955-fps Real-time Shape Measurement of a Moving/Deforming Object using High-speed Vision for Numerous-point Analysis

Yoshihiro WATANABE, Takashi KOMURO and Masatoshi ISHIKAWA¡-this
Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.
E-mail: Yoshihiro_Watanabe@ipc.i.u-tokyo.ac.jp

Abstract— **This paper describes real-time shape measurement using a newly developed high-speed vision system. Our proposed measurement system can observe a moving/deforming object at high frame rate and can acquire data in real-time. This is realized by using two-dimensional pattern projection and a high-speed vision system with a massively parallel co-processor for numerous-point analysis. We detail our proposed shape measurement system and present some results of evaluation experiments. The experimental results show the advantages of our system compared with conventional approaches.**

I. INTRODUCTION

There has been an increasing need for shape measurement techniques in a wide range of application fields. In particularly, approaches using vision are useful because of their contactless nature and flexibility. On the other hand, there are demands for improved performance such as accuracy, measurement range, time, cost and so on, and many kinds of systems have been developed to meet these demands [1].

In this paper, we focus on the improvement of the measurement time and realize high-speed time-sequential shape measurement. This type of measurement will have a great impact on a wide variety of applications, such as control tasks for robotics; surgery support, for example, observation of a beating heart; rapid visual inspection of products; vehicle application, for example, checking road surfaces; and human-machine interaction with high accuracy and flexibility.

On the other hand, in such application fields, we need to observe a high-speed moving rigid-body and a deforming or vibrating non-rigid body. Conventional shape measurement, however, is mainly performed in situations where the relationships between the objects being measured and observation devices are completely known by ensuring stable or controllable conditions. Also, the performance of conventional systems in terms of throughput and latency is usually not good enough for real-time control applications; for example, throughput on the order of a thousand frames per second and latency on the order of milliseconds is generally required for robotic control.

Based on these backgrounds, the feasible system is considered to have three requirements: (A) the ability to observe a moving object, (B) continuous observation at high frame rate, and (C) real-time acquisition of shape information. To realize these requirements, our proposed system employs an approach in which multiple spots in a single image of a shape are acquired using high-speed vision for numerous-point analysis. As a result, the constructed system achieved 955-fps measurement of shape information with 4.5 ms

latency. This paper details the system architecture and the proposed operations, and demonstrates its advantages with some experimental results.

II. CONCEPT AND STRATEGY

Our system has the following three functions: two-dimensional pattern projector, high-frame-rate imaging, and high-speed image processing system. This section shows that the integration of those functions newly realizes effective measurement for a moving/deforming object.

There have been various methods for vision-based shape measurement, such as the time of flight method, active triangular method, moire fringes, interferometry, and so on[1]. Among those, the active triangular method, which basically uses a single slit light, achieves sufficient accuracy on the order of millimeters with a simple setup. Therefore, most applications requiring three-dimensional information employ this approach.

However, this method involving laser scanning needs typically tens to hundreds of frames to construct a single range map, the exact number being proportional to the resolution and inversely proportional to the throughput. With this method, it is necessary to assume that the measured object remains stationary until measurement is completed, which causes a critical problem if the object moves between frames. In order to avoid this problem, this method requires extremely high-frame-rate imaging, which is not feasible in terms of application development because it involves a reduction in image quality and great increase in projected light intensity.
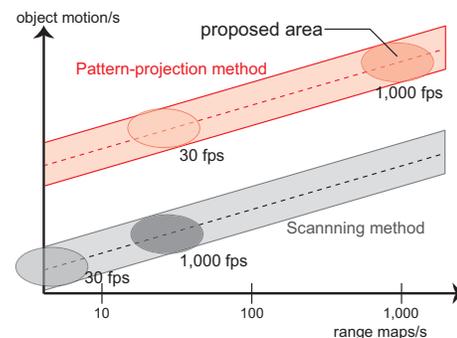
Fig. 1. Performance comparison of shape measurement approaches.

Compared with this, there is another method using two-dimensional pattern projection acquiring shape information in a single frame by projecting a pattern to whole measured space. This can be a useful solution to avoid this

motion constraints and realize high-throughput and low-latency shape measurement of deforming or moving objects. Fig. 1 illustrates the comparison between a scanning method and a pattern projection method with its performance about throughput and motional allowability. This figure also shows our target field, where shape information feedback rate is 1 kHz. In this field, the advantages of the pattern projection method are considered to work quite effectively. Although it cannot keep constant resolution about space in wide range, the following two reasons keep its usefulness. First, there are various applications even in case that the range is limited. Second, it can be solved by changing the projected pattern dynamically, which has a high feasibility.

However, in order to achieve high resolution about both of space and time, the method based on the pattern projection requires to observe high-speed changes in the fine pattern. Therefore, the high-frame-rate imaging, that is the second function mentioned above, is an essential component.

In addition, this method requires quite high amount of processing. For the performance about high throughput and low latency, the processing speed needs to be dramatically improved. The high-speed image processing system, that is the third function, plays an important role about this requirement. Especially, in the proposed system, the second and third function is provided by our newly developed high-speed vision system.

III. REAL-TIME SHAPE MEASUREMENT USING HIGH-SPEED VISION FOR NUMEROUS-POINT ANALYSIS

This section describes the whole structure of our system and the essential components including the measurement principle and high-speed vision for numerous-point analysis.

A. Whole System
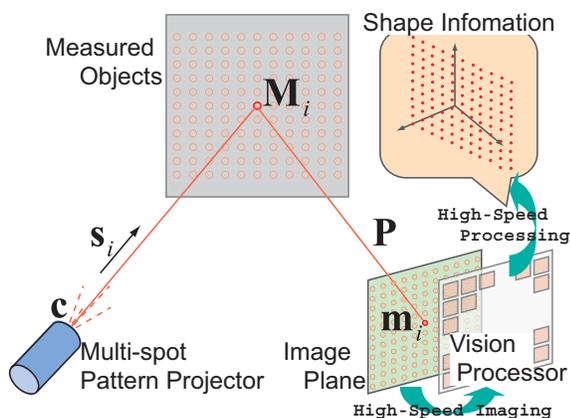
Fig. 2. Shape measurement using high-speed vision for numerous-point analysis.

Fig. 2 shows the basic concept of our system. As the measurement method, we use a multi-spot projection method in which numerous pre-calibrated spots are projected into the measurement space using a laser. In the remainder of this paper, the number of frames per second (fps) is the same as the rate used for acquiring a range map because the shape information is obtained in a single frame.

In the related method, other patterns may be used, such as multiple slits or a grid. Also, it is possible to use color patterns by using an LCD projector[2]. There are two reasons for using a focused multi-spot pattern from a laser: to achieve high accuracy in each spot to avoid intensity noise, and to simplify the image processing for improving the speed, resulting in high-frame-rate imaging and high-throughput, low-latency visual feedback, respectively. This multi-spot method has also been used to develop a compact range sensor for a robot[3]. However, it has been difficult for this method to be used in real-time shape measurement at 1 kHz.

B. Measurement Principle

The principle of this method is shown in Fig. 2. Each measured spot lies on the intersection of two lines, a projection line and a vision constraining line. Therefore, if geometric information about the projected line is known, the three-dimensional point $M_i = [x_w, y_w, z_w]^t$ can be obtained from the image point $m_i = [X_v, Y_v]^t$. The suffix i shows the spot number. The expression for the projection line is:

$$M_i = c + \delta s_i \quad (i = 1, \ldots, N_p). \tag{1}$$

This is a line with gradient s_i passing through the projector center c and on which the measured spot i lies. Here, N_p is the total number of projected spots.

The expression for the vision constraining line is:

$$P\tilde{M}_i = \omega\tilde{m}_i. \tag{2}$$

This expresses the relationship between the image point $\tilde{m}_i = [m_i^t, 1]^t$ of spot i and a three-dimensional point $\tilde{M}_i$ connected by perspective projection matrix P.

When applying this measurement, c, s_i, and P are known parameters, and $\tilde{m}_i$ is the observed data. The three-dimensional point M_i is obtained from the above two equations form the observed image points.

C. High-speed Vision for Numerous-point Analysis

The method using multi-spot pattern involves extraction of position information of the numerous spots from an image, which requires long processing time. This bottleneck causes a drop in the frame rate in real-time operation. Therefore, to avoid this bottleneck, we introduce a newly developed high-speed vision for numerous-point analysis.

There have been some studies using real-time vision at high frame rates such as Column Parallel Vision (CPV)[4] and CMOS+FPGA vision[5] in manipulation tasks[6] and microscopic applications[7]. CPV has a general-purpose, massively parallel SIMD processor. However, the operated processing has been very simple, for example, just calculating the centroid of a single target because its processor architecture limits the types of image processing. Also those vision systems connected a processing unit to a camera directly. This is considered to yield disadvantages in terms of scalability of system functions about module exchanging and coordination with other heterogeneous processors.

Our newly developed system enables high-speed image processing requiring a large amount of calculations by installing a massively parallel and dedicated vision processing

Fig. 3.   High-speed vision system with an massively parallel processor.

Fig. 4.   Operation flow.

IV. OPERATION FOR SHAPE MEASUREMENT BASED ON NUMEROUS-POINT ANALYSIS

The proposed flow of operations from extracting the image point s_i to obtaining the three-dimensional point M_i is shown in Fig. 4. Each operation is detailed below.

A. Calibration

First of all, the geometrical relationship between the camera and each projected spot must be determined in advance. We call this operation calibration. As an example of a simple method, the calibration can be achieved by determining three functions $[x_w, y_w, z_w]^t = [f_1^i(z_w), f_2^i(z_w), f_3^i(X_v)]^t$ from known pairs of three-dimensional points M_i and image points m_i of each projected spot i without obtaining intrinsic parameters c, s_i, and P.

The functions f_1^i and f_2^i determine the x_w and y_w coordinates of the three-dimensional point for spot i from the depth distance, that is the z_w coordinate. The relationships are expressed as linear function:

$$f_j^i(z_w) = \alpha_{j,1}^{(i)} z_w + \alpha_{j,0}^{(i)} \quad (j = 1, 2) \qquad (4)$$

The function f_3^i determines the depth distance z_w from the X_v coordinate of an image point. Strictly speaking, this function is expressed as a hyperbola about X_v and Y_v. However, over a small range, this function can be accurately approximated by a polynomial expression as below:

$$f_3^i(X_v) = \sum_{k=1}^{n} \alpha_{3,k}^{(i)} X_v^k \qquad (5)$$

In the actual experiment, we used a two-dimensional polynomial approximation. Also those functions were determined by obtaining multiple spot patterns to $x_w y_w$ planes at the known different distance z_w.

B. Initialization

The processing system needs to know which image point corresponds to each projected spot. Our method uses the information of the previous frame at the high-frame-rate imaging and achieves the matching between them. This tracking-based method can easily make the dynamic modification of the search area according to pattern changes. On the other hand, at beginning of the measurement, it requires another operation. We call this operation initialization.

In the applications focusing on moving objects, imaging all unknown spots at one time and stopping the motion of the measured object have the drawback of feasibility although they easily solve the problem. We therefore control the start time $t(i)$ of projecting about each spot i as follows:

unit as a co-processor in a PC. Its configuration diagram is shown in Fig. 3. The co-processor also is exchanged according to applications. The ability to swap the co-processor and the bus expansion of the co-processor and other devices allow reconfigurability of the system and ensures high performance in various applications. Based on this configuration, we construct high-speed vision for numerous-point analysis by providing effective modules to co-processor. In the remainder of this section, the details of each component are described.

We mounted the co-processor on a large-scale FPGA board (KAC02-A, made by Mitsubishi Electric Engineering Company). This board had two FPGA (XILINX XC2V6000) for implementing co-processor cores. Also, we used a high frame rate camera (DALSA CA-D6) with a resolution of 256×256. The 8-bit gray-scale images were captured and transferred in real-time at 955 fps. New sensors that are capable of high-frame-rate image transfer could be developed based on parallel A-D conversion technology[8]. The host PC was a Dell Precision 670 with an Intel Xeon 2.80-GHz main CPU. The co-processor communicated with the host via a 64 bit/66 MHz PCI bus.

In this paper, the co-processor calculated image moments as spot information. The image moments are parameters that can be converted to various geometric features, such as size, centroid, orientation, rough shape information, and so on [9]. The $(i + j)$th image moments m_{ij} are calculated from:

$$m_{ij} = \sum_x \sum_y x^i y^j I(x, y), \qquad (3)$$

where $I(x, y)$ is the value at pixel (x, y). Normally, when image moments of multiple objects are calculated in an $N \times N$ image, the amount of calculations following moment calculation and region division is proportional to the number of objects n and $O(N^2)$. In contrast, by providing a massively parallel architecture, our processor requires only $O(\sqrt{n})$[10]. This allows observation of a few thousand objects at frame rates on the order of thousands of frames per second.

$$t(i) = T_\delta \quad (i \in A_\delta : \delta = 1, \ldots, N_e) \qquad (6)$$

Here, A_δ is a class of projected spots, whose epipolar lines $l_i(Y_v = l_i(X_v))$ constraining movement of spot i in image space don't intercross. N_e is the number of divided classes. Fig. 4 shows the case under the condition that the points at the same line are regarded as the same class. This enables ambiguity in the correspondence to be avoided, which ensures high versatility. Moreover, since this spot pattern is already projected when commencing sequential frame operation, as detailed in IV-C, there is no loss of shape information after beginning the measurement.

C. Sequential frame operation

This section describes the operation in normal frames. After initialization, shape information is measured only in input frames. When the frame rate is relatively high compared to changes in the measured phenomena, it can be assumed that the difference between spots projected on a smooth surface between successive frames is small. Using this assumption, the operation for corresponding an image point to a spot i could be expressed as a tracking operation between frames, in which a point $m_i(t-1)$ corresponding to a point $m(t)$ is searched for using corrected points at time $t-1$ based on the evaluation $\min\{|m_i(t-1) - m(t)| + |M_i(t-1) - \tilde{M}(t)|\}$.

Such searching of neighbour points in two-dimensional image space can be performed using the bucket method, which can efficiently perform the search operation of the nearest point to an input point by dividing the search space into grids and accessing only neighbour areas, instead of accessing all points. This makes the number of calculations required linear with respect to the number of measured image points to be corresponded, so long as the set of points is distributed equally, which results in an equal number of points included within each grid. We call this operation the tracking in Fig. 4.

On the other hand, there are points that move discontinuously because they are on points of contact between the measured object and the projected line of the spot. These points are mapped exceptionally by using the epipolar line. This search operation is based on the evaluation $\min\{|Y_v(t) - l_i(X_v(t))|\}$.

The relative number of these discontinuously moving points can be assumed to be small, and this operation does not cause a crucial problem in terms of processing time. We call this operation the new point mapping operation in Fig. 4. It is necessary, however, to define constraints for the speed at which these points jump in the depth direction between frames in order to avoid the overlapping of spots in image space.

D. Performance of Real-time shape measurement

This section describes the performance about measurement speed using the proposed system and operations. The timing chart is shown in Fig. 5. First, images were captured (Ex) and stored in main memory through the capture board (R1/R2).
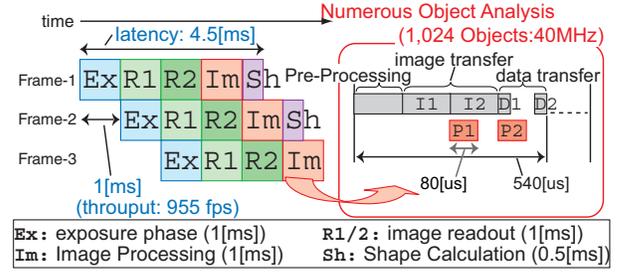
Fig. 5.   Timing in system operation.

Next, the image was converted to a binary image on the main CPU (Pre-Processing) and transferred to the two co-processors (I1/I2). Each co-processor received 256×128 image and extracted image moments of numerous spots (P1/P2) and stored the data in the host (D1/D2). Finally, the main CPU calculated shape information (Sh) based on the method described in IV.

The co-processor was operated at 40 MHz and extracted 0th and 1st order image moments of 1,024 spots from an image. Also, the number of regions and the order of image moments can be changed.

Each stage was pipelined and operated at 955 fps. This means that our system could achieve measurement with a throughput of 955 fps and a latency of 4.5 ms. It is considered enough for most applications. However, this is not the performance limit of the system because there is still room for improvement in the processing time of the co-processor and the total latency. Also we can include simple operations for improvement of the image quality such as sharpening at pre-processing on the CPU.

V. EVALUATION EXPERIMENT

A. Experimental Setup
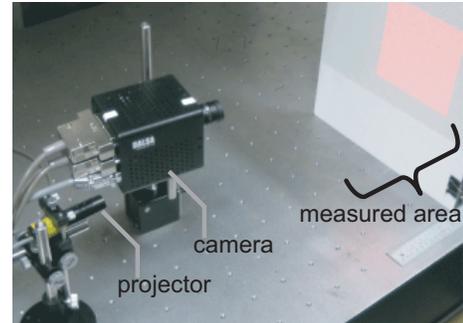
Fig. 6.   Experimental setup.

This section describes the experimental setup. Fig. 6 shows a photograph of the system. As the laser projector for the multi-spot pattern, we used SNF-533X (33×33 dots, line pitch 0.45 deg) made by Moritex.

The measurement results for planes are shown in Fig. 7. The $x_w y_w$ plane is defined as a plane perpendicular to the base. Fig. 7(a) shows the three-dimensional points of each measured spot at distances of $z_w = -60$, 0, 40 mm. The measured area in this experimental setup is also shown in the figure. The distance between the image plane and the center of the measured area was about 40 cm. This setup

(a) Measured Result

(b) Error Evaluation

Fig. 7. Measurement results for planes.

(a) 0 ms

(b) 20 ms

(c) 40 ms

(d) 60 ms

(e) 80 ms

(f) 100 ms

Fig. 9. Sequential measurement results of the waving flag.
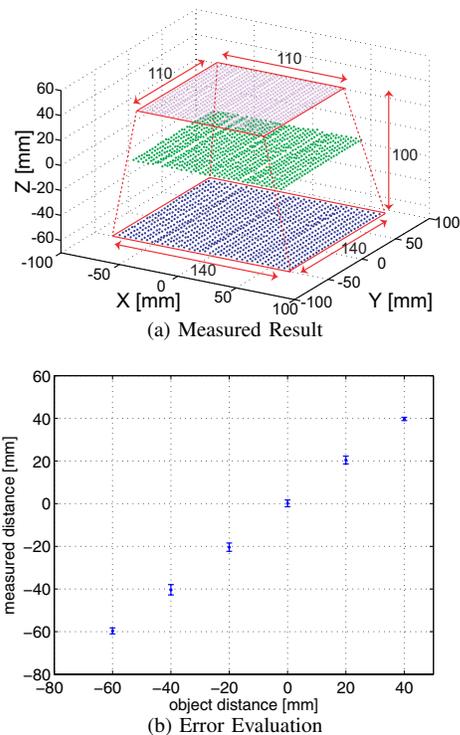
is an example and not the measurement limitation. Fig. 7(b) shows the average object distance, that is, z_w coordinate, at each plane, together with maximum and minimum bars. From this figure, it is clear that the calibration was performed correctly with about the order of millimeter accuracy. Also, the allowable difference of each spots in the space between frames was defined as a maximum of 50 mm/ms in this experiment.

B. Experimental Results

Here, we describe the measurement results for three different types of object: a waving flag, a flying ball, and moving hand. The motion of these objects is usually fast and is difficult to measure with conventional systems.
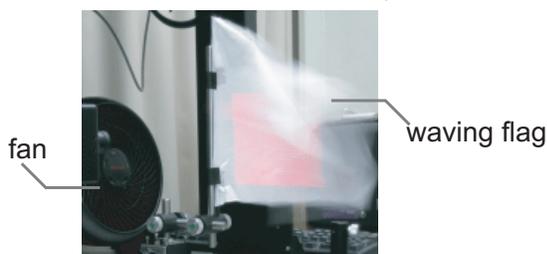
Fig. 8. Waving flag.

1) Waving Flag: In this experiment, we used a piece of cloth as a flag and fixed it at one side. The cloth was then blown from the fixed side by a fan and to produce waves like a flag blowing in the wind. A photograph of the waving flag is shown in Fig. 8.

Fig. 9 shows time-sequential measurement results for the waving flag. In this figure, the points show three-dimensional points of the measured spots. The estimated surface is also plotted. The measurement results show that the system

correctly measured the motion of a non-rigid body in a high-speed fluid. These results suggest the possibility of using this system in applications such as inspection of dynamic deformations and control of complex vibrations.
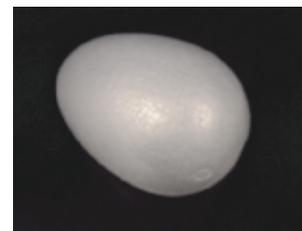
Fig. 10. Photograph of measured ball.

2) Flying ball: The next measurement object was a moving rigid body, namely, a flying ball. In this measurement, the background plane shown in Fig. 6 was set between the camera and measurement space. Fig. 10 shows a photograph of the measured egg-shaped ball.

The measurement results are shown in Fig. 11. The results show that the shape of the ball was measured almost correctly while it was moving.

One possible application of this time-sequential shape and motion measurement is dynamic manipulation of a rigid body, which plays an important role in robot tasks. To develop such an application, however, the system will need to handle more complex shapes. This will be the subject of future work.

3) Moving Hand: The third measurement object was a moving hand. In this experiment, an open hand came into the measurement space between the camera and background plane described above, and it was then closed. Fig. 12 shows
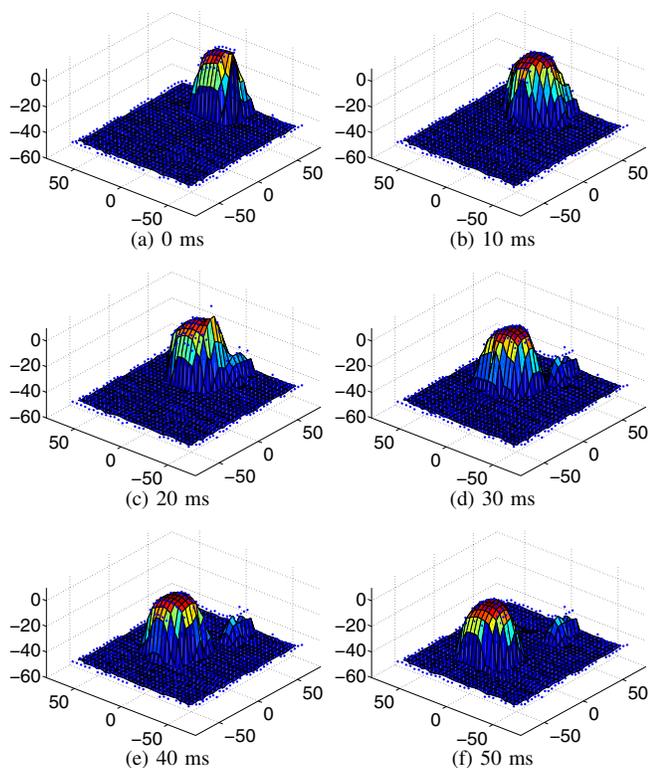
Fig. 11. Time-sequential results for flying ball.

Fig. 12. Photograph of the measured hand.

Fig. 13. Time-sequential results for moving hand.

a photograph of the measured hand.

Fig. 13 shows the measurement results. We can see that the shape and motion of the hand was correctly measured. These results suggest the possibility of using this system in a flexible, high-precision man-machine interface.

VI. CONCLUSION

In this paper, we have described a time-sequential high-frame-rate shape measurement system for moving and deforming objects. The system is based on multi-spot projection measurement in a single image using a high-speed vision with a co-processor for numerous-point analysis. We demonstrated a throughput of 955 fps and a latency of 4.5 ms. Experimental results for three kinds of objects showed that the objects were measured correctly.

This type of measurement system is expected to find uses particularly in feedback applications such as robotic manipulation, inspection, vibration suppression, human-machine interfaces, and so on. We believe that our system is highly promising for these kinds of applications. As a next step, in parallel with development of applications, we should consider improving the accuracy by increasing the number

of spots, introducing dynamically changeable projector, integrating shape information between frames, and so forth. This improvement should make the system even more feasible for the applications mentioned above.

REFERENCES

[1] F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Journal of Optical Engineering*, vol. 39, no. 1, 2000.

[2] P. Fong and F. Buron, "High-resolution three dimensional sensing of fast deforming objects," in *Proceedings of International Conference on Intelligent Robotics and Systems*, 2005, pp. 1606–1611.

[3] K. Umeda, "A compact range image sensor suitable for robots," in *Proceedings of International Conference on Robotics and Automation*, 2004, pp. 3167–3172.

[4] Y. Nakabo, M. Ishikawa, H. Toyoda, and S. Mizuno, "1ms column parallel vision system and its application of high speed target tracking," in *Proceedings of 2000 IEEE International Conference on Robotics and Automation*, 2000, pp. 650–655.

[5] K. Shimizu and S. Hirai, "CMOS+FPGA vision system for visual feedback of mechanical systems," in *Proceedings of 2006 IEEE International Conference on Robotics and Automation*, 2006, pp. 2060–2065.

[6] N. Furukawa, A. Namiki, T. Senoo, and M. Ishikawa, "Dynamic regrasping using a high-speed multifingered hand and a high-speed vision system," in *Proceedings of International Conference on Robotics and Automation*, 2006, pp. 181–187.

[7] N. Ogawa, H. Oku, K. Hashimoto, and M. Ishikawa, "Microrobotic visual control of motile cells using high-speed tracking system," *IEEE Transactions of Robotics*, vol. 21, no. 4, 2005.

[8] S. Kleinfelder, S. Lim, X. Liu, and A. E. Gamal, "A 10000 frames/s CMOS digital pixel sensor," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 12, pp. 2049–2058, 2001.

[9] R. J. Prokop and A. P. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," *Graphical Models and Image Processing*, vol. 54, no. 5, pp. 438–460, 1992.

[10] Y. Watanabe, T. Komuro, S. Kagami, and M. Ishikawa, "Parallel extraction architecture for image moments of numerous objects," in *Proceedings of IEEE 7th International Workshop on Computer Architecture for Machine Perception (CAMP'05)*, 2005, pp. 105–110.