.

Fig.3 shows the two microscopic top views in the typical micro-tele-operation, putting the second glass ball $\phi 4\mu m$ onto the first glass ball $\phi 4\mu m$ with a micro griper. Left

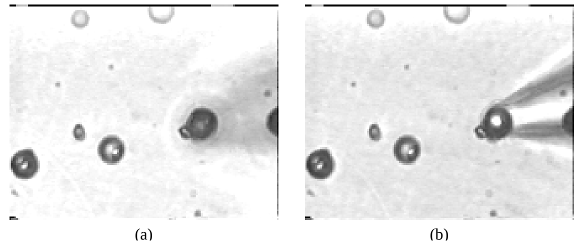


Fig. 3. Typical Microscopic Views in Micro Actuation.

figure (a) in Fig.3 shows the first glass ball in focused, but the gripper is blurring at almost same position in the different depth. And right figure (b) shows the gripper in focused. Therefore, the operator should be a professional who can operate the microscope and the micro-actuator, simultaneously.

As a results, two informations seem to be required in the micro observation;

1. all-in-focus image,
2. and depth information of objects.

"All-in-focus image" gives in-focused information all over the image, no blurring, but no depth information. And "depth information of objects" is the geometric information about objects, which could allow us to change the view point to operate the micro three dimensional objects.

Generally speaking, for the vision system in the micro environments, the "small depth of a focus" is one of the drawback for the micro-operation with the microscope, but, as the matter of fact, it is one of the benefits for the "shape from focus" criteria in the 3D modeling technique with computer vision [8]-[11]. There is another criteria;"depth from defocus", which could only obtain the depth information of object.

To realize the real-time operation in micro environments with the vision system, which could give us the all-in-focus image and depth information of object with the "shape from focus" criteria, the "integration" technique with high-speed vision system is required.

Actually, this is our motivation on this work.

IV. System with Vision Chip

To realize a real-time micro VR camera with the depth from focus criteria, the high-speed image capturing and processing system should be required for the integration. For example, if eight images are applied to integrate one all-in-focus image with $30 frame/sec$, $240 frame/sec$ image sequence is necessary to capture and process.

In our past work[14], we had proposed and developed a micro VR camera system shown in Fig.4 with the dynamic focusing lens [12] and smart sensor; IVP C-MOS vision chip (MAPP2200), which has the resolution of $256 * 256 pixel$, a column parallel ADC architecture and column parallel DSP processing [6].

A sample object in Fig.5, of which sample images at four particular focal distances are shown in Fig.6, was constructed in four steps' pyramid shape; first stage $\phi 10mm$

Fig. 4. First System Overview.

height 10mm, second $\phi 7mm$-10mm, third $\phi 4mm$-10mm, and top $\phi 3mm$-5mm, with textured paper printed with $h0.8mm$ and $w0.4mm$ fonts. In the case of real usage, such as less than 1mm object size, the image quality measurement(IQM) value could be calculated with the original texture on the object without any artificial texture with the equation;

$$IQM = \frac{1}{|D|} \sum_{x=x_i}^{x_f} \sum_{y=y_i}^{y_f} \left(\sum_{p=-L_c}^{L_c} \sum_{q=-L_r}^{L_r} |I(x,y) - I(x+p, y+q)| \right), \quad (1)$$

where $(-L_c, -L_r) - (L_c, L_r)$ and $(x_i, y_i) - (x_f, y_f)$ are the area for the evaluation of frequency and the smoothing, respectively[13]. And D is the total pixel number to make standard the image quality measure value with the number of pixels in the area $(-L_c, -L_r) - (L_c, L_r)$ and $(x_i, y_i) - (x_f, y_f)$. The spatial resolution is depending on the optical setting. For this demonstration, the view area is almost 16mm square with $256 \times 256 pixel$, then the spatial resolution is $62.5\mu m = 16mm/256$. The depth resolution is 1.67mm (21 frames with 35mm depth range, the each $3V$ input voltage from $-30V$ to $+30V$ to charge the PZT), which directly depends on the number of input frames in the range of variable focusing.

The all-in-focus image and the micro VR environments from one image sequence shown in Fig.7 and 8, respectively. The all-in-focus image is quite reasonable to observe. But the resolution of the depth without any interpolation in Fig.8 seems not enough. To increase the resolution of depth, any interpolation technique must be required.

The processing time with IVP chip is almost 2$sec.$ for one final VR output. This is caused by not enough ADC and processing performance for the gray level intensity on the vision chip MAPP2200. Actually, MAPP2200 has a good performance for the binary image more than 2000$frame/sec.$

See more detail in [14].

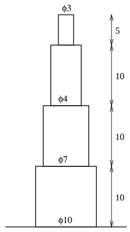


Fig. 5. Sample Object 1.

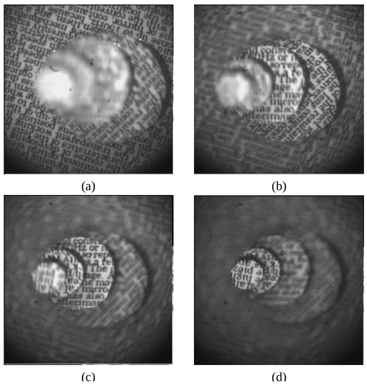


Fig. 6. Samples of the Single Focus Image.

Fig. 7. A Sample of the All-in-Focus Image.

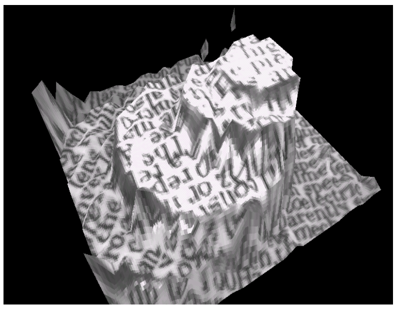


Fig. 8. A Sample of the Micro VR Environments.

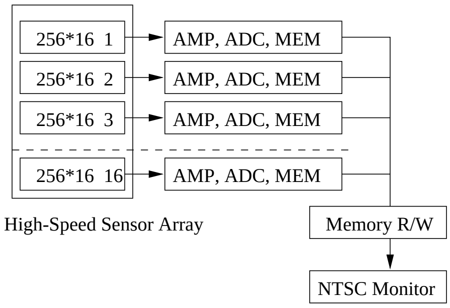


Fig. 9. A Sample of High Speed Video Configuration.

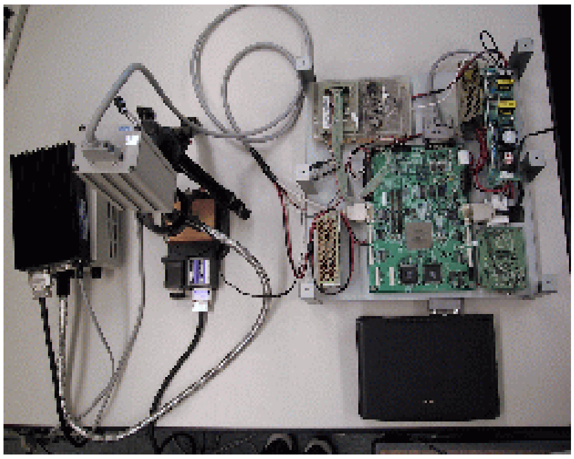


Fig. 10. Latest System Overview.

V. System with High-Speed Video (HSV)

A. HSV

The high-speed video (HSV) is widely used to observe the high speed phenomena in fluidics and physical field at more than $10,000,000 frame/sec$. To capture the image in high-speed on HSV, column parallel image capture is realized as shown in Fig.9. In this figure, the image 256*256 is divided into 16 of $256 * 16$ column, and each column captures each image data with $25MHz$, simultaneously. With this architecture, $4,500 frame/sec$ for $256 * 256$ full frame, and $40,000 frame/sec$ for $16 * 16$ was realized in the *FASTCAM* by *PHOTRON*. To increase the video-rate, the high S/N ratio with the large sensor area is also required.

For our system, almost the eight images are necessary to use for each all-in-focus image at $30 frame/sec$ output, i.e. $512 * 512$*pixel 8bits* $240 frame/sec$ image capture and processing is finally required. To realize this, we decided to use the HSV as the high-speed image capture system, and construct the image processing system based on this HSV.

B. Processing Part

Recently, the large scale FPGA (Field Programmable Gate Array) is drastically getting good performance, and widely used because of the programmable capability.

Then, in the latest system shown in Fig.10, one *FPGA (APEX EP20K600E, ALTERA)* and *SDRAM* in the image processing test board *(iLS-BEV800, INNOTECH Co.)* are applied to use to calculate the IQM value at each pixel all over the image with $240Hz$, which has the TMDS (Transition Minimized Display Signaling) architecture interface to connect between the sensor part and processing part as shown in Fig.11. Then, the image data $512 * 480$ is captured with two parallel interface and realize the high-speed data traffic $60 Mbyte/sec (512 * 480 * 240Hz)$ from HSV to the dual-port SDRAM. As the results, the performance of the FPGA is good enough to calculate the IQM value with $240Hz$, and the total execution performance is less than 20% of the performance of FPGA.

Up to now, the all-in-focus image and the depth image are stored in each memory space, and could be observed with VGA monitor through the analog RGB output, separately. The VR display might be realized in PC after transmit the all-in-focus and depth image into the PC, directly.

C. Optics Part

An micro zoom lens (Asahi co. MZ-30, f=27.2-58.5mm, F7.5) with the dynamic focusing lens [12] is attached on the HSV. The dynamic focusing lens is controlled by FPGA with ramp input through DAC and Amp. The relation between the input voltage to PZT and the objective distance was evaluated to be linear in the range from $147mm$ to $180mm$ corresponding to the input voltage from $-30V$ to $+30V$ in our past work[14]. And the resolution of the objective distance seems to increase with the number of frames. Actually, we had eight images for each video-frame, but the first image is not stable for the vibration of the variable focusing lens. Finally, only seven images are applied to integrate.

The spatial resolution in this system is $31.25\mu m$. The depth resolution is $5.0mm$ (7 frames with $35mm$ depth), which can be improved with the input frame number.

D. Experimental Results

Fig.12 shows the example results on the all-in-focus and depth image with the latest system. The sample 1 is the same object as in the past work shown in Fig.5-8, and the sample 2 is the almost same size as the sample 1, but has the cone slope. In the figure 12(d), you can see the slope in the depth.

Actually, this output results is **real-time movie** on this system. Even though, the operator put the gripper in the sight, in-focus image could allow us to observe the object and the gripper, simultaneously, which is located in the different depth.

VI. Micro Application

A. System

For the real micro applications, microscopic system is developed with HSV and processing part mentioned before as shown in Fig.13. Instead of using the variable focusing

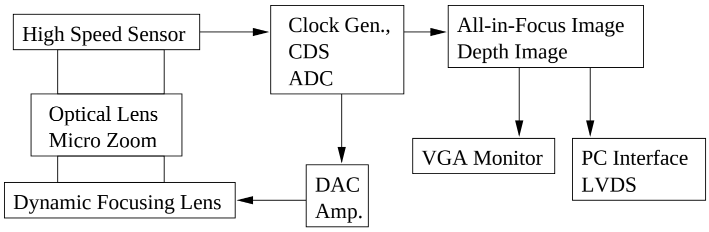
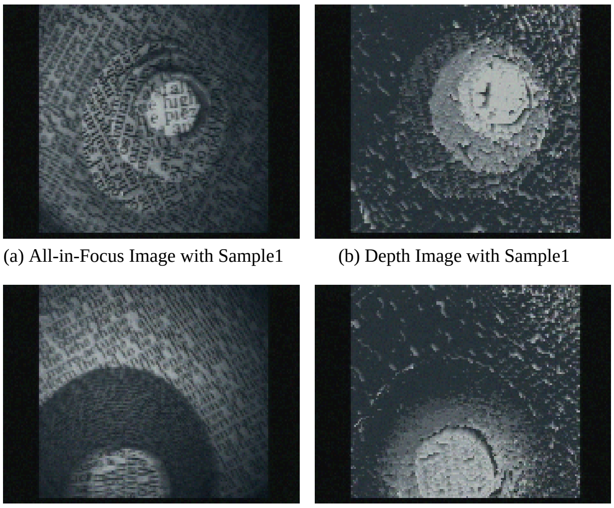


Fig. 11. Block Diagram on Latest System.



(a) All-in-Focus Image with Sample1 (b) Depth Image with Sample1

(c) All-in-Focus Image with Sample2 (d) Depth Image with Sample2

Fig. 12. Sample Movies with All-in-Focus and Depth Image.

lens, which was used in the previous system, a *PIFOC microscope objective nano positioners and scanners P-721.20, PI-Polytec Co.* is controlled by *a high-speed nano automation controllers E-612.C0, PI-Polytec Co.*, which could be controlled the object lens in the microscope, *BX60, Olympus Co.*. The focus length is controlled $0 - 100\mu m$ maximum as the actuator input voltage $0 - 10V$ $30Hz$ ramp input from the FPGA. And the real position could be observed with the sensor output from the controller.

We apply to observe two glass fabrics $\phi 4\mu m$ each located in micro 3D environments with the object lens $50X$ in the microscope. One fabric is located in near distance, and the other is in far distance. Fig.14 show the usual focus images scanning the focus length from $0\mu m$ to $90\mu m$. Near fabric is in focus in the figure 14(e), and far fabric is in figure 14(h).

The spatial resolution in this system is $0.195\mu m$, $100\mu m/512pixel$. The maximum depth resolution is $0.781\mu m$, $100\mu m/128bits$.

Fig.15 shows the all-in-focus image in microscope. Comparing with the Fig.14, you can clearly see the both fabrics in-focused in one sight.

B. Ghost Problem

In the detail analysis in Fig.15, several blurring edge could be observed just around the objects in Fig.15. These

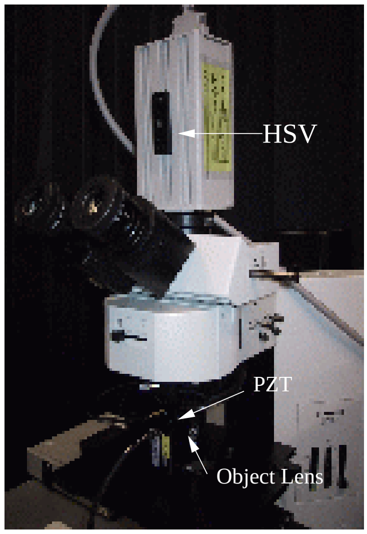


Fig. 13. Micro Scopic System.

blurring ghost is caused by the several out-of-focus images. In the microscope, the out-of-focus image could make large blurring region around the real object as you can see in Fig.14. This blurring region may cause the miss-recognition of the all-in-focus area around the objects.

C. Ghost Filtering Technique

To solve the ghost problem, the reliability of the IQM value should be evaluated to eliminate the blurring area. Then, the minimum IQM value;IQM_{min} is pre-defined, which could hold the in-focus clearly in the particular image sequences.

$$in-focus \quad : \quad IQM(x,y,f) \geq IQM_{min}, \qquad (2)$$

$$background \quad : \quad otherwise \qquad (3)$$

where $IQM(x,y,f)$ is the image quality value at image location;(x,y) with the focus length;f. Fig.16 shows the validity with this ghost filtering technique.

VII. Conclusion

In this paper, there was several discussion on the high-speed vision system. Then, a real-time micro VR sensor system was developed to achieve the micro 3D model and all-in-focus images at video-rate with the criteria of "depth from focus". This system is constructed with the high-speed video and FPGA hardware, and realize the real video-rate $30frame/sec.$ observation. And to apply this system to the microscope, the ghost filtering technique is proposed. Microscope system shows the validity in the microscopic application.

Acknowledgments

This research was done as the cooperative research between the National Institute of Advanced Industrial Science and Technology and Photron Co., based on the past cooperative research with DHT Corporation, Kawasaki

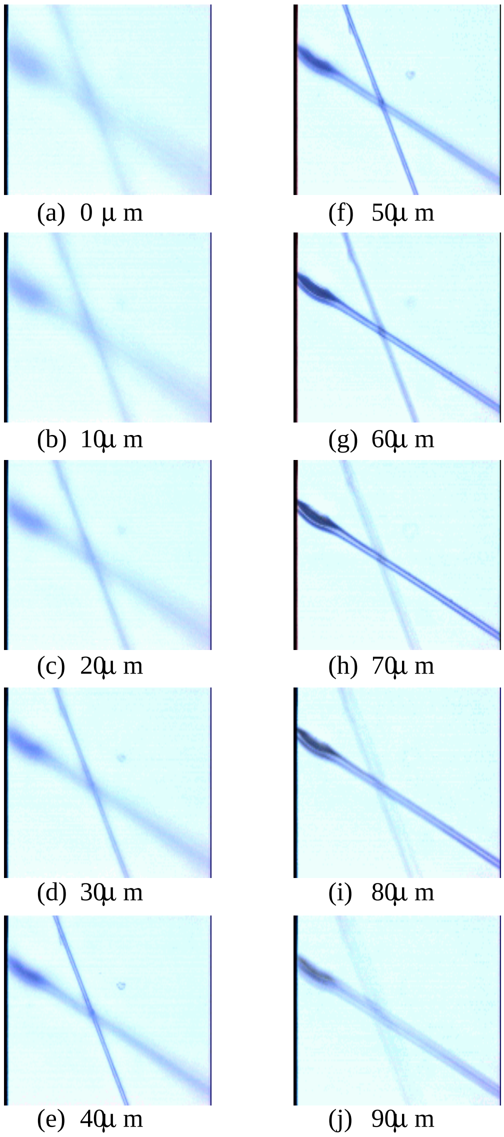


Fig. 14. Sample Images of Single Focus Image in Micro Application.

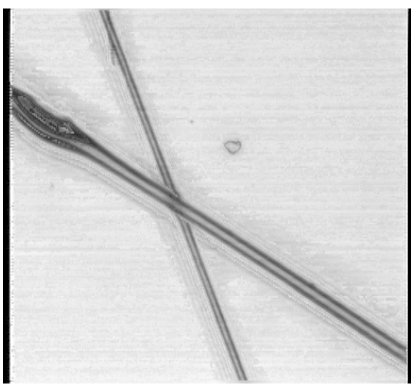


Fig. 15. A Sample of All-in-Focus Image in Micro Application.

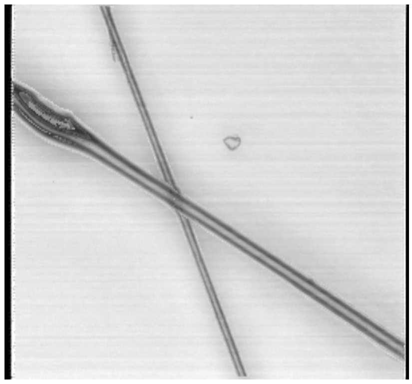


Fig. 16. A Sample of All-in-Focus Image with Ghost Filtering.

Steel Techno-Research Corporation and DENSO Corporation.

And, a part of this work for the dynamic focusing lens was performed under the management of the Micromachine Center as the Industrial Science and Technology Frontier Program, "Research and Development of Micromachine Technology", of MITI supported by New Energy and Industrial Technology Development Organization.

References

[1] Kohtaro Ohba, Hitoshi Soyama, Sho Takeda, Hikaru Inooka and Risaburo Oba, "High-Speed Observations on Highly Erosive Vortex Cavitation Using Image Processing", Journal of Flow Visualization and Image Processing, Vol.2,No.2,pp.161-172, May,1995
[2] Akio Namiki, Yoshihiro Nakabo, Idaku Ishii and Masatoshi Ishikawa, "1ms Sensory-Motor Fusion System", IEEE Trans. on Mechatoronics, Vol.5, No.3, pp.244-252, 2000.
[3] Bladimir Brajovic and Takeo Kanade, "Sensory Attention: Computational Sensor Paradigm for Low-Latency Adaptive Vision", DARPA Image Understanding Workshop 1997.
[4] K.Kyuma, E.Lange, J.Ohta, A.Hermanns, B.Banish, and M.Oita, "Artificial Retinas -Fast, Versatile Image Processors", Nature, 372, 6502, pp.197-198, Nov.1994
[5] Alireza Moini, "Vision Chips or Seeing Silicon", http://www.eleceng.adelaide.edu.au/Groups/GAAS/Bugeye/visionchips/index.html, March 1997.
[6] "User Documentation MAPP2200 PCI System", IVP, 1997.
[7] Takashi Matsuyama, "Cooperative Distributed Vision –Dynamic Integration of Visual Perception, Action, and Communication –", Proc. of Image Understanding Workshop, Monterey CA, 1998.11
[8] Kazuya Kodama, Kiyoharru Aizawa, and Mitsutoshi Hatori, "Acquisition of an All-Focused Image by the Use of Multiple Differently Focused Images", The Trans. of the Institute of Electronics, Information and Communication Engineering Engineers, D-II, Vol.J80-D-II, No.9, pp.2298-2307, 1997.
[9] Masahiro Watanabe and Shree K. Nayer, "Minimal Operator Set for Passive Depth from Defocus", CVPR'96, pp.431-438, 1996.
[10] Shree K. Nayer, Masahiro Watanabe, and Minoru Noguchi, "Real-Time Focus Range Sensor", ICCV'95, pp.995-1001, 1995.
[11] Shree K. Nayer, and Yasuo Nakagawa, "Shape from Focus", IEEE Trans. on PAMI, Vol.16, No.8, pp.824-831, 1994.
[12] Takashi Kaneko, Takahiro Ohmi, Nobuyuki Ohya, Nobuaki Kawahara, and Tadashi Hattori, "A New, Compact and Quick-Response Dynamic Focusing Lens", Transducers'97, 1997.
[13] Sridhar R. Kundur and Daniel Raviv, "Novel Active-Vision-Based Visual-Threat-Cue for Autonomus Navigation Tasks", Proc. CVPR'96, pp.606-612, 1996.
[14] Kohtaro Ohba, Jesus Carlos Pedraza Ortega, Kazuo Tanie, Gakuyoshi Rin, Ryoichi Dangi, Yoshinori Takei, Takashi Kaneko and Nobuaki Kawahara, "Implementation of Real Time Micro VR Camera", Trans. IEE Japan, Vol.120-E, No.6, pp.264-271, 2000.