2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

2000 fps Real-time Vision System with High-frame-rate Video Recording

Idaku Ishii, Tetsuro Tatebe, Qingyi Gu, Yuta Moriue, Takeshi Takaki, and Kenji Tajima

Abstract—**This paper introduces a high-speed vision system called IDP Express, which can execute real-time image processing and high frame rate video recording simultaneously. In IDP Express, a dedicated FPGA (Field Programmable Gate Array) board processes 512 × 512 pixel images from two camera heads by implementing image processing algorithms as hardware logic; the input images and processed results are transferred to standard PC memory at a rate of 2000 fps or more. Owing to the simultaneous high-frame-rate video processing and recording, IDP Express can be used as an intelligent video logger for long-term high-speed phenomenon analysis even when the measured objects move quickly in a wide area. We applied IDP Express to a mechanical target tracking system to record a high-frame-rate video at high resolution for a crucial moment, which is magnified by tracking the measured objects with visual feedback control. Several experiments on moving objects that undergo sudden shape deformation were performed. The results of the experiments involving the explosion of a rotating balloon and the crash of falling custard pudding have been provided to verify the effectiveness of IDP Express.**

I. Introduction

Many high-speed robot vision systems that operate at higher frame rates than the video rates for conventional video cameras (e.g. NTSC 30 fps) have been developed to recognize high-speed phenomena in the real world. Vision chips [1]–[3] execute real-time processing at rates of 1000 fps or more by integrating sensors and processors compactly. Recently, several research groups have attempted to implement high-speed image processing with circuits on an FPGA (Field Programmable Gate Array) board, which is directly connected to a high-speed digital camera head. For example, massively parallel coprocessors were implemented for multitarget tracking [4]; Hough transforms were implemented on an FPGA [5]; and a high-speed vision platform for 1024 × 1024 pixel images was developed to execute real-time image processing algorithms such as color marker tracking, feature point tracking, and optical flow at rates of 1000 fps [6].

Most high-speed robot vision systems can be used as robot sensors for visual feedback control at hundreds of hertz or more; this is because they can output image features at high frame rates by reducing their data sizes through hardware processing. Several studies on visual feedback control have been reported; for instance, high-speed target tracking [7], high-speed grasping [8], and virtual stillness for beating heart surgery [9] have been investigated. Although they are useful as sensors for long-term monitoring of instantaneous high-speed phenomena, they cannot store and output high-frame-rate videos at crucial moments, which promotes human analysis of high-speed phenomena that occur in video logs.

Therefore, we developed a high-speed vision system called IDP Express; this system functions as an intelligent high-speed video logger for long-term monitoring of high-speed phenomena. This is possible because of real-time video processing and high-frame-rate recording. In this paper, we introduce the configuration of IDP Express and verify its effectiveness by showing the results of experiments in which moving objects that undergo sudden shape deformation are mechanically tracked by using two-axis active vision system.

II. Intelligent High-speed Video Logger

Offline high-speed cameras are widely used as video recorders at high frame rates to analyze high-speed phenomena for time spans as short as several seconds due to their insufficient memory. In addition, image features of measured objects that can be used to control camera positions or decide trigger timings cannot be obtained in real time. High-frame-rate video recording is limited to short time spans at a fixed camera position.

Such constraints in the offline high-speed cameras make long-term observation of high-speed phenomena difficult in cases such as (a) observation of an object moving in a wide area (for example, as the beating wings of a flying insect) and (b) long-term observation of unpredictable behaviors such as abnormal behaviors in product machines. In case (a), zooming out to capture all the motions at a fixed camera position degrades the spatial resolution of the recorded video. In case (b), observers are forced to monitor high-speed phenomena over a long time span to manually time the trigger; consequently, they may often miss crucial moments because their visual powers are limited in speed.

In this study, these problems were solved by introducing the concept of an intelligent high-speed video logger, which realizes effective high-frame-rate video recording by automatically selecting only required frames at crucial moments. Such video loggers are required to monitor and analyze high-speed phenomena over long time spans in various applications such as abnormality analysis for assembly tasks, endurance tests for mechanical machines, discovery of sudden cellular behaviors in biology, incident recording for surgery accidents at a clinical site, and impact motion analysis in sports.

High-speed phenomena involving a moving object can be recorded as high-resolution videos at a high frame rate by implementing functions such as (a) a sensor for high-speed visual feedback control, (b) a real-time recognition sensor to judge behaviors at crucial moments, and (c) a video recorder with the same frame rate as the high-speed vision system. By

I. Ishii, T. Tatebe, Q. Gu, Y. Moriue, and T. Takaki are with Hiroshima University, Hiroshima 739-8527, Japan. (corresponding author (I. Ishii) Tel: +81-82-424-7692; e-mail: iishii@robotics.hiroshima-u.ac.jp). K. Tajima is with Photron Inc., Tokyo 102-0071, Japan.

integrating these functions into an intelligent high-speed video logger, we can track a moving object with visual feedback control to zoom in on the object and reduce memory consumption of high-frame-rate videos by selecting only the required frames for crucial moments. Such intelligent high-speed video loggers function as high-frame-rate recording and high-resolution video logging systems for moving objects or machines, even when the target object moves quickly in a wide area.

III. High-speed Vision System, IDP Express

A. Outline of System

IDP Express is a high-speed vision platform for high-frame-rate video processing and recording. IDP Express was designed to implement various types of real-time image processing algorithms and record both images and features at a high speed to a standard personal computer memory. Figure 1 shows the general configuration of IDP Express.

This platform consists of two compactly designed camera heads, a dedicated FPGA image processing board (IDP Express board) with two camera inputs, and a PC. Images captured by the camera heads are transferred to the IDP Express board at a high speed. The IDP Express board executes arbitrary image processing, which is hardware-implemented by the user. The input images and processing results are transferred to a standard PC memory through a PCI-e bus. By executing simultaneous video processing and recording functions at a high frame rate, IDP Express can function as an intelligent high-speed video logger.

B. Components

The camera head of a commercially available high-speed camera (Fastcam MH4-10K; Photron Ltd.) was used in IDP Express. Figure 2 (left) shows the camera head, and Table 1 lists the specifications. Color/gray 8-bit images were transferred at 2000 fps for 512 × 512 pixels and 10000 fps for 512 × 96 pixels with digital serial communication in parallel with 6 pixels. The camera head was also compactly designed for mounting on movable objects; the dimensions and weight were 35 mm × 35 mm × 34 mm and 300 g, respectively.

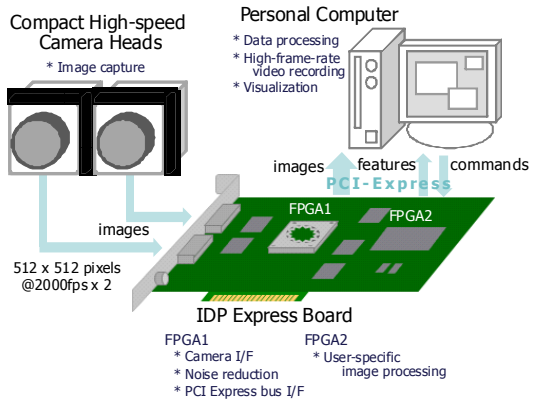


Fig. 1: Configuration of high-speed vision: IDP Express

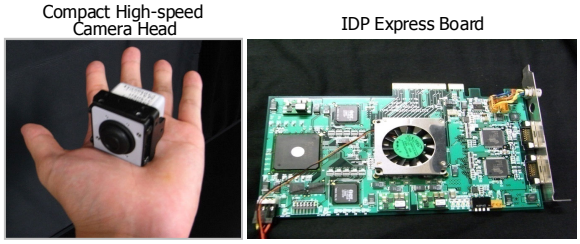


Fig. 2: Photos of IDP Express, the high-speed vision system

The IDP Express board dedicated to IDP Express was designed for high-speed video processing and recording of two 512 × 512 pixel images transferred at rates as high as 2000 fps. Figure 2 (right) shows the board, and Table 2 lists the specifications, and Figure 3 is a function block diagram.

This board consists of an FPGA (hereafter called FPGA1) for camera I/O and PCI-e bus controls, an FPGA (hereafter called FPGA2) for hardware implementation of the algorithms by the user, serial-to-parallel convertors for camera inputs, FIFO (First-In First-Out) memories for data transfer between the FPGAs, and configuration PROMs for the FPGAs. This board has two inputs for the camera heads and connects through 8 lanes of PCI-e buses to a PC; 8-bit images of 512 × 512 pixels can be transferred at 4000 fps. External sync and trigger I/O were also implemented. The path for image data processing and transfer in the IDP Express board is as follows.

(1) Converting image data from serial to parallel

Serial image data transferred from a camera head are converted into 48-bit parallel data (8 bits/pixel) with 6 pixels constituting a block at the camera head's clock speed of 100.8 MHz. These conversions are executed for the two camera heads in parallel.

(2) Data format conversion for image processing

On FPGA1, the 48-bit parallel data for 6 pixels at 100.8 MHz are converted into 40-bit parallel data for 4 pixels at the FPGA2's clock speed of 151.2 MHz. The conversions are executed in parallel for two camera heads, and 96-bits parallel data are converted into 80-bits parallel data for 8 pixels at 151.2 MHz.

Table 1: Specifications of IDP Express camera head

resolution	max. 512 × 512 pixel
bit depth	Color/Gray 8 bit (Bayer)
frame rate	2000 fps (512 × 512 pixel) 10000 fps (512 × 96 pixel)
imager size	5.12 mm × 5.12 mm
pixel size	10 μm × 10 μm
camera mount	NF mount or C mount
interface	digital 8ch
size (W × D × H)	35 mm × 35 mm × 34 mm
weight	0.3 kg

Table 2: Specifications of IDP Express board

board size	112 mm × 180 mm
camera I/F	digital serial 8 ch (×2 camera heads)
FPGA	Xilinx XCVFX60 (FPGA1) Xilinx XC3S5000-4FGG900 (FPGA2)
output	max. 4000 fps (512 × 512 pixel)
memory	Block RAM 1872 kb (FPGA Internal)
I/F	Ext. Trig. IN/OUT, Ext. Sync. IN/OUT
Bus I/F	PCI-e 8 Lanes

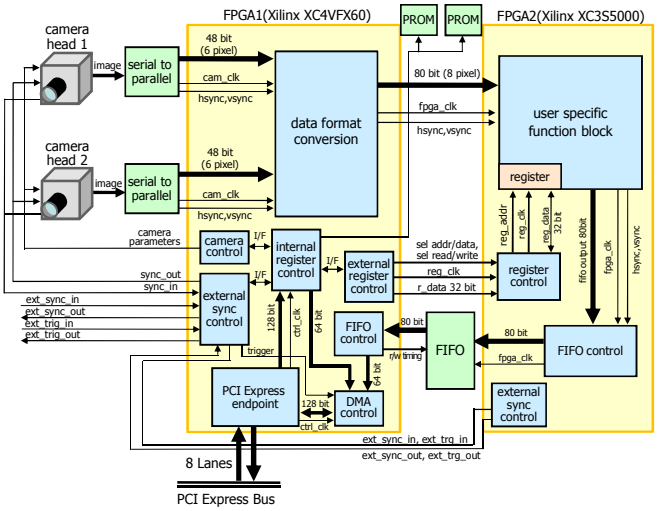


Fig. 3: Function block diagram of IDP Express board

(3) Executing user-specified image processing

The 80-bit image data from FPGA1 are processed in a processing block in FPGA2, where a user-implemented image processing hardware algorithm is executed in real time for the two camera head inputs. The processing block can use register values for parameters in the algorithm; the register values are updated at a clock speed of 50.4 MHz on commands from a PC.

(4) Transferring image data and processing results to the PC

Both the input images and processed results are transmitted through FIFO memories from FPGA2 to FPGA1 in parallel with 80 bits at 151.2 MHz; all data are outputted via a PCI Express endpoint on FPGA1 to a PCI-e bus.

For the PC, a computer with a PCI-e × 8 bus and processor chipset with a DMA (Direct Memory Access) function to transfer data from the PCI-e bus to standard memory such as DRAMs was adopted. By using the DMA function, the input images and processed results from the IDP Express board are memory-mapped onto the allocated memories on the PC at a rate of 4000 images of 512 × 512 pixels or more in a second; these images are overwritten onto the same memories in an endless loop, even when the memory storage is full.

Various API functions associated with board control and data access to memory-mapped data are prepared for a 32-bit Windows XP OS as middleware to develop application programs. With these API functions, only the required pixels or processed results can be accessed in real time; a high-frame-rate video for a crucial moment can be saved by stopping the endless loop for memory-mapping with a trigger determined from the accessed data. A memory of only 4 GB for 7296 frames of 512 × 512 pixel images is presently available as preallocated memory due to limitations of the 32-bit OS.

The CPU and memory specifications can be selected according to the purpose. In this study, we used a PC with the following specifications: ASUSTeK P5E main board; Core 2 Quad Q9300 Bulk CPU; 4 GB memory; Windows XP Professional OS; and 2 × PCI-e 2.0 × 16, 3 × PCI-e × 1, and 2 × PCI × 1 buses I/F.

C. *Implementation*

As an example of hardware implementation on FPGA2 of the IDP Express board, a moment calculation module was implemented in hardware logic to calculate the 0th and 1st moments and output the input images simultaneously. The 0th and 1st moments of a binarized image $B(x,y)$ for an input image $I(x,y)$ are calculated as follows:

$$M_0 = \sum_x \sum_y B(x,y), \quad (1)$$

$$M_x = \sum_x \sum_y xB(x,y), \quad M_y = \sum_x \sum_y yB(x,y). \quad (2)$$

These moment features are useful in determining the size and position of a measured object in an image; their application to high-speed target tracking with high-frame-rate video recording are described in the following sections.

Figures 4 and 5 show the block diagram for a moment calculation module and the timing diagram for its control signals, respectively; the module deals with the parallel input of 8-bit images in units of 8 pixels from the two camera heads. Here, a 512 × 512 pixel image is raster-scanned to the module in units of 4 pixels from the upper left to lower right using X and Y address signals with a 151.2 MHz clock, as shown in Figure 6 (a). In the module, 8 pixels are transmitted at a time, and two moment calculation circuits are prepared in parallel for the two camera heads.

The moment calculation circuit is composed of gray-to-binary conversion circuits, 4-pixel adders, moment accumulators, and a data selector for FIFO output.

By using four gray-to-binary conversion circuits, $B(x,y)$ is obtained as a binarized image for the input image $I(x,y)$; the 4-pixel adders calculate the added values ${}^4M_0$, ${}^4M_x$ and ${}^4M_y$ for $B(x,y)$, $xB(x,y)$ and $yB(x,y)$ for 4 pixels. In the accumulators, the moment features M_0, M_x and M_y for the whole area are calculated by adding the summed values for ${}^4M_0$, ${}^4M_x$ and ${}^4M_y$, sequentially. The four input adders and moment accumulators were implemented in parallel for all three moment features. In the data selector, the output data for FIFO are selected with X and Y address signals from the input images $I(x,y)$ and moment features M_0, M_x and M_y, as shown in Figure 6(b). Input images are outputted to FIFO in the same manner as the input image format except for the three blocks in the lower right of the last 12 pixels. Moment features are calculated for 512 × 512 pixel images, except for the last 8 pixels in the lower right, and outputted into the three last blocks by addition to the input images.

The delay time in calculating the moment features is 6 clocks (1 clock = 6.6 ns) after raster-scanning all pixels in an input image: one clock for gray to binary conversion; three for the 4-pixels adder; one for moment accumulators; and one for data selectors. Table 4 lists the resource consumption of FPGA2 on the IDP Express board when the moment calculation modules for the two camera heads are implemented on FPGA2. We confirmed that all input images and moment features from the two camera heads can be outputted to the PC at 2000 fps.

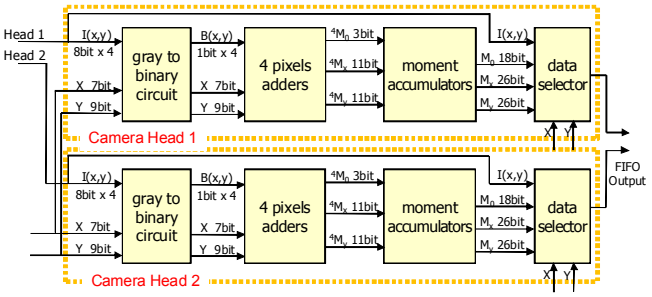


Fig. 4: Schematic data flow of moment calculation module

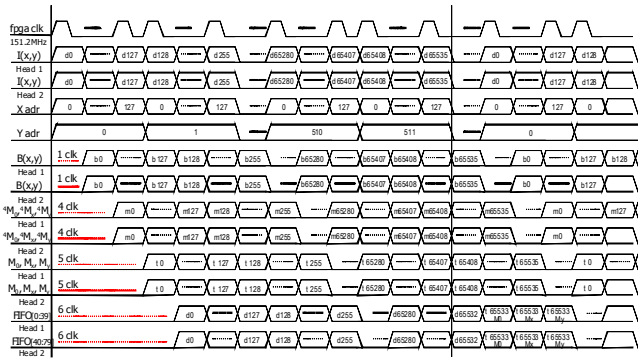


Fig. 5: Timing chart of moment calculation module

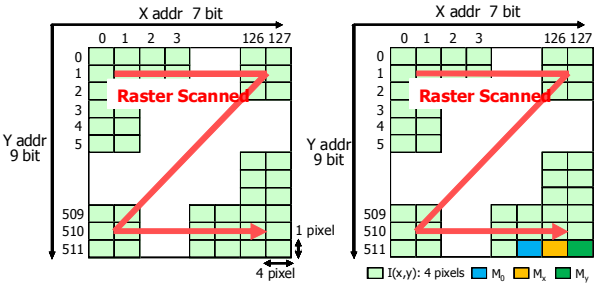


Fig. 6: Image input format and output data format

TABLE 4: RESOURCE CONSUMPTION OF USER-SPECIFIED FPGA

Device Type	Xilinx XC3S5000-4FGG900
Slice	767/33,280 (2%)
Slice Flip Flop	1,179/66,560 (1%)
4-input LUT	897/66,560 (1%)
Bounded IOB	250/633 (39%)
Block RAM	5/104 (4%)
GCLK	2/8 (25%)

IV. HIGH-SPEED TARGET TRACKING SYSTEM WITH HIGH-FRAME-RATE VIDEO RECORDING

A. Configuration

To demonstrate the effectiveness of IDP Express as an intelligent high-speed video logger, a high-speed target tracking system was developed. Figure 7 shows its configuration. The target tracking system consists of two 2-DOF (Degrees Of Freedom) active visions and IDP Express. The active vision system is moved by pan and tilt motors, which are the compact and high-speed motors RSF-5A-50 (Harmonic Drive Systems Inc.). The sizes of the active vision system are 12 cm × 12 cm × 7 cm without a camera head. The camera heads

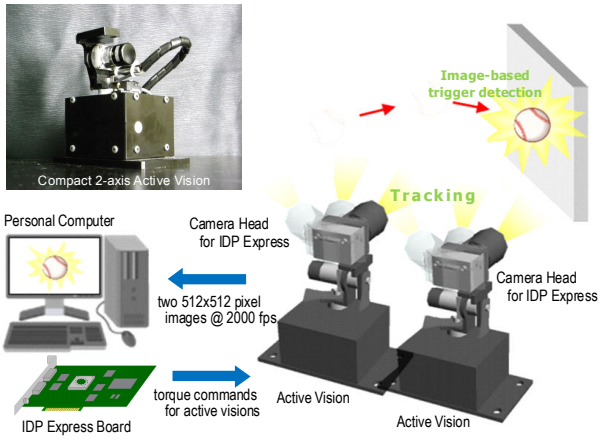


Fig. 7: Target tracking system with high-frame-rate video recording

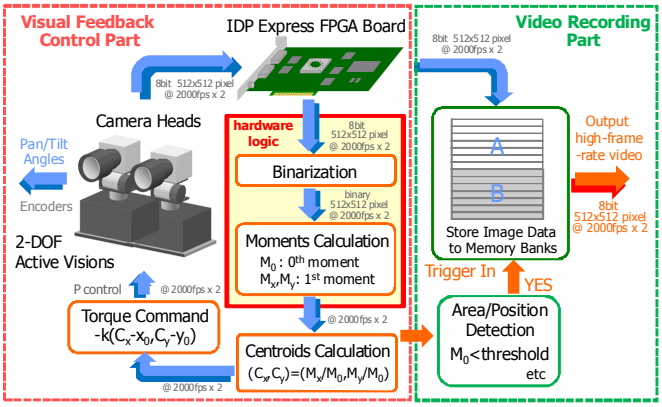


Fig. 8: Flowchart of tracking control/video recording algorithm

mounted on both active vision systems are controlled by using high-frame-rate visual information from the left and right cameras in order to track a moving object.

B. Algorithm

The target tracking system is an intelligent high-speed video logger that records high-resolution videos at high frame rates to monitor crucial moments of moving objects, which were always zoomed in the camera views. Figure 8 shows the flow chart of the implemented algorithm. The algorithm consists of (a) a visual feedback control part for tracking objects at the center of the camera views and (b) a high-frame-rate video recording part with image-based triggers.

In the visual feedback control part, the 0th and 1st moment features are used to calculate (1) the image centroid $(C_x, C_y) = (M_x/M_0, M_y/M_0)$ as the object's position for tracking control, and (2) the summation M_0 as its area to detect sudden changes in images. These features were calculated for 512 × 512 pixels images from the two camera heads at 2000 fps by using the hardware-implemented moment calculation module on the IDP Express board. Here, the pan and tilt motors on the two active visions are controlled at a feedback rate of 2000 Hz to correspond to the calculated image centroids for the centers of the camera views.

In the video recording part, the input images and calculated moment features are memory-mapped at 2000 fps in the

preallocated PC memory on an endless loop unless there is a trigger. Image-based triggers to stop the endless loop are automatically determined by using the calculated moment features and encoder values for tilt or pan angles of the active visions from the camera heads: e.g., an image-based trigger for a measured object's size is determined by judging whether the 0th moment M_0 is lower than a threshold, or an image-based trigger for the direction of the motion of the object is determined by checking the angular velocities.

Once an image-based trigger is generated, only the frames around the trigger timing are selected and outputted to external systems as a high-frame-rate video, which stops the overwriting of the endless loop in the current memory bank. By applying the same process to the other memory banks, high-frame-rate videos can be continually processed and recorded. We also confirmed that target tracking and video recording of 512 × 512 pixel images for the two camera heads could be simultaneously executed at 2000 fps.

V. Experiments

We show the experimental results for automatic high-frame-rate video recording at crucial moments when the measured objects are being tracked for zooming in by using the target tracking system. In the experiments, the two active vision systems were used in parallel and located at a distance of 40.6 cm from each other; they had identical lenses. The measuring area for each camera was 10 cm × 10 cm at 50 cm in front of their lenses; the threshold for binarization was set to 40; and the center of the camera view was set to x = 256 pixels, y = 256 pixels.

A. Crash of a Falling Custard Pudding

When the custard pudding was free-falling, the high-frame-rate videos were automatically recorded just during the time of its crash onto a table. Figure 9 shows the experimental environment. The custard pudding was 4 cm in height and 6 cm in diameter; it was allowed to fall from 30 cm above the table. The target tracking system was set 57 cm ahead of the free-falling pudding, which was tracked in the centers of the camera views. An image-based trigger was generated by judging the crash as when the angular velocity of tilt on the right active vision was inverted.

Figure 10 shows the 3-D position of the pudding, the image centroids, the angle of pan and tilt, the angular velocity of tilt, and the 0th moment for the active vision on the right over 2 s, which was the time of the crash. t = 0 s was the start of the observation time. The pudding began to fall at t = 19.8500 s, and an image-based trigger was generated at t = 20.1795 s. The 3-D position was calculated by triangulation by using image centroids and encoder values for the tilt and pan angles on both the active vision systems. Figure 11 shows a pair of 8-image sequences from the left and right cameras taken at intervals of 10 ms. Here, these images were selected from 384 frames of 512 × 512 pixel images recorded at 2000 fps, which were center-triggered with the image-based trigger.

By tracking the free-falling pudding in order to zoom in on the target, the sudden shape deformation occurring in time spans on the order of subseconds was automatically recorded as high-frame-rate videos at high resolution. Different views

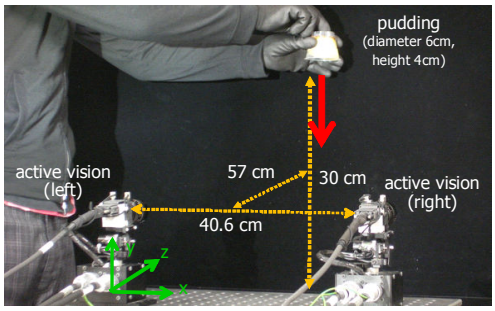


Fig. 9: Experimental environment to observe crash of pudding

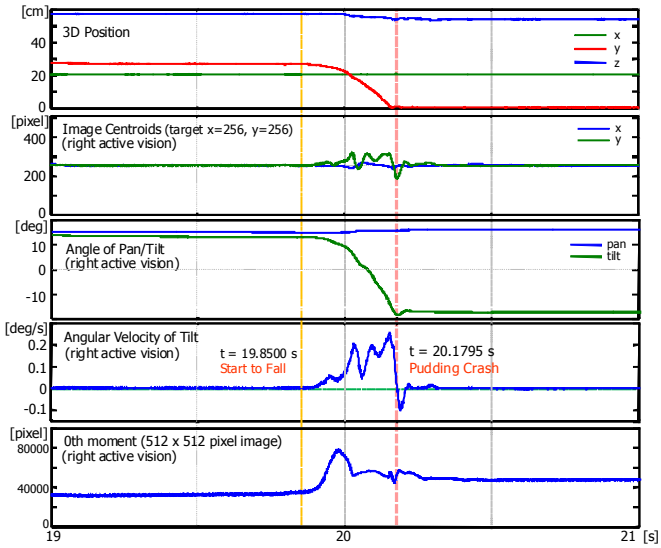
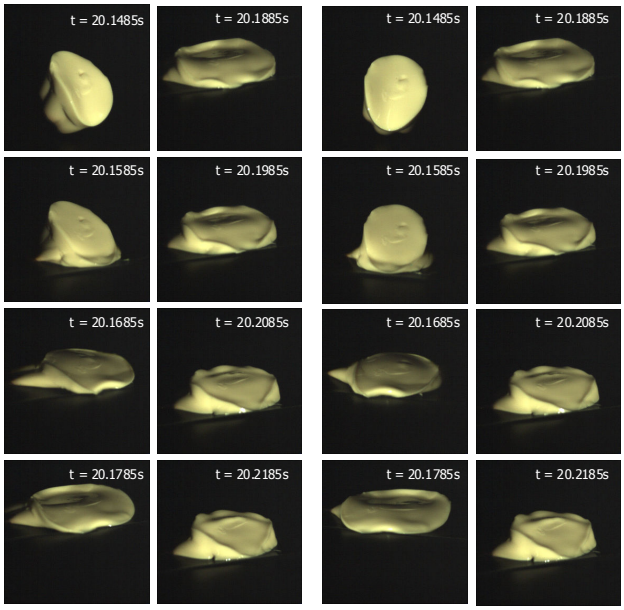


Fig. 10: 3D position, image centroids, pan/tilt angles, angular velocity of tilt, and 0th moment on the right active vision (crash of pudding)



(a) left camera (b) right camera

Fig. 11: High–frame-rate image sequence of the crash of pudding

were recorded from the left and right cameras at a rate too high for human vision. The pudding was always observed at a width of 300 pixels or more on the target tracking system, while it was observed at a width of 100 pixels with a 512 × 512 pixel camera located at a fixed position.

B. *Rotating Balloon's Explosion*

When a rotating balloon was tracked in the centers of the camera views, high-frame-rate videos were recorded just the time of its explosion. Figure 12 shows the experimental environment. The balloon was 16 cm in size, and it was rotated at 3.5 rps in an orbit of radius 15 cm. The target tracking system was set in front of the balloon 114 cm away from it. An image-based trigger was generated by judging that there was an explosion when the 0th moment on the right camera was less than 80000 pixels.

Figure 13 shows the 3-D position of the balloon, the pan and tilt angles, and the image centroids and 0th moment on the right active vision over 2 s, when the explosion occurred. Here, the image-based trigger was generated at $t = 23.6425$ s. Figure 14 shows a pair of 8-image sequences from the left and right cameras at intervals of 0.5 ms. Even when the balloon was rotating quickly, its sudden explosion within time spans on the order of subseconds was automatically recorded as high-frame-rate videos from different views at high resolution as if there was no rotation.

These experimental results indicate that IDP Express can execute both real-time video processing for target tracking and high-frame-rate video recording simultaneously; the system can be applied as an intelligent high-speed video logger for objects moving quickly in a wide area.

VI. Conclusion

This paper introduced the high-speed vision system IDP Express, which is capable of both processing and recording 512 × 512 pixel images at 2000 fps in real time for two cameras; it functions as an intelligent high-speed video logger. We verified its effectiveness by conducting experiments on moving objects that undergo sudden shape deformation; a stereo high-speed target tracking system was used in the experiments. On the basis of the results of these experiments, we will improve the image-based trigger function to detect crucial moments in more complicated real world and extend the system applications for factory automation, robot control, and biomedical fields as an automatic video logging system for monitoring high-speed phenomena of moving objects.

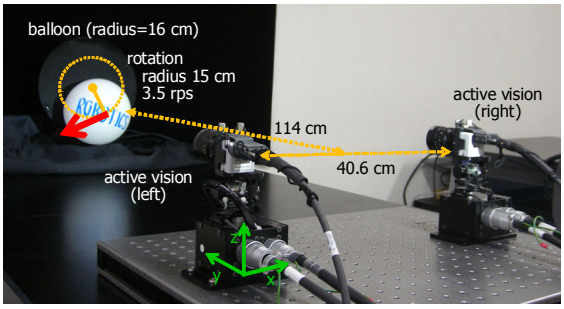


Fig. 12: Experimental environment to observe explosion of a balloon

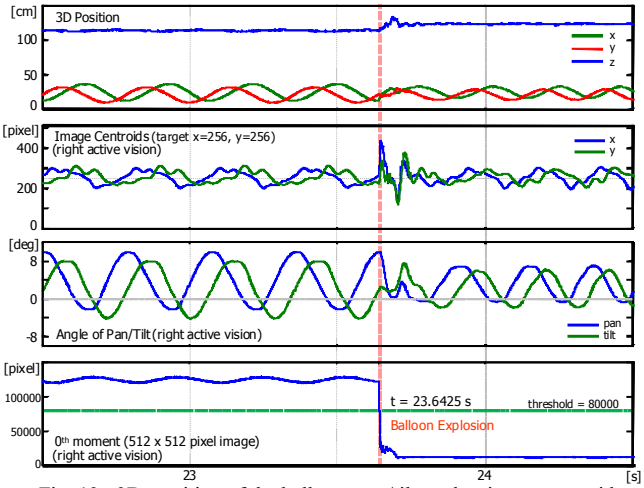
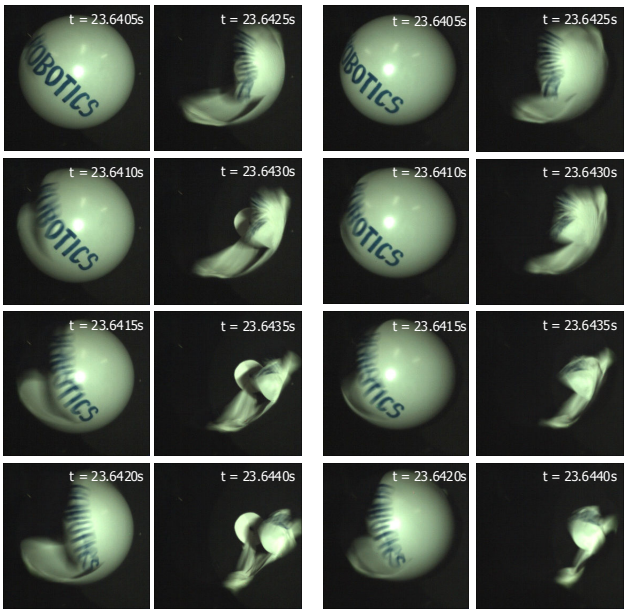


Fig. 13: 3D position of the balloon, pan/tilt angles, image centroids, and 0th moment for the right active vision (balloon explosion)



(a) left camera (b) right camera

Fig. 14: High-frame-rate image sequence of the explosion of the balloon

References

[1] T.M. Bernard et al, "A Programmable Artificial Retina," *IEEE J. Solid-State Circuits*, 28(7), pp. 789–797, 1993.

[2] J. Eklund et al, "VLSI Implementation of a Focal Plane Image Processor - A Realization of the Near-sensor Image Processing Concept," *IEEE Trans. VLSI Systems*, 4(3), pp. 322–335, 1996.

[3] T. Komuro et al, "A Dynamically Reconfigurable SIMD Processor for a Vision Chip," *IEEE J. Solid-State Circuits*, 39(1), pp. 265–268, 2004.

[4] Y. Watanabe et al, "955-Fps Real-Time Shape Measurement of a Moving/Deforming Object Using High-Speed Vision for Numerous-Point Analysis," *Proc. ICRA*, pp. 3192–3197, 2007.

[5] S. Hirai et al, "Realtime FPGA-Based Vision System," *J. Robotics and Mechatronics*, 17(4), pp. 401–409, 2005.

[6] I. Ishii et al, "Development of High-speed and Real-time Vision Platform, H^3 Vision," *Proc. IROS*, pp. 3671–3678, 2009.

[7] I. Ishii et al, "Target Tracking Algorithm for 1ms Visual Feedback System Using Massively Parallel Processing," *Proc. ICRA*, pp. 2309–2314, 1996.

[8] A. Namiki et al, "Development of a High-speed Multifingered Hand System and Its Application to Catching," *Proc. IROS*, pp. 2666–2671, 2003.

[9] Y. Nakamura et al, "Heartbeat Synchronization for Robotic Cardiac Surgery," *Proc. ICRA*, pp. 2014–2019, 2001.