International Journal of Automation and Computing 13(4), August 2016, 305-318
DOI: 10.1007/s11633-016-1024-0

Review of Some Advances and Applications in Real-time High-speed Vision: Our Views and Experiences

Qing-Yi Gu Idaku Ishii

Robotics Laboratory, Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Hiroshima 738-8527, Japan

Abstract: The frame rate of conventional vision systems is restricted to the video signal formats (e.g., NTSC 30 fps and PAL 25 fps) that are designed on the basis of the characteristics of the human eye, which implies that the processing speed of these systems is limited to the recognition speed of the human eye. However, there is a strong demand for real-time high-speed vision sensors in many application fields, such as factory automation, biomedicine, and robotics, where high-speed operations are carried out. These high-speed operations can be tracked and inspected by using high-speed vision systems with intelligent sensors that work at hundreds of Hertz or more, especially when the operation is difficult to observe with the human eye. This paper reviews advances in developing real-time high speed vision systems and their applications in various fields, such as intelligent logging systems, vibration dynamics sensing, vision-based mechanical control, three-dimensional measurement/automated visual inspection, vision-based human interface, and biomedical applications.

Keywords: Real-time high-speed vision, target tracking, abnormal behavior detection, behavior mining, vibration analysis, 3D shape measurement, cell sorting.

1 Introduction

The human eye and its brain interface, the human visual system, can process 10 to 12 separate images per second, by perceiving them individually[1]. The time resolution of the human visual system is low, compared with the touch and auditory senses of humans. The human auditory system can perceive sounds that are between hundreds of Hertz and several thousand Hertz, which are impossible to perceive by the human visual system. In the field of robotics, conventional vision systems designed for human eyes are used as visual sensors to perceive spatial information. The frame rate of a conventional vision system is sufficient for a humanoid robot. However, if the frame rate of a vision system can be improved to the level of sound, we may be able to develop various hyper-human applications.

Recently, in order to overcome the restrictions posed by conventional video signals, many offline high-speed vision systems that can operate at high frame rates of 1 000 fps or more have been developed. Such offline high-speed camera systems are used as tools for the analysis of high-speed phenomena, such as sports motion analysis, auto crash testing, and motion modeling for industrial machinery. In aerospace, life science, and automotive applications, it is often necessary to capture sequences of high-speed images for analysis. In automotive applications, the dynamics of airbag deployment are regularly captured with high-speed cameras. In these applications, high-speed phenomena are replayed in slow motion for human eyes and processed by using offline programs. However, these systems are not real-time vision sensing systems that can process all image frames captured by the system in real time. Here, "real-time" refers to the process of image frames obtained from the image sensor being processed immediately (in a few microseconds). In addition, image features extracted from the image frames can be transferred to a control system with very little delay (of millisecond level) for rapid pattern recognition, fast feedback control, etc. Therefore, high-frame-rate (HFR) real-time vision-based feedback control is difficult to achieve by using an offline high-speed vision system. To overcome the restrictions of offline high-speed vision systems, several real-time high-speed vision systems have been developed recently, which can be used as real-time vision sensors working at several thousands of Hertz. In such systems, image processing algorithms are accelerated by using hardware circuits. Further, real-time high-speed vision feedback control can be achieved at HFR. In Section 2, the recently developed real-time high-speed vision systems and integrated algorithms are discussed.

During the past two decades, many visual feedback control systems[2, 3] have already been realized based on conventional vision systems. Recently, the demand for a real-time high-speed vision sensor has increased rapidly in many application fields where high-speed operations are carried out and high-speed visual feedback control is required. With the use of a real-time high-speed vision system, high-speed phenomena can be observed and analyzed in real time. A highly accurate speed and acceleration distribution can be calculated in real time by using the small time difference in high-speed vision. The dynamics of an object, such as force distribution and mechanical impedance, can

Review
Manuscript received July 27, 2015; accepted March 2, 2016
Recommended by Associate Editor Jangmyung Lee

be measured by using high-speed vision. For example, in the equation of motion, $F = ma$, if the mass of an object is known, the acceleration of the object can be calculated by using high-speed image pairs. In this case, a real-time high-speed vision system can be used as a force sensor. Compared with conventional dynamic sensor technology, this system is almost a revolution brought by using high-speed vision as a sound sensor, an acceleration sensor, etc. It can be used as a dynamic sensor to detect the displacement and vibration distribution in the time and space domains. Further, such real-time high-speed vision technology has a huge application potential in fields such as biological dynamics, structure dynamics, and material dynamics. In Section 3, several hyper-human applications using a real-time high-speed vision system are reviewed.

High-speed vision systems have two properties: 1) small displacement between frames, and 2) complex timing and limited resources for fast image processing. To realize high-speed vision-based feedback control, a huge number of image frames need to be processed in real time. Unfortunately, the currently used image processing algorithms including object tracking and recognition have been designed for conventional video signals, which work at several dozen frames per second. The key idea of real-time high-speed vision-based applications is to utilize the small displacement between frames for optimizing or simplifying image processing algorithms. After optimization, appropriate development environments are selected for the implementation of the improved algorithm by considering the application purpose. An algorithm with high parallelism, low complexity, and low memory consumption is suitable for field-programmable gate arrays (FPGAs). The drawbacks are high development time and high cost. Further, an algorithm with high parallelism, medium complexity, and high memory consumption is suitable for graphics processing units (GPUs). However, the processing speed is slower than that in the case of FPGAs. A complex algorithm with high memory consumption algorithm is suitable for personal computers (PCs). Its advantage is the fast development speed, and its drawback is an execution speed lower than that of FPGAs and GPUs. In an actual implementation, an image processing algorithm is broken down into several sub-modules and implemented on different development platforms according to their properties.

Several review papers[4, 5] on high-speed vision have been published recently. These review papers focus on high-speed vision architecture and its applications in robot control, dynamic manipulation and handling, human-machine interface, interactive display, micro visual feedback control, etc. In this paper, based on our experience in the field of real-time high-speed vision, we focus on the design of high performance algorithms and their implementations on FPGAs according to the properties of high-speed vision. Additionally, we illustrate several novel real-time applications by utilizing the properties of high-speed vision, which include vibration dynamics sensing, biomedical applications, etc. We expect that with this article, the readers will be able to generate new ideas and algorithms in the field of real-time high-speed vision.

2 Real-time high-speed vision systems

2.1 Vision systems

High-speed imaging and processing are both important for real-time vision-based feedback control and the analysis of high-speed phenomena. Owning to the technological developments in the field of electronics, various real-time high-speed vision systems have been developed recently, which can work at around 1 kHz.

Vision chips[6−12] are one-chip vision systems with resolutions no more than tens of thousands of pixels for HFR video processing that uses integrated sensors and processors on a compact die. Vision chips have a general-purpose massively parallel single instruction, multiple data (SIMD) processor. This is considered to lead to disadvantages in terms of scalability of the system functions related to module exchange and coordination with other heterogeneous processors. However, the processing operations are very simple, e.g., it calculates the centroid of a signal target because its processor architecture is limited in terms of the type of image processing.

FPGA-based high-speed vision platforms have been developed for the hardware implementation of various types of image processing algorithms such as a gravity extraction function on an FPGA[13], a high-speed Hough transform processor on an FPGA[14], massively parallel co-processors for multi-target tracking[15], and a high-speed vision platform for real-time video processing of $1\,024\times1\,024$ images at 1 000 fps[16]. More complicated image processing algorithms can be implemented in FPGA-based vision systems, e.g., multi-object feature extraction[17], higher-order local autocorrelation (HLAC) features[17, 18], Harris corner detection[19], and color histograms[20, 21]. The drawback is that only low memory consumption and non-iterative algorithms are suitable for such FPGA-based vision systems. There are two reasons: 1) The number of embedded block random access memories (BRAMs) in an FPGA is limited, and it is slightly difficult to store image data of 1 MB or more without using any external memory with the present integrated FPGA technology. Thus, image processing algorithms with a high memory consumption are difficult to accelerate by using such vision systems. 2) FPGA-based high-speed vision systems work at 1 000 fps or more. This implies that image processing algorithms should be executed immediately after receiving parallel image pixel data from an image sensor. There are no blank times for iterative algorithms.

A common problem of standalone real-time high-speed vision systems[14, 22] is the limited on-board memory for caching image frames. Thus, there are limited applications of standalone real-time high-speed vision systems as these systems cannot work as long-term HFR video logging systems. Recently, a PC-based real-time HFR vision system that operates at 1 000 fps or more has been developed. It can execute an arbitrary image processing algorithm on a dedicated FPGA board and transfer raw image and processed image features to a PC directly via the PCI-e bus (IDP Express)[23]. Fig. 1 shows the configuration and photographs 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. Color/Gray 8-bit images captured by the camera heads were transferred to the PC at 2 000 fps for 512×512 pixels and 10 000 fps for 512×96 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. Further, sound-level vision-based feedback control could be realized by using an IDP Express system and executing simultaneous video processing on an FPGA, GPU and PC.

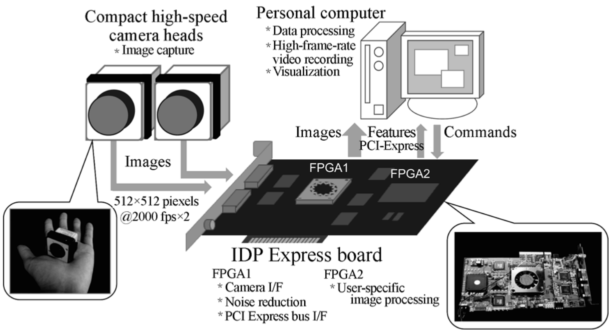


Fig. 1 High-speed vision system: IDP Express

2.2 Integrated algorithms

In recent years, researchers have conducted many trials for developing various types of integrated algorithms on the FPGAs embedded in a high-speed vision system. The basic idea of such research is to simplify the conventional image processing algorithms by considering the hardware limitation and the small displacement between frames. By using high-speed vision, we can assume that the movement of an object is a linear motion because the displacement between frames is very small. By utilizing this property, we can simplify many image processing algorithms. Meanwhile, algorithm design for hardware implementation should consider the consumption of hardware resources in FPGAs; e.g., BRAMs and multiples are limited in an FPGA.

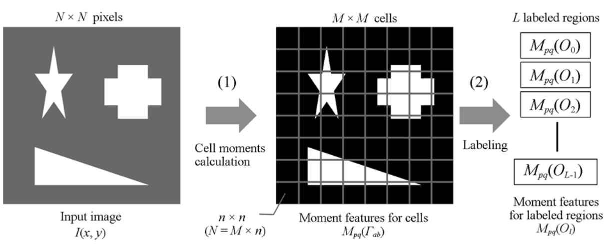


Fig. 2 Concept of a cell-based labeling algorithm

Multi-object extraction is a fundamental operation in image-based target tracking and pattern recognition. However, conventional multi-object extraction methods are not always suitable for hardware implementation because most of them require a large memory area, depending on the image size and the iterative processes that are applied to all the pixels in an image. This makes it difficult to accelerate multi-object extraction by using a high-speed vision platform based on the hardware implementation. Watanabe et al.[24] proposed a labeling algorithm using binary search. This algorithm was specially designed for a unique vision chip for high frame rate multi-object extraction and tracking. However, only simple functions such as the extraction of an object centroid can be implemented using such architecture. A moment-based analysis of numerous objects[15] has been reported. It can extract the 0th, 1st and 2nd moment features from 1 024 objects in a 256×256 pixel image at 955 fps by using FPGAs as co-processors. However, the latency to obtain moment features is 4 ms, and this is mainly caused due to the caching of the input image in random access memory (RAM). Moreover, an additional merge processing is required in the PC after transferring the moment features of the blocks from the FPGA boards to the PC. To overcome these problems, a cell-based labeling method[25] was reported. It can accelerate multi-object recognition to extract the locations and features of multiple objects in an image, and multi-object feature extraction is performed at 2 000 fps for 512×512 images by implementing the cell-based labeling algorithm as the hardware logic on a high-speed vision platform[17]. In this system, images are immediately processed on an FPGA after receiving pixel data from the image sensor. The latency of this system is only 1 ms. The concept of cell-based labeling can reduce the number of scanned pixels for labeling and memory size to store label equivalences without accuracy degradation in the space resolution by dividing an image into sub-image regions as cells and assuming the additivity in the image feature calculation, as shown in Fig. 2. The cell-based labeling algorithm has two sub-processes: 1) cell-level image feature calculation for the divided cells, and 2) object-level labeling and updating of cell-based image features. In the cell-based labeling algorithm, the computational complexity and memory consumption of the labeling process can be reduced by exchanging the computational sequence in the labeling process of the divided cells after calculating the moments for cells of $n \times n$ pixels on the basis of the additivity of the moment calculation. The results labeled with the cell-based labeling algorithm are not perfectly matched with those labeled with pixel-based connected component labeling algorithms when $n > 1$, because there are cases in which pixels belonging to different connected components in the same cell or neighboring cells are identified as the same regions on the basis of the cell connectivity. Although we have to consider the trade-off relationship between its computational complexity and its equivalence to pixel-based connected component labeling, the computational complexity and memory consumption of the labeling process in the cell-based labeling algorithm can be reduced to the cell-level complexity of the order of $O(M^2)$ when $n > 1$, i.e., $\frac{1}{n^2}$ of the pixel-level complexity of the order of $O(N^2)$ in pixel-based connected components labeling algorithms. Thus, the cell-based labeling algorithm is suitable for hardware implementation on a high-speed vision platform to extract multiple objects in a binary image at a relatively high frame rate. In the implementation, the additive image features of 1 024 objects in an image can be simultaneously extracted for multi-object recognition and tracking by dividing the image into 8×8 cells concurrently with the calculation of the 0th-order and 1st-order moments to obtain the sizes and locations of multiple objects. Several types of object features have been extracted by using the hardware implementation of cell-based labeling. For example, higher-order local autocorrelation (HLAC) features were extracted for shift-invariant object recognition[17]. The schematic data flow of the implemented circuit for 25 HLACs is shown in Fig. 3.

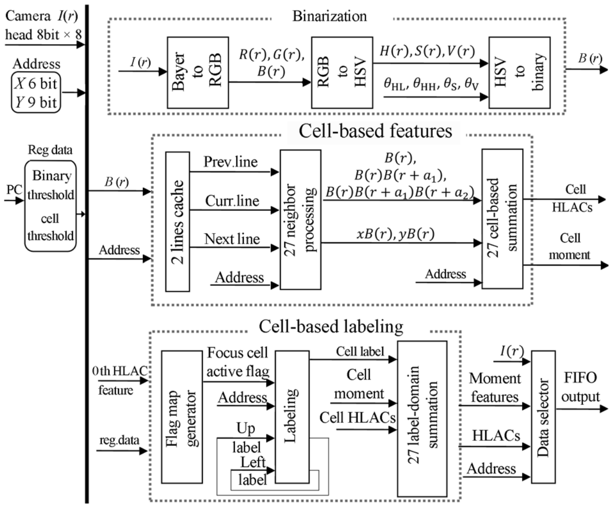


Fig. 3 Schematic data flow of the implemented circuit

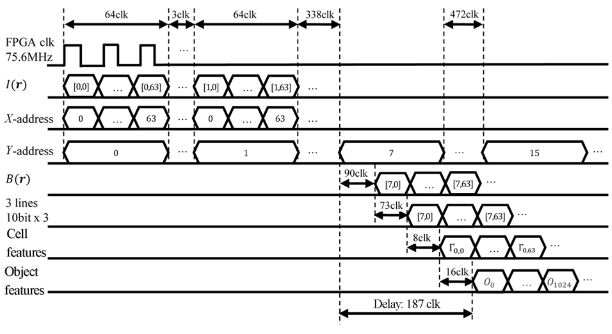


Fig. 4 Timing chart of the implemented circuit

The timing chart of the implemented circuit is shown in Fig. 4. The 512×512 input images are scanned in units of eight pixels from the upper left to the lower right by using X and Y address signals with a 75.6 MHz clock. Immediately after all the pixels in a cell are scanned, the submodules (binarization, cell-based features, and cell-based labeling submodules) output 25 cell-based HLACs and the two 1st-order moments with a delay time of 187 clocks (=2.47 μs) for a color image in parallel. The hardware circuits have already been implemented in a commercial FPGA (Xilinx XC3S5000-4FG900) embedded in IDP Express. This FPGA was designed for high-volume and cost-sensitive consumer electronic applications and has a block RAM of 239 KB. The BRAM resource of the cheap FPGA was almost exhausted (87 %) in this hardware implementation.

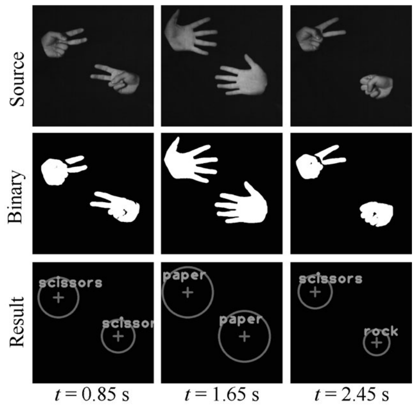


Fig. 5 Recognition of quick hand gestures in the RPS game

System performance during the simultaneous execution of multi-object feature extraction at 2 000 fps for 512×512 images has already been verified by showing several experimental results for high-speed moving objects, as shown in Fig. 5. We conducted an experiment on real-time gesture recognition by using 25 HLACs: The shapes of two human hands were recognized in the rock-paper-scissors (RPS) game. In the experiment, two persons played the RPS game with three battle times of more than 3 s. It can be observed that the hand shapes of the two persons were correctly classified. The authors also implemented six parallel cell-based labeling circuits for multiple pie-shaped marker tracking[26].

A high-speed vision system that can be applied to color-histogram-based tracking at 2 000 fps by hardware implementation of an improved CamShift algorithm was reported[20]. In the improved CamShift algorithm, the size, position, and orientation of the color-patterned object to be tracked in an image can be simultaneously extracted by using only the hardware implementation of a color histogram circuit module for calculating the moment features of binary images quantized by 16 hue-based color bins. The search area of the improved CamShift algorithm is limited in a small region by considering the first property of high-speed vision. By the hardware implementation of the color histogram circuit modules on a high-speed vision platform, IDP Express, the improved CamShift algorithm enables color histogram-based tracking at 2 000 fps for 512×511 pixel images in real time. By installing the tracking system on a fast robot arm, the authors demonstrate the effectiveness of 2 000 fps color-histogram-based tracking by performing several experiments of color-patterned objects, which are always tracked in the camera views even when they move rapidly in complicated backgrounds. Fig. 6 shows the three-dimensional (3D) color-histogram-based tracking of fast human motion with a complex background.

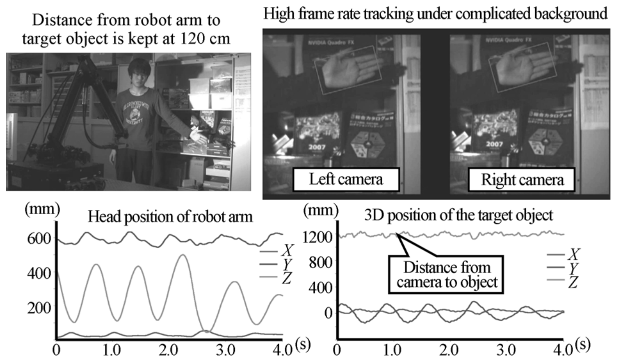


Fig. 6 3D color-histogram-based tracking of fast human motion

A multi-object color-histogram-based tracking system was implemented[21] by introducing the cell-based labeling algorithm[25].

Further, an HFR vision system that can estimate the optical flow in real time at 1 000 fps for 1 024×1 024 pixel images via the hardware implementation of an improved optical flow detection algorithm on a high-speed vision platform was developed[27]. Based on the Lucas-Kanade method, the authors adopted an improved gradient-based algorithm that can adaptively select a pseudo-variable frame rate according to the amplitude of the estimated optical flow to accurately detect the optical flow for objects moving at high and low speeds in the same image. In the improved algorithm, the product sums of only the spatial partial derivatives (S_{xx}, S_{xy} and S_{yy}) are calculated in the same manner as that in the Lucas-Kanade method, whereas the product sums of the spatial and temporal partial derivatives (S_{xt} and S_{yt}) are calculated by accumulating the product sums in a variable interval n, which is automatically adjusted according to the speed of the objects being observed, n is small for objects moving rapidly and large for objects moving slowly. Ishii et al.[16] implemented the improved algorithm on a high-speed vision platform H^3 Vision and realized 1 000 fps real-time optical flow detection for gray-level 1 024×1 024 pixel images. Further, because of the limitation of the hardware resources in the FPGA embedded on H^3 Vision, a full-pixel optical flow estimation is difficult to implement by hardware circuits. In this study, the authors divided a 1 024×1 024 pixel image into block regions of 32×32 pixels and calculated the optical flows at 1 000 fps for all the 1 024 divided blocks.

3 Hyper-human applications

In order to establish hyper-human robotics technology, which is a technology that far exceeds human abilities, we are principally promoting the research and development of sensing and manipulation technologies based on high-speed vision that has real-time processing rates of 1 000 fps or more. In order to accomplish our mission of producing something that will be beneficial for people in our society, we will extend these technologies to applications in a wide variety of fields. Based on our experience in the field of real-time high-speed vision, the following paragraphs of this section introduce applications in different fields, as shown in Fig. 7. In this section, we introduce several applications, which are designed by utilizing the properties of high-speed vision. Some of those applications are accelerated by cell-based labeling algorithm[17] for fast multi-object feature extraction, such as dynamics-based visual inspection systems[28−30], blink-spot projection method[31], and fast cell analysis systems[32, 33].

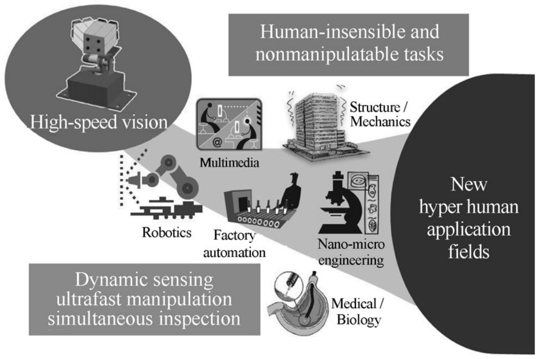
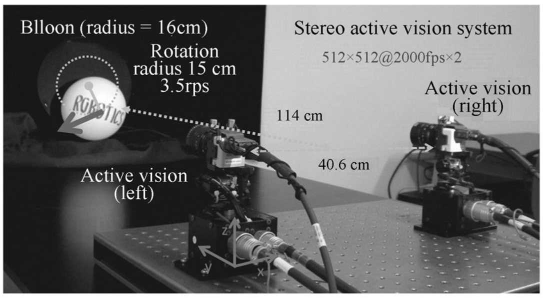


Fig. 7 Hyper-human applications

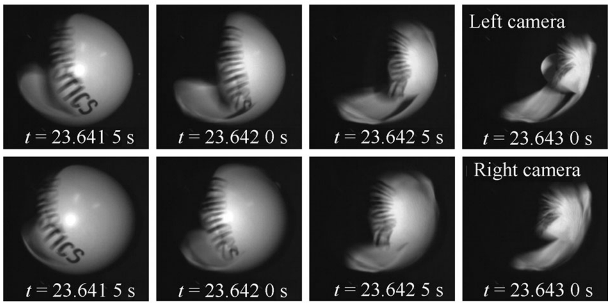
3.1 Intelligent logging system

High-speed cameras can be used as HFR video recording systems for analyzing high-speed phenomena that occur within a short time span. However, it is difficult for conventional offline high-speed cameras to carry out long-term observations of unpredictable high-speed phenomena, because most of them do not have an image-based trigger function to determine the trigger timing in real time with automatic image processing. However, there are strong demands for high-speed long-term unpredictable abnormal behavior detection and logging systems.

An HFR object tracking system with video recording function was reported[23]. It can record high-speed phenomena of moving objects by using mechanical pan-tilt tracking. This target tracking system consists of two 2-degree of freedom (DOF) active visions and IDP Express, as showed in Fig. 8 (a). The pan and tilt motors on the two active visions are controlled at a feedback rate of 2 000 Hz to correspond to the calculated image centroids for the centers of the camera views. The centroid position of the target object is calculated for 512×512 pixel images from the two camera heads at 2 000 fps by using the hardware-implemented moment calculation module on the IDP Express board. In the video recording part, the input images and the calculated moment features are memory-mapped at 2 000 fps in the preallocated PC memory in an infinite loop unless there is a trigger. Image-based triggers to stop the infinite loop are automatically determined by using the calculated moment features and encoder values for the tilt or pan angles of the active visions from the camera heads. Once an image-based trigger is generated, only the frames around the trigger timing are selected and output to external systems such as an HFR video, which stops the overwriting of the infinite loop in the current memory bank. By applying the same process to the other memory banks, we can continually process and record HFR videos. For instance, consider a balloon explosion experiment in front of the two active vision systems. 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 at a distance of 114 cm. An image-based trigger was generated by judging that there was an explosion when the 0th-order moment on the right camera was less than 80 000 pixels. Fig. 8 (b) shows the HFR image sequence of the explosion of the balloon.



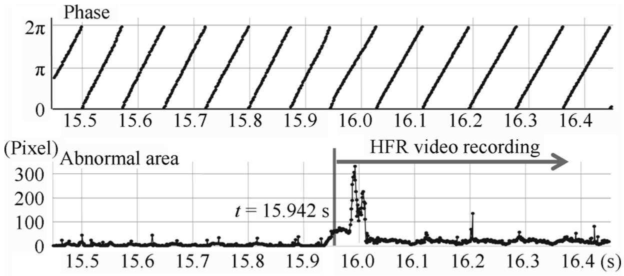
(a) Overview of tracking system for HFR recording



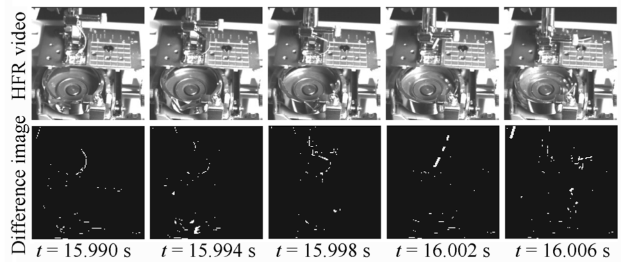
(b) High-frame-rate image sequence for balloon explosion experiment

Fig. 8 Tracking system for HFR recording

It is important to perform long-term monitoring for detecting unpredictable behaviors of high-speed repetitive mechanical motion because abnormal behaviors of mechanical motion produce defective products, which are difficult to detect by using a normal vision system. An intelligent image-based video recording system can help factory workers to find out the reason of abnormal behaviors within a short period of time and enable them to improve the product quality. Wang et al.[34] introduced a form of vision-based machinery surveillance for abnormal behavior detection in high-speed periodic operations. To allow periodic machine operations to be monitored, the surveillance algorithm can estimate the phase of a periodic operation by inspecting the temporal changes in the brightness at several significant pixels in an input image from a single camera; abnormal behavior in the periodic operations can be intelligently detected at a crucial moment. The system can judge and count the abnormal pixels by matching an input image with the reference images. When the abnormal behavior is occurring, the system can detect it and generate a trigger signal for HFR video recording. For instance, an abnormal behavior was detected in a periodic operation and recorded on HFR video during a crucial period, while the needle of a sewing machine was vertically operated at 12.5 Hz. In the experiment, the needle of the sewing machine was suddenly dislocated after it was significantly deformed and its broken fragment flew rapidly from the camera′s view. Fig. 9 (a) shows the summation of an abnormal area in a differential image. Fig. 9 (b) shows the automatically recorded HFR video when abnormal behavior occurred.



(a) Summation of an abnormal area in a differential image



(b) Automatically recorded HFR video when abnormal behavior occurred

Fig. 9 Abnormal behavior detection system

3.2 Vibration dynamics sensing

In the case of an image comprising vibrating objects, temporal periodic changes can be observed in the image intensities at the pixels around the objects, depending on their vibration frequencies. The image intensity at every pixel in an image can be considered to be a time-sequence signal, including periodic changes caused by vibrating objects, when the frame rate of a vision system is sufficiently high for vibration measurement. Recently, an HFR vision-based tracking system was developed[35], which can detect spatiotemporal changes in image intensities as a vibration distribution by implementing digital filters at all the pixels in order to pass signals through a specific band of frequencies, as shown in Fig. 10. The concept can be applied to various types of vibration-based image processing techniques such as vibration source localization and vibration imaging, which cannot be realized by using image features based on a single-frame image.

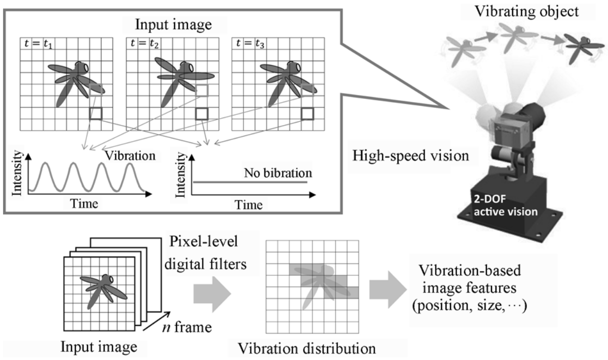


Fig. 10 Vibrating object tracking system

Recently, Yang et al.[28] proposed the concept of dynamics-based visual inspection for the verification of the structural dynamics of a vibrating object, as shown in Fig. 11. In dynamics-based visual inspection, the dynamics represented by the modal parameters are simultaneously estimated from the HFR video and can be used as input-invariant image features for robust structure recognition. These features are clearly different from appearance-based image features obtained from a single-frame image. Even when the static appearance of the structure with and without a crack is similar, a small defect can be detected by examining the difference in the dynamic properties of the structure between the presence and the absence of the crack. Yang et al.[29] applied this idea to a high-speed vision-based structural damage quantification system. The modal parameters of an excited object are simultaneously estimated to determine its input-invariant dynamic properties by using a fast output-only modal analysis algorithm, SSI-CPAST. The algorithm is implemented on a 2 000 fps vision platform and facilitates non-destructive monitoring of the structure of beam-shaped objects vibrating at dozens of hertz. The algorithm detects small changes in the dynamic properties of the objects caused by internal defects such as fatigue cracks. The modal parameters of resonant frequency and mode shape were actually estimated for beam-shaped objects with artificial cracks excited by human finger tapping to verify the performance of the 2 000 fps real-time dynamics-based visual inspection. The authors also expanded their research to dynamics-based stereo visual inspection using a multidimensional modal analysis[30]. Steel beams having a diameter of 0.98 mm were inspected using a high-speed stereo vision system (IDP Express). The 3D shape measurement was accelerated by implementing column-moment calculation circuits on an FPGA on the IDP Express board. Further, the output-only modal parameter estimation was executed for the 72 directions by using the SSI-CPAST method. When excitation was initiated by tapping the object with a human finger, the 72 directional resonant frequencies and mode shapes of the steel beams were estimated in real time. The 1st-order resonant frequencies of the no-crack steel beam were axisymmetric around 26.7 Hz (within 1% deviation), whereas those of the cracked beam had a 9% deviation around 25.0 Hz in the 72 directions, corresponding to the crack location.

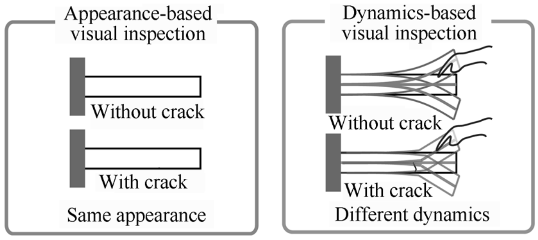


Fig. 11 Concept of dynamics-based visual inspection

3.3 Vision-based mechanical control

On the basis of basic robotic technologies for rapid and flexible responses in the real world, many researchers are studying vision-sensor-based mechanical control combined with high-speed vision, and robot mechanisms that have built-in intelligence in order to realize manipulation/control tasks that differ from those of humans.

Namiki et al.[36] developed a high-speed caching robot hand more than a decade ago. Dynamic grasping using a newly developed high-speed hand system and high-speed vision is proposed. With high-speed visual feedback at the rate of 1 KHz, the hand can grasp and handle a dynamically moving object.

Recently, Namiki et al.[37] developed a novel air-hockey robot system that switches strategies according to the playing styles of its opponent. The system consists of a four-axis robot arm and two high-speed vision sensors. The authors control the robot by using the visual information received at the rate of 500 Hz. The control system consists of three layers: motion control, short-term strategy, and long-term strategy. By integrating the three control layers, the robot exhibits human-like reactions, which increase the appeal of the game.

Senoo et al.[38] designed a tweezers manipulation system to achieve dexterous manipulation of a human tool. First, the contact state between the fingers and the tweezers is analyzed in order to determine the preferable fingers for tool manipulation. Next, the control method based on high-speed visual servoing is presented. Experimental results show that a high-speed hand can grasp a tiny grain with tweezers in two-dimensional (2D) and 3D situations.

Recently, Aoyama et al.[39] applied stereo high-speed vision to passive object manipulation. The objective of this study was to control an object′s orientation with respect to the gravitational force direction by using an active plate for realizing hitherto unrealized object motion. The controlled object is an acrylic ball, which is tumble doll charged clay. The acrylic ball has a mass of 0.3 kg, and two markers are attached for angular measurement of the controlled object. An aluminum plate is attached at the end effector of the parallel link robot, and then, the plate is used as an active plate. A marker tracking system is hardware implemented onto the high-speed vision platform. The system can determine the 3D position of the marker at 1 000 fps by using stereo measurement.

Broadcasting contents of sport games (e.g., the FIFA World Cup and the Olympics) is very popular. However, it is often difficult for camera operators to keep tracking their camera′s direction on a dynamic object such as a particular player or a ball because the camera head and lens are too heavy to motor for ultra-fast movements. Saccade Mirror, which is lightweight and can move in a wide range at several hundred hertz, has been employed to solve this issue. Based on this concept, Okumura et al.[40] developed "1 ms Auto Pan-Tilt" technology, which can automatically control the camera′s pan-tilt angles to always keep an object at the center of the field, just like "autofocus" keeps an object in focus. Saccade Mirror controls a camera′s gazing direction not by moving the camera itself but by rotating small two-axis galvanometer mirrors. It controls the gaze by 60 deg, the widest angle, for both pan and tilt, and steering the gaze by 40 deg takes only 3.5 ms. The "1 ms Auto Pan-Tilt" technology has been applied to a ping-pong game, in which vision-based feedback control is conducted at 1 000 fps and a full-HD high-speed video for broadcasting is captured at 500 fps. The ping-pong ball is always kept in the center of the camera view by using the "1 ms Auto Pan-Tilt" technology. This technology can be used by a broadcasting service for sport games and for recording the detailed dynamics of a flying bird, an insect, a car, an aircraft, etc.

3.4 3D measurement

Nowadays, vision-based shape inspection technology plays an important role in accurate product quality management in the field of industrial manufacturing. A high frame-rate 3D shape measurement system[41] is developed by introducing a temporally coded multi-slit projection method, which can work at 3 350 fps for a 256×256 pixel image. However, it is an offline system, which cannot be used as a real-time 3D sensor. Recently, Gao et al.[42] proposed a novel light-section method that can accurately obtain a differential shape from a given reference 3D shape at a high frame rate by projecting a self-projected light pattern, i.e., a curved-line light pattern generated by the 3D shape of a reference object. An HFR 3D shape measurement system based on the concept was also developed, and its effectiveness was demonstrated by measuring the 3D shapes of cylinder-shaped objects in real time at 10 000 fps. The authors also expanded this self-projection concept to the structured light method. Recently, a self-projected structured light depth measurement system[43] was proposed. It can simultaneously inspect depth images at a high speed even when there are large height differences in the measured scene. Different shapes are acquired from a 3D reference shape by processing multiple straight-stripe-like patterns projected in the camera view, which are projections of multiple curved-stripe patterns generated by the reference shape. Compared with straight-stripe projections in most structured light methods, the number of curved-stripe projections can be reduced by using this method, which facilitates rapid processing and the acquisition of different 3D shapes with little degradation of the accuracy when a 3D shape in the background scene is provided as a reference. Using this method, the authors developed a structured light system that can acquire and process 512×512 depth images in real time at 500 fps by using a high-speed vision platform synchronized with a high-speed digital light processing (DLP) projector.

The frame rate of the image sensor and projector restricts the accuracy of measuring 3D shapes of moving objects because of the synchronization errors encountered in the projection of multiple light patterns with different timing. To avoid these synchronization errors, Liu et al.[44] proposed a motion-compensated structured-light coding method, which can be performed in real time at 500 fps by estimating the velocity of a target object in a 3D scene. When the object is assumed to move at uniform speed on the image plane during projection times, the corresponding motion leads to pixel displacements at each point of the object among the captured images. The estimated velocity of the target object can be used for predicting the corresponding positions in the previous frames and compensate for the pixel displacements among the different captured images. Thus, the synchronization errors can be further reduced and minimized.

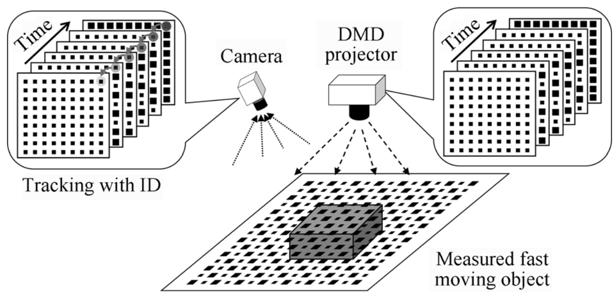


Fig. 12 Blink-spot projection method

However, 3D measurement errors of moving objects remain even when a tracking based motion compensation method is used. Watanabe et al.[45] proposed a novel multi-spot tracking method to measure the 3D shape of a fast moving and deformable object by using high-speed vision and laser dot matrix. However, the latency of the system was 4.5 ms, which is too high to perform fast feedback control. Based on Watanabe′s idea[45], Chen et al.[31] proposed a multi-spot projection method called the "blink-spot projection method". In this method, a certain number of multi-spot light patterns, where large and small spots alternatively blink at different times corresponding to their ID numbers, are projected by an HFR projector onto the scenes being measured. The image sequences captured by a camera synchronized with the projector are simultaneously processed to extract the spots′ ID numbers to achieve robust and accurate 3D shape measurement. This method performs frame-to-frame correspondence of multiple spots and generates a robust ID number for each dot, as shown in Fig. 12. The position and size of multiple spots in each frame are extracted by using hardware acceleration[17]. As a result, the 3D positions of 16×16 spots can be generated at 500 fps without any synchronization errors. Compared with Microsoft Kinect[46], this method measures the scene with gap-depth accuracy without imposing restrictions on the surface of the scene. Further, compared with that of sequential projection techniques, the accuracy of the blink-spot projection method is stable even when the measured object moves at a high speed. To measure 3D information of a fast moving or deformable object, another reasonable approach is to reduce the projected pattern. Recently, Tabata et al.[47] proposed a high-speed 3D sensing system with three-view geometry using a segmented pattern, which can work at 500 fps. In this research, only one well-designed pattern with 16 different primitives is projected by the digital micromirror device (DMD) projector and two calibrated high-speed cameras are used to determine the corresponding primitives between two views. After this sparse identification, more dense 3D positions are obtained by the coarse-to-fine approach.

3.5 Vision-based human interface

Projection mapping is a well-known spatial augmented reality technology for projecting computer-generated light patterns from projectors onto the real environment. Many illusion works based on projection mapping, such as Virtual Showcase, projector-guided painting, and architectural projection mapping, have been reported thus far. Most of them have only been conducted for static 3D scenes because they assume that the geometry of the real environment has been measured beforehand by using offline depth sensing such as laser-scan sensing. Cassinelli and Ishikawa[48] developed a Khronos projector based on high-speed 3D vision and projection mapping technique, which is a real-time interactive augmented reality system. Based on Cassinelli and Ishikawa′s method[48], Watanabe et al.[49] developed a deformable workspace for manipulating 3D virtual objects by using a membrane between the real and virtual space. Chen et al.[50] reported the development of an HFR camera-projector system that can acquire and process 512×512 depth images in real time at 500 fps and project computer-generated light patterns onto time-varying 3D scenes. The depth image processing is accelerated by installing a GPU board to process the 8-bit gray-code structured light method using infrared (IR) light patterns projected at 1 000 fps. Red-green-blue (RGB) light patterns are interactively generated so that the patterns are projected onto the 3D scene and enhanced with pixel-wise correspondence, as shown in Fig. 13 (a). A plaster lion relief (height 31 cm, width 27 cm, and depth 10 cm) to be enhanced was moved above the level plane with periodic up-and-down motion once or less per second. In Fig. 13 (b), we can see that the 3D shape information of the relief was correctly measured in the depth images, even when the relief was moved up-and-down with slight rotation. The white-surface relief was enhanced with a cyclic jet color map, which could directly visualize its detailed height information for the human eye, and the RGB light patterns were correctly projected for pixel-wise projection mapping on the moving lion relief.

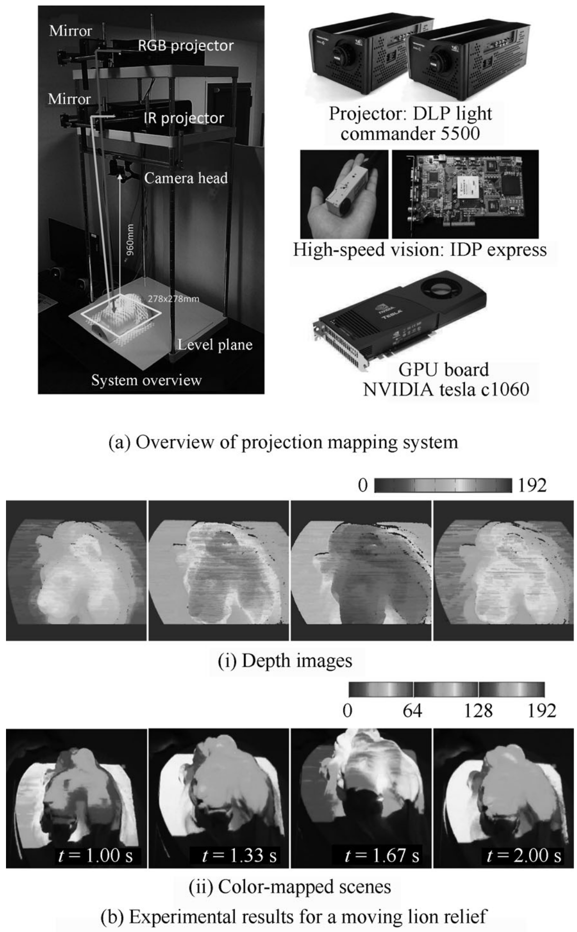


Fig. 13 Projection mapping system

The projection mapping area of the moving object is limited by the view of fixed camera and projectors in [50]. A large delay, caused by the time between measuring the object and projecting images on the object, results in significant misalignment in the case of dynamic objects. Recently, to solve this problem, Sueishi et al.[51] developed a Lumipen system, which can provide projected images that are fixed on dynamic objects such as bouncing balls. The "1 ms Auto Pan-Tilt" technology was employed to solve the time-geometric inconsistency caused by the delay observed when using dynamic objects. However, the robustness of the tracking was sensitive to the simultaneous projection on the object, as well as the environmental lighting. In order to achieve robust dynamic projection mapping, the authors introduced a retroreflective background to the Lumipen system. As a result, the object appeared darker than the background during projection, which was observed using a high-speed camera. Therefore, the tracking would be robust to changes in the content of the projection such as movies, and changes in the environmental lighting in the object′s vicinity and associated with its motion. This opens up avenues for new applications of projection mapping, such as the visualization of a pitched ball as a fire ball.

In the case of rapid human motion, video images at frame rates of several dozen frames per second are often insufficient for tracking a human face responsively. Therefore, there is a strong demand for high-speed face tracking at frequencies of the order of several hundred hertz. Recently, Ishii et al.[52] developed a high-speed vision system that can be applied to real-time face tracking at 500 fps by using the GPU acceleration of a boosting-based face tracking algorithm. By assuming a small image displacement between frames, which is a property of HFR vision, the authors developed an improved boosting-based face tracking algorithm for fast face tracking by enhancing the Viola-Jones face detector. In the improved algorithm, the size and position of a face pattern to be tracked in an image can be efficiently extracted by reducing the number of window searches for Haar-like features and by combining skin color extraction with Haar classifiers. The improved boosting-based face tracking algorithm is implemented on a GPU-based high-speed vision platform, and face tracking can be executed in real time at 500 fps for an 8-bit color image of 512×512 pixels. The proposed face tracking algorithm is an enhanced version of the Violajones face detector, developed by combining skin-color-based extraction with Haar classifiers. The proposed algorithm involves: (a) a learning process, (b) a searching process, and (c) a tracking process. In the learning process, Haar classifiers are trained offline by using positive and negative samples, as in the case of the Violajones face detector. After the learning process, the searching process is executed for detecting face patterns in the entire image region. Then, the tracking process is executed for efficiently selecting region of interests (ROIs) by assuming HFR color vision.

3.6 Biomedical applications

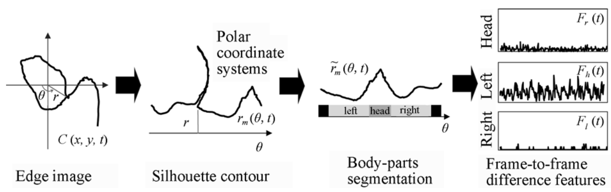


Fig. 14 Repetitive behavior analysis system

Laboratory animals such as mice and rats are used in the applied research for new drug development, biomedical research, and toxicology testing. There are many types of high-speed model behaviors such as scratching and grooming, involving repetitive motions of their fore or hind limbs at several or dozens of hertz, which are much faster than those of the human leg motion. Nie et al.[53] proposed an HFR-video-based behavior recognition algorithm for laboratory mice to quantify multiple model behaviors including repetitive behaviors at dozens of hertz. The algorithm can segment repetitive limb movements into several regions by using a mouse′s silhouette in polar coordinate systems, and multiple model behaviors can always be quantified independent of a mouse′s position and orientation, as shown in Fig. 14. Automatic discrimination of the six behaviors ("immobility", "moving", "left-side scratching", "right-side scratching", "head grooming", and "rearing") is performed in 10-min experiments for several mice. The correspondence ratios for all the behaviors are larger than 80%, and these ratios are sufficiently high for most quantification requirements of animal testing.

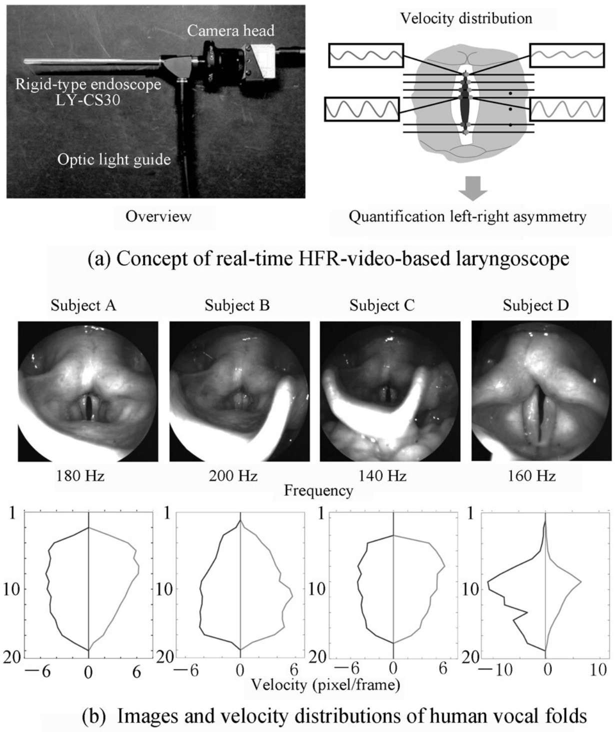


Fig. 15 Real-time HFR-video-based laryngoscope

Ishii et al.[54] developed an HFR laryngoscope that can measure the vibration distribution of a vocal fold in real time at hundreds of hertz. The developed laryngoscope consists of an IDP Express high-speed vision platform and a rigid-type endoscope, as shown in Fig. 15 (a). A contour-extraction algorithm was implemented on a PC to obtain a velocity distribution of 20 pairs of border points on the left and right edges of a vocal fold in real time. Based on the vibration velocities, the amplitude ratio and correlation ratio are introduced as quantification indexes for the left-right asymmetry of vocal-fold vibrations to indicate the degree of laryngeal disease. The authors examined the vocal-fold vibrations of four human subjects under clinical conditions (512×512 images at 2 000 fps). Subjects A, B and C were healthy males in their twenties. Subject D was a female patient in her fifties with a polypous vocal cord. For subjects A, B and C, the amplitude ratios A were around 0 and the correlation ratios R were around -1. These tendencies indicated that their vocal folds vibrated almost left-right symmetrically. For subject D, the authors observed that the left-right symmetry in the vocal-fold vibrations was strongly disrupted because a polyp on the right side of the vocal fold disturbed the vocal-fold vibrations, as shown in Fig. 15 (b).

Several non-vision cell analysis systems have been developed to extract the shape and motion of cells in lab-on-a-chip (LOC)[55, 56]. However, the detectable geometrical properties of cells have been limited by several technical factors, such as poor spatial resolution and difficulties in conducting quantitative measurements at such a small scale. Offline high-speed cameras have been used for cell-shape analysis. However, the measurable speeds of cells in microchannel flows are limited by their frame intervals because their frame intervals are not sufficiently small for capturing the fast apparent motion of cells magnified in the microscopic view. Gu et al.[32] proposed a high-speed vision-based shape and motion analysis system for cells, which can overcome this restriction, and operates as a real-time inspection tool for cells flowing in microchannels at several meters per second, where their shapes are deformed corresponding to their mechanical properties. Lab-on-a-chip practitioners can easily apply this system to cells fast-flowing in microchannels, such as red blood cells or iPS cells, which need to be inspected by checking their size, shape, stiffness, and other image-based parameters for automated mass production of good-quality cells. The system can synchronize two camera inputs with the same view with only a tiny time delay on the sub-microsecond timescale, as shown in Fig. 16. Real-time video processing is performed using hardware logic by extracting the moment features of multiple cells in 512×256 images at 4 000 fps for the two camera inputs. Their frame-straddling time can be adjusted from 0 to 0.25 ms in steps of 9.9 ns. By setting the frame-straddling time within a certain range to avoid large image displacements between the two camera inputs, we can use the proposed frame-straddling high-speed vision platform to perform a simultaneous shape and motion analysis of cells in fast microchannel flows of 1 m/s or greater.

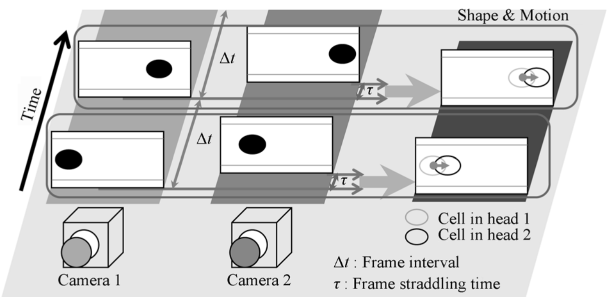


Fig. 16 Frame-straddling multi-object feature extraction system

For cells in fast-flowing microchannels at 2 m/s, the displacement between frames is 500 pixels/frame (at 4 000 fps; 1 pixel = 1 m), which is smaller than the resolution of a high-speed vision system in the direction of the flow. Thus, all passed cells can be detected in a 2 000-l/min flow.

Recently, a high-speed vision-based morphological analysis system for fast-flowing cells in a microchannel was proposed[33]. Real-time video processing is performed using hardware logic by extracting the moment features and bounding boxes of multiple cells in 512×256 pixel images at 2 000 fps. By extracting the bounding boxes of the cell regions using hardware logic and shrinking the cell region to a certain size to reduce the processing time, the proposed high-speed vision system can perform a fast morphological analysis of cells at 2 ms/cell in fast microchannel flows.

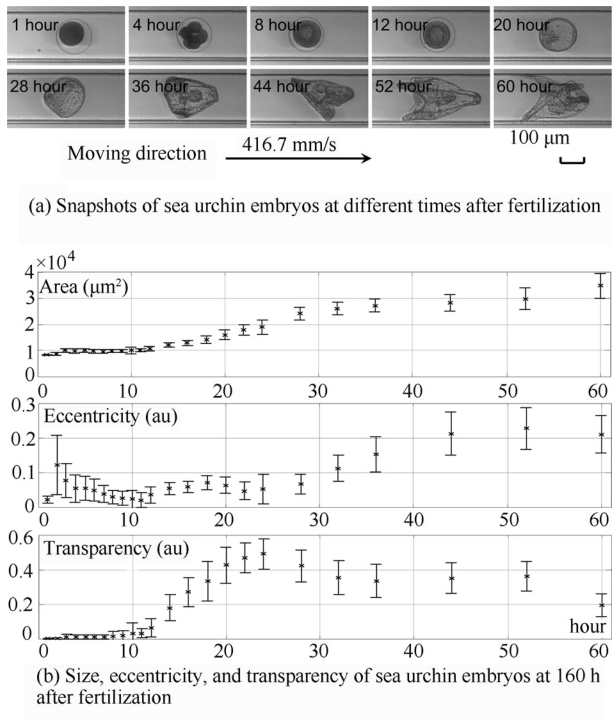


Fig. 17 High-speed vision-based morphological analysis system

Fig. 17 (a) shows snapshots of sea urchin embryos at different time points after fertilization. Fig. 17 (b) shows the average values and standard deviations of the size, eccentricity, and transparency of the sea urchin embryos from 1 h to 60 h after fertilization.

Red blood cells (RBCs) circulate in the human body several hundred thousand times during their entire life span. Therefore, their deformability is really important, particularly when they pass through a local capillary whose diameter can be as narrow as 3 μm. Sakuma et al.[57] presented a new concept of RBC fatigue evaluation. The fatigue state is defined by the time of reciprocated mechanical stress when the extensibility and the recoverability characteristics meet each other. To achieve stable and accurate control of RBCs in a microchannel, two fundamental components are employed in this study. One is a robotic pump capable of manipulating a cell with the accuracy of $\pm 0.24\,\mu$m in an equilibrium state with a maximum response time of 15 ms. The other is an online high-speed camera capable of chasing the position of RBCs with a sampling rate of 1 kHz. By utilizing these components, the authors achieve continuous observation of the length of an RBC over a 1 000 times reciprocated mechanical stress.

4 Conclusions

In this paper, we reviewed the recent progress of a real-time high-speed vision system and its applications, based on our experience. High-speed vision has already applied in various application fields as a powerful dynamic sensing tool, as shown in Fig. 3. In particular, several hyper-human applications were introduced in this paper by utilizing the properties of high-speed vision, such as dynamics-based visual inspection, fast vision-based manipulation, and cell flow analysis and sorting. Further, many conventional image processing algorithms can be simplified by considering the small displacement between frames of high-speed vision. Simplified algorithms can be implemented and accelerated on different platforms (FPGA GPU and CPU) according to the purpose of application. The real-time high-speed vision technology is almost a technological revolution in the field of vision sensors. The small frame interval of high-speed vision enables humans to observe and analyze ultra-fast phenomena, and the real-time processing technology of huge frame data enables a fast reaction speed. Moreover, real-time high-speed vision is expected to be a very exciting and attractive field in the next ten years. Currently, the sensitivity of a high-speed vision system is low when working at a high frame rate, and the resolution of the high-speed vision system is limited to 1 megapixel when working at 1 000 fps or more. We expect a new generation of image sensors with full resolution and high sensitivity. In the future, the target application fields of real-time high-speed vision will include biomedicine, micro-/nano- manipulation, infrastructure monitoring, and high-frequency object detection and tracking.

References

[1] P. Read, M. P. Meyer. *Restoration of Motion Picture Film*, Oxford, UK: Butterworth-Heinemann, pp. 24–26, 2000.

[2] E. Altug, J. P. Ostrowski, R. Mahony. Control of a quadrotor helicopter using visual feedback. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Philadelphia, USA, pp. 72–77, 2002.

[3] P. C. Zhang, D. Xu. Tracking and guiding multiple Laser beams for beam and target alignment. *International Journal of Automation and Computing*, vol. 12, no. 6, pp. 600–610, 2015.

[4] Y. Watanabe, H. Oku, M. Ishikawa. Architectures and applications of high-speed vision. *Optical Review*, vol. 21, no. 6, pp. 875–882, 2014.

[5] T. Senoo, Y. Yamakawa, Y. Watanabe, H. Oku, M. Ishikawa. High-speed vision and its application systems. *Journal of Robotics and Mechatronics*, vol. 26, no. 3, pp. 287–301, 2014.

[6] T. M. Bernard, B. Y. Zavidovique, F. J. Devos. A programmable artificial retina. *IEEE Journal of Solid-State Circuits*, vol. 28, no. 7, pp. 789–797, 1993.

[7] J. E. Eklund, C. Svensson, A. Astrom. VLSI implementation of a focal plane image processor-A realization of the near-sensor image processing concept. *IEEE Transactions on Very Large Scale Integration Systems*, vol. 4, no. 3, pp. 322–335, 1996.

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

[9] S. Kleinfelder, S. Lim, X. Liu, 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.

[10] T. Komuro, S. Kagami, M. Ishikawa. A dynamically reconfigurable SIMD processor for a vision chip. *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 265–268, 2004.

[11] I. Ishii, K. Yamamoto, M. Kubozono. Higher order autocorrelation vision chip. *IEEE Transactions on Electron Devices*, vol. 53, no. 8, pp. 1797–1804, 2006.

[12] W. C. Zhang, Q. Y. Fu, N. J. Wu. A programmable vision chip based on multiple levels of parallel processors. *IEEE Journal of Solid-State Circuits*, vol. 46, no. 9, pp. 2132–2147, 2011.

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

[14] S. Hirai, M. Zakoji, A. Masubuchi, T. Tsuboi. Real-time FPGA-based vision system. *Journal of Robotics and Mechatronics*, vol. 17, no. 4, pp. 401–409, 2005.

[15] Y. Watanabe, T. Komuro, M. Ishikawa. A high-speed vision system for moment-based analysis of numerous objects. In *Proceedings of IEEE International Conference on Image Processing*, IEEE, San Antonio, USA, pp. V177–V180, 2007.

[16] I. Ishii, R. Sukenobe, T. Taniguchi, K. Yamamoto. Development of high-speed and real-time vision platform, H^3 Vision. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, St. Louis, USA, pp. 3671–3678, 2009.

[17] Q. Y. Gu, T. Takaki, I. Ishii. Fast FPGA-based multi-object feature extraction. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 30–45, 2013.

[18] I. Ishii, R. Sukenobe, K. Yamamoto, T. Takaki. Real-time image recognition using HLAC features at 1000 fps. In *Proceedings of IEEE International Conference on Robotics and Biomimetics*, IEEE, St. Louis, USA, pp. 954–959, 2009.

[19] Q. Y. Gu, S. Raut, K. Okumura, T. Aoyama, T. Takaki, I. Ishii. Real-time image mosaicing system using a high-framerate video sequence. *Journal of Robotics and Mechatronics*, vol. 27, no. 1, pp. 12–23, 2015.

[20] I. Ishii, T. Tatebe, Q. Y. Gu, T. Takaki. Color-histogram-based tracking at 2000 fps. *Journal of Electronic Imaging*, vol. 21, no. 1, Article number 013010, 2012.

[21] Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. High framerate tracking of multiple color-patterned objects. *Journal of Real-time Image Processing*, vol. 11, no. 2, pp. 251–269, 2016.

[22] T. Komuro, T. Tabata, M. Ishikawa. A reconfigurable embedded system for 1000 f/s real-time vision. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 496–504, 2010.

[23] I. Ishii, T. Tatebe, Q. Y. Gu, Y. Moriue, T. Takaki, K. Tajima. 2000 fps real-time vision system with high-framerate video recording. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Anchorage, USA, pp. 1536–1541, 2010.

[24] Y. Watanabe, T. Komuro, S. Kagami, M. Ishikawa. Real-time visual measurements using high-speed vision. In *Proceedings of SPIE 5603, Machine Vision and Its Optomechatronic Applications*, SPIE, Philadelphia, USA, pp. 234–242, 2004.

[25] Q. Y. Gu, T. Takaki, I. Ishii. A fast multi-object extraction algorithm based on cell-based connected components labeling. *IEICE Transactions on Information and Systems*, vol. E95-D, no. 2, pp. 636–645, 2012.

[26] Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. Fast tracking system for multi-colored pie-shaped markers. *International Journal of Optomechatronics*, vol. 7, no. 3, pp. 60–80, 2013.

[27] I. Ishii, T. Taniguchi, K. Yamamoto, T. Takaki. High-framerate optical flow system. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 1, pp. 105–112, 2012.

[28] H. Yang, T. Takaki, I. Ishii. Simultaneous dynamics-based visual inspection using modal parameter estimation. *Journal of Robotics and Mechatronics*, vol. 23, no. 1, pp. 180–195, 2011.

[29] H. Yang, T. Takaki, I. Ishii. A structural damage quantification method for HFR-video-based modal testing. *Journal of System Design and Dynamics*, vol. 5, no. 4, pp. 624–641, 2011.

[30] H. Yang, Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. Dynamics-based stereo visual inspection using multidimensional modal analysis. *IEEE Sensors Journal*, vol. 13, no. 12, pp. 4831–4843, 2013.

[31] J. Chen, Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. Blink-spot projection method for fast three-dimensional shape measurement. *Journal of Robotics and Mechatronics*, vol. 27, no. 4, pp. 430–443, 2015.

[32] Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. Simultaneous vision-Based shape and motion analysis of cells fast-flowing in a microchannel. *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 204–215, 2015.

[33] Q. Y. Gu, T. Kawahara, T. Aoyama, T. Takaki, I. Ishii, A. Takemoto, N. Sakamoto. LOC-Based high-throughput cell morphology analysis system. *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1346–1356, 2015.

[34] Y. D. Wang, I. Ishii, T. Takaki. HFR-video-based machinery surveillance for high-speed periodic operations. *Journal of System Design and Dynamics*, vol. 5, no. 6, pp. 1310–1325, 2011.

[35] I. Ishii, I. Ohara, T. Tatebe, T. Takaki. 1000 fps target tracking using vibration-based image features. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Shanghai, China, pp. 1837–1842, 2011.

[36] A. Namiki, Y. Imai, M. Ishikawa, M. Kaneko. Development of a high-speed multifingered hand system and its application to catching. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Las Vegas, USA, pp. 2666–2671, 2003.

[37] A. Namiki, S. Matsushita, T. Ozeki, K. Nonami. Hierarchical processing architecture for an air-hockey robot system. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Karlsruhe, Germany, pp. 1187–1192, 2013.

[38] T. Senoo, D. Yoneyama, A. Namiki, M. Ishikawa. Tweezers manipulation using high-speed visual servoing based on contact analysis. In *Proceedings of IEEE International Conference on Robotics and Biomimetics*, IEEE, Karon Beach, Thailand, pp. 1936–1941, 2011.

[39] T. Aoyama, Y. Harada, Q. Y. Gu, T. Takaki, I. Ishii. Cyclic motion design and analysis for a passive object manipulation using an active plate. *Advanced Robotics*, vol. 29, no. 7, pp. 493–503, 2015.

[40] K. Okumura, H. Oku, M. Ishikawa. High-speed gaze controller for millisecond-order pan/tilt camera. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Shanghai, China, pp. 6186–6191, 2011.

[41] J. Takei, S. Kagami, K. Hashimoto. 3,000 fps 3D shape measurement using a high-speed camera-projector system. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, San Diego, USA, pp. 3211–3216, 2007.

[42] H. Gao, Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. A self-projected light-section method for fast three-dimensional shape inspection. *International Journal of Optomechatronics*, vol. 6, no. 4, pp. 289–303, 2012.

[43] H. Gao, T. Aoyama, T. Takaki, I. Ishii. Self-projected structured light method for fast depth-based image inspection. In *Proceedings of International Conference on Quality Control by Artificial Vision*, Fukuoka, Japan, pp. 175–180, 2013.

[44] Y. J. Liu, H. Gao, Q. Y. Gu, T. Aoyama, T. Takaki, I. Ishii. High-frame-rate structured light 3D vision for fast moving objects. *Journal of Robotics and Mechatronics*, vol. 26, no. 3, pp. 311–320, 2014.

[45] Y. Watanabe, T. Komuro, M. Ishikawa. 955-fps real-time shape measurement of a moving/deforming object using high-speed vision for numerous-point analysis. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Roma, Italy, pp. 3192–3197, 2007.

[46] Microsoft. *Kinect*, [Online], Available: https://en.wikipedia.org/wiki/Kinect, June 29, 2015.

[47] S. Tabata, S. Noguchi, Y. Watanabe, M. Ishikawa. High-speed 3D sensing with three-view geometry using a segmented pattern. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Hamburg, Germany, pp. 3900–3907, 2015.

[48] A. Cassinelli, M. Ishikawa. Khronos projector, emerging technologies. In *Proceedings of SIGGRAPH*, ACM, Los Angeles, USA, pp. 4506–4511, 2004.

[49] Y. Watanabe, A. Cassinelli, T. Komuro, M. Ishikawa. The deformable workspace: A membrane between real and virtual space. In *Proceedings of IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, IEEE, Amsterdam, Netherlands, pp. 145–152, 2008.

[50] J. Chen, T. Yamamoto, T. Aoyama, T. Takaki, I. Ishii. Simultaneous projection mapping using high-frame-rate depth vision. In *Proceedings of IEEE International Conference on Robotics and Automation*, IEEE, Hong Kong, China, pp. 4506–4511, 2014.

[51] T. Sueishi, H. Oku, M. Ishikawa. Robust high-speed tracking against illumination changes for dynamic projection mapping. In *Proceedings of IEEE Conference on Virtual Reality*, IEEE, Arles, France, pp. 97–104, 2015.

[52] I. Ishii, T. Ichida, Q. Y. Gu, T. Takaki. 500-fps face tracking system. *Journal of Real-Time Image Processing*, vol. 8, no. 4, pp. 379–388, 2013.

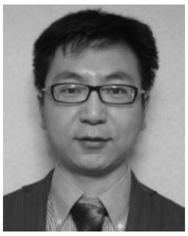
[53] Y. M. Nie, T. Takaki, I. Ishii, H. Matsuda. Algorithm for automatic behavior quantification of laboratory mice using high-frame-rate videos. *SICE Journal of Control, Measurement, and System Integration*, vol. 4, no. 5, pp. 322–331, 2011.

[54] I. Ishii, S. Takemoto, T. Takaki, M. Takamoto, K. Imon, K. Hirakawa. Real-time laryngoscopic measurements of vocal fold vibrations. In *Proceedings of IEEE International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, Boston, USA, pp. 6623–6626, 2011.

[55] A. Drochon. Use of cell transit analyzer pulse height to study the deformation of erythrocytes in microchannels. *Medical Engineering & Physics*, vol. 27, no. 2, pp. 157–165, 2005.

[56] Y. Katsumoto, K. Tatsumi, T. Doi, K. Nakabe. Electrical classification of single red blood cell deformability in high-shear microchannel flows. *International Journal of Heat and Fluid Flow*, vol. 31, no. 6, pp. 985–995, 2010.

[57] S. Sakuma, K. Kuroda, C. Tsai, W. Fukui, F. Arai, M, Kaneko. Red blood cell fatigue evaluation based on the close-encountering point between extensibility and recoverability. *Lab on a Chip*, vol. 14, no. 6, pp. 1135–1141, 2014.



Qing-Yi Gu received the B. Eng. degree in electronic and information engineering from Xi'an Jiaotong University, China in 2005. He received the M. Eng. degree, and Ph. D. degree in engineering, from Hiroshima University, Japan in 2010 and 2013, respectively. He is currently a designated associate professor in Graduate School of Engineering, Hiroshima University, Japan.

His research interests include high-speed image feature extraction, and applications in industry and biomedicine.

E-mail: gu@robotics.hiroshima-u.ac.jp (Corresponding author)

ORCID iD: 0000-0001-8332-5350

Idaku Ishii received the B. Eng. the M. Eng. and the Ph. D. degrees from the University of Tokyo, Japan in 1992, 1994, and 2000, respectively. He is currently a professor in Graduate School of Engineering, Hiroshima University, Japan.

His research interests include high-speed robot vision, sensory information processing, sensor-based robot manipulation, and applications in industry and biomedicine.

E-mail: iishii@robotics.hiroshima-u.ac.jp