Review Article

Yoshihiro Watanabe*

High-speed optical 3D sensing and its applications

Abstract: This paper reviews high-speed optical 3D sensing technologies for obtaining the 3D shape of a target using a camera. The focusing speed is from 100 to 1000 fps, exceeding normal camera frame rates, which are typically 30 fps. In particular, contactless, active, and real-time systems are introduced. Also, three example applications of this type of sensing technology are introduced, including surface reconstruction from time-sequential depth images, high-speed 3D user interaction, and high-speed digital archiving.

Keywords: camera; high-speed sensing; optical sensing; shape measurement.

1 Introduction

There is an increasing demand for application systems that can realize complicated tasks at high speed. To meet this demand, optical sensing using an image sensor is an essential elemental technology. In particular, for new applications, it is important not to aim to substitute the recognition function of the human eye, but to observe the real-world at speeds exceeding human capabilities. A target performance level of 1000 fps will enable various applications. This rate greatly exceeds the speed of conventional imaging systems, which typically operate at standard video rates (30 fps). In addition, system architectures that can simultaneously execute not just image capturing and recording but also image processing at a level of 1000 fps will result in technological evolutions. Recently, it has become possible to obtain the 3D shape of a target at such speeds. This paper reviews recent

technologies related to high-speed 3D sensing, meaning sensing for obtaining 3D shapes.

The methods of realizing 3D sensing can be categorized as follows. First, there are the contact-dependent type and the contactless type. Contactless-type 3D sensing is based on optical sensing. These types include a passive method in which the target is observed by capturing the environmental incident illumination reflected at the target surface and an active method in which a special light pattern is projected onto the target. This paper, in particular, focuses on contactless, active 3D sensing.

One of the applications of 3D sensing is the generation of a 3D model. In this application, the targets are mainly static objects, and the sensing obtains high-accuracy shapes over time. For example, a laser rangefinder sequentially scans a static target with light to obtain the whole shape [1]. In other approaches, camera images are stored, and the shape is reconstructed in the form of batch processing. For example, slow-motion 3D video has been reconstructed from 1000-fps video images captured by a high-speed camera specially designed for recording [2]. These approaches require scanning and a large amount of calculation, and are therefore designed for offline operation.

In contrast to this offline 3D sensing, real-time 3D sensing for obtaining a shape at the same speed as the imaging speed can be used in a wider range of applications. This type of 3D sensing is strongly desired in applications such as robot control, map generation using a mobile robot, human-machine interfaces for computers and games, inspection, medical support, and vehicle control. In particular, the Kinect game controller [3] is widely used in such applications.

However, the speed of most real-time 3D sensing systems is around 30–60 fps. As mentioned above, this speed is not high enough for real-time applications. In fact, many high-speed optical sensing systems are effective in various application fields [4], although most systems cannot obtain 3D shapes.

Against this background, in this paper, we introduce contactless, active, and real-time 3D sensing technologies. In particular, we focus on the systems that exploit

---
*Corresponding author: Yoshihiro Watanabe, Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan, e-mail: Yoshihiro_Watanabe@ipc.i.u-tokyo.ac.jp

the speed. This paper is organized as follows. First, we introduce two approaches for high-speed 3D sensing with multi-pattern projection and single-pattern projection. In multi-pattern projection, the speed of both a camera and a projector are required to be high. Also, the speed is required to be high enough to assume that the target hardly moves. If these requirements are met, the achieved spatial resolution can be high. On the other hand, in single-pattern projection, the limitation about the target motion can be relaxed. Using a laser with diffractive optical elements, we can make a projector compact. However, 3D reconstruction method becomes difficult because the obtained amount of information in a single captured image is small. This leads the calculation cost to be high and becomes a critical problem for real-time and high-speed 3D sensing. The introduced sensing technologies solve such problems in various approaches and achieve high-speed performance.

Also, examples of applications based on this type of 3D sensing are introduced later in this paper. The high speed, which is far superior to that of conventional sensing, provides not only quantitative shifts, such as enhancing the performance of existing applications, but also qualitative shifts, such as a key technology that allows the creation of new applications. This paper introduces three application examples. The first example is surface reconstruction from time-sequential depth images. Toward the speed up of the acquisition of detailed all-around 3D modeling and wide range environment, the high-speed 3D sensing has a strong impact in this application. The second example is high-speed 3D user interaction. In order to observe the user's high-speed 3D motion and reduce the latency from the user motion to display, 3D sensing needs to be high speed. The third example is high-speed digital archiving. The number of targets for the conservation and share through the Internet has been increased more and more in recent years. To meet this demands, the 3D sensing speed requires to be improved drastically.

2 High-speed optical 3D sensing

2.1 3D sensing with multi-pattern projection

A representative method for contactless, active 3D sensing is based on triangulation. A simple system configuration is used to project slit light generated by a laser, capture the laser light reflected at the target from a different angle from the projection angle, and determine the 3D positions of the locations on the target where the slit light is reflected

[1]. If the slit light is scanned such that it passes through all points on the target surface and is imaged at each scanning position, the whole shape can be obtained by collecting all measured results from multiple images. In order to achieve high-speed 3D sensing of the whole shape, it is necessary to improve both the scanning speed and the frame rate of the camera. For example, there is a system that can achieve 30-fps 3D sensing using a 50- to 260-kfps camera [5, 6]. As demonstrated with this system, it is difficult to improve the sensing speed with this type of scanning. Moreover, if the target moves, the sensing results in significant errors because the timing of the result at each scanning angle is shifted, and the connected shape is thus distorted.

One of the solutions to this problem is to project a two-dimensional pattern and reduce the number of captured images required to obtain the whole shape. This kind of pattern is called structured light in the field of active 3D sensing. In this method, the captured image contains the two-dimensional pattern reflected at the target and distorted according to the surface shape. Therefore, it is necessary to determine the correspondence between the camera image and the projected image.

This can be solved by projecting multiple, different patterns and imaging them. This solution uses the fact that the time change caused by switching the patterns at the corresponding points in the camera image and projected image is the same. However, similar to the above scanning method, the target needs to be static until projection of all of the patterns is completed.

To overcome this limitation and allow 3D sensing of moving objects, there is a system in which the frame rates of the camera and the projector are improved, thus, reducing the total time for projecting multiple patterns and reducing the measurement errors caused by the movement [7]. In this system, 16 binary stripe patterns, in which Gray codes are embedded, are projected. The shape is obtained at 500 fps in the form of 512×512 depth images in which each pixel has depth information. Another method using a high-frame-rate camera utilizes dithering in a DLP projector [8], although the shape is not calculated in real-time. If the image projected by the DLP projector is captured by a camera whose frame rate is higher than that of the projector, the pixel values in the camera are changed temporally and differ according to the brightness value set at each pixel in the projector image. This time change can be used for calculating the correspondence. This system was used to reconstruct a 3D shape from 20 images captured at 3000 fps. There is also a system projecting the structured light using laser speckle [9], instead of using off-the-shelf projectors. In this system, the multiple speckle patterns

are projected by moving the diffuser in front of a laser, and two cameras capture about 10 images for a single depth image. The correspondence matching for each pixel between the cameras is based on the temporal correlation technique. The acquisition rate of depth image is reported as 57 images per second when the frame rate of the camera is 207 fps, although the calculation for the measurement is not done online. These methods are based on binary codes and use binary patterns. One advantage of this is the ability to robustly detect a pattern in the camera image even when the target has spatially varying textures.

On the other hand, structured light can use not only binary patterns but also gradation patterns. Using gradation images, the total number of patterns can be reduced. A well-known example of this technique is the phase-shift method [1]. The phase-shift method uses fringe patterns in which the brightness is changed sinusoidally. The minimum number of patterns is three. One of the bottlenecks for the method using gradation-image projection is the projection speed. In the recent research, a high-speed projector has been developed, which can project 8-bit images at 1000 fps with 3-ms latency from image generation in a computer to projection [10]. This is realized by closely coordinating mirror control based on a DMD and the on/off blinking of a high-brightness LED and can exceed the speed achievable using DMD control alone. This high-speed projector is expected to be a key technology for the next high-speed 3D sensing. As the other approach to improve the projection speed, a system consisting of a lamp and a rotating wheel with a fringe pattern is proposed [11]. As an accurate sinusoidal pattern cannot be projected in this system, the target scene is captured by two cameras, and the depth image is reconstructed based on the temporal correlation technique. In the reported system, the camera frame rate was 12,000 fps, the used images for a single depth image was nine, and the depth image acquisition rate was 1333 fps, although the calculation was considered to be done offline because the used camera could not transfer images to a computer in real time.

Even when the frame rates of the camera and projector are significantly improved, it is hard to meet the conditions for precise 3D sensing of moving objects using the multi-shot structured-light method where the target can be approximated as static. To solve this problem, motion compensation methods are proposed, although the achieved speeds are not high because these methods involve additional calculation costs.

For example, in one proposed method, the boundary of a binary stripe pattern, where the white and black parts are reversed, is tracked in the camera image, and the

shape is obtained at 60 fps with less noise caused by the target motion [12]. Also, there is a system that achieves 3350-fps 3D sensing with motion compensation [13]. This system assumes that the motion between successive frames is small. Based on this assumption, the code information is propagated to neighboring pixels such that the collected codes originate from the same position on the target surface. Although the processing is done offline, the paper mentions that there is a possibility of real-time processing. In another method, the motion is compensated by assuming that the motion in the camera image is uniform and can be obtained from the speed of the centroid of the whole target; with this approach, 500-fps real-time 3D sensing was achieved [14]. A similar approach is also employed in Ref. [15]. In this system, a 15×15 dot pattern is projected, and the size of each dot is modulated at a different timing. By tracking dots through captured images, the dots are identified based on their unique modulations, and the 3D positions are obtained at 1000 fps.

These are techniques using binary-pattern structured light. There is also a technique using a gradation pattern. For example, in the phase-shift method, assuming that in a local region of the camera image the motion has a constant velocity and the shape is planar, motion compensation is achieved at 17 fps [16].

2.2 3D sensing with single-pattern projection

A solution to the noise problem caused by motion is to reduce the number of used structured-light patterns to just one and reconstruct the shape from a single camera image. In this approach, the pattern must have embedded clues to solve the problem of searching for correspondence between the camera image and the projector image using spatial features or brightness features [17]. In general, however, the resolution of the obtained shape becomes higher, the projected pattern needs to be dense, and the correspondence search becomes difficult and needs a large amount of calculations. In order to realize high-speed 3D sensing, this problem needs to be solved.

For example, there is a method using a single fringe pattern projection [18]. The 3D data is reconstructed based on a Fourier method. Although this method can provide a high-resolution depth image in a single captured camera image, the slopes on the target surfaces are limited because the calculation is based on the frequency analysis [19]. High-speed sensing system based on this method is reported in Ref. [20]. The reported performance was 4000 fps for 3D acquisition. However, the calculation was

considered to be done offline because the used camera was specialized only for recording.

As an other approach, a pattern in which spots are randomly placed is a feasible way to solve the correspondence search. In Ref. [21], by applying efficient block matching between the camera and projector image based on spatial correlation technique, the depth image is reconstructed at 100 fps. Also, a more efficient calculation method combined with machine learning techniques has been proposed, and this approach achieved 3D shape sensing at 375–1000 fps [22]. However, in this approach, 'blocky' noise is difficult to avoid, and smooth filtering is needed to remove such noise.

Another high-speed system utilizes parallel processing and an efficient algorithm for the correspondence search [23]. In this system, a 33×33 multi-spot pattern is projected, and each position in a camera image is calculated using a specialized parallel processing module. Assuming that motions of the spots are small between successive camera images, the correspondence search is efficiently solved by combining the correspondence search with spot tracking processing. The 3D positions of the spots were obtained at around 1000 fps. A demonstration video is shown in Ref. [24]. Also, a system for 200-fps 3D sensing using a multi-spot pattern has been developed [25]. In this system, instead of using time-sequential information like above, a spatial geometry constraint, which limits the measured range to within about 1.0 m, is used to perform the correspondence search. In another approach, a system consisting of a projector and two cameras has been proposed [26]. By projecting a well-designed pattern called a segmented pattern and by using three-viewpoint epipolar constraints, the system efficiently solves the correspondence search problem. Although the performance was reported as 500 fps in Ref. [26], the latest demonstration has demonstrated a 1000-fps 3D shape sensing [27]. Figure 1 shows a demonstration of the system in which a rigid body in the form of a bunny and a fluttering flag were captured. On the other hand, another proposed special active 3D sensing method is occlusion-robust 3D sensing [29]. Conventionally, active 3D sensing systems do not work well when other objects get between the measurement target and the measurement equipment, occluding the line of sight. The proposed system solves this problem using a light field created using aerial imaging. Figure 2 shows the developed system. Even if the top of the measuring device is messy and even if the user's hand takes a complex form, the 3D position and the inclination of his or her palm are obtained. A demonstration video of the sensing at around 100 fps is shown in Ref. [31].

Another approach beside triangulation using a projector and a camera is the time-of-flight method for obtaining the depth from the time taken for light emitted from the sensor to reach the target, be reflected at the target, and return to the sensor. Recently, a sensor has been developed to obtain a depth image without scanning the light with mirrors. However, it is difficult to improve the speed because, as the light receiving time becomes shorter, the depth accuracy becomes worse. At present, 50- to 160-fps sensors are commercially available [32].

3 Applications of high-speed 3D sensing

3.1 Surface reconstruction from time-sequential depth images

Besides the ability of high-speed sensing to observe high-speed moving objects, another nonobvious advantage is that a performance level that exceeds the speed of the target dynamics can contribute to an enhancement in spatial resolution. In this application, by moving the sensing system or observing a moving target, the system can acquire multi-view depth images composed of a number of frames acquired at different depths and from
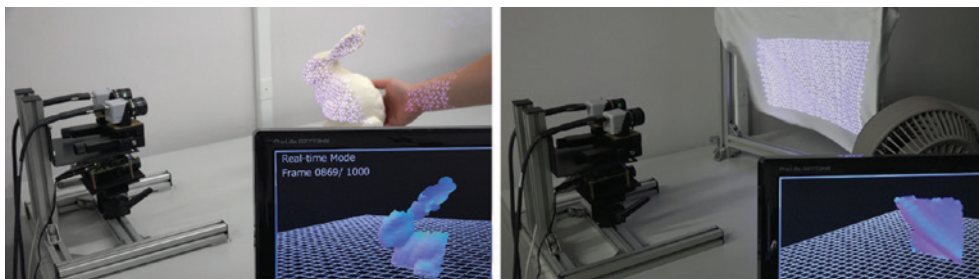

Figure 1: High-speed 3D sensing with three-view geometry using a segmented pattern [28].
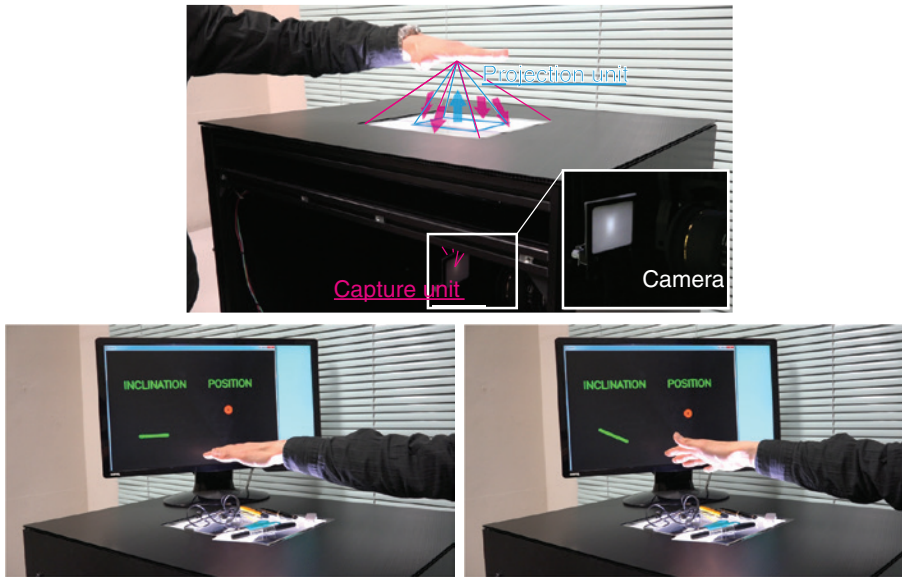
Figure 2: Occlusion-robust 3D sensing using aerial imaging [30].

different viewpoints. Those images are integrated into a single surface based on the 3D motion of the target. Compared with a single depth image obtained by a wide-field-of-view depth sensor, this type of application is highly promising because it can achieve high resolution with less noise, and a shape can be generated from all directions. Moreover, using high-speed 3D sensing, it is possible to capture many images within a certain time, to enhance the resolution even when the target moves at high speed, and to finish the acquisition of detailed all-around 3D modeling and wide-range environment in a short time.

This application is normally applied to a single rigid body and requires two tasks: alignment of depth images with motion estimation and surface reconstruction. In general, the obtained depth image is degraded through sensing, and the spatial sampling intervals on the target surface are non-uniform. Also, the features in the shape need to be effectively used for alignment. These problems could make this application difficult. For example, in order to enhance the accuracy of alignment, the depth images are aligned to the final reconstructed shape [33], color camera images are combined [34], and simultaneous estimation of alignment and surface reconstruction is incorporated [35, 36]. For the reconstruction, the surface is represented with voxels and as an implicit surface, and the depth images are approximately averaged to estimate the final shape [33, 37]. In another method, the surface is represented as a continuous domain and as an implicit surface using a shape representation proposed in the field of computer graphics [35, 36].

However, almost these methods perform the reconstruction using low-speed 3D sensing. The effectiveness of the combination with such method and high-speed 3D sensing is shown in Refs. [35, 36]. In this works, depth images obtained at a high speed, namely, 500–1000 fps, are used, and this method works even with low-resolution input images capturing a high-speed moving target as shown in Figure 3.

These types of techniques are mainly based on surface matching and image feature matching for the alignment task. Therefore, they involve a high computational cost, and the accuracy highly depends on the target conditions. In particular, when the target has no texture on the surface and no distinctive structural features, the method inevitably fails.

There is a system that incorporates 3D motion sensing for this application, namely, rotation and translation in case of a rigid body, instead of the image-based motion estimation [38]. In order to avoid the problems arising in the image-based approach and realize high-speed and contactless 3D motion sensing, a laser-based 3D rigid-body motion sensor has been newly developed and used for surface reconstruction from time-sequential depth images. The 3D motion sensing system consists of a laser rangefinder, a laser Doppler velocimeter, and a beam control unit. The system also introduces formularization and linearization techniques for deriving the motion velocity of a moving target from fragmentary information, such as speed and distance. The system can work at 410 Hz. A demonstration video is shown in Ref. [39]. The developed system and example reconstructed results are
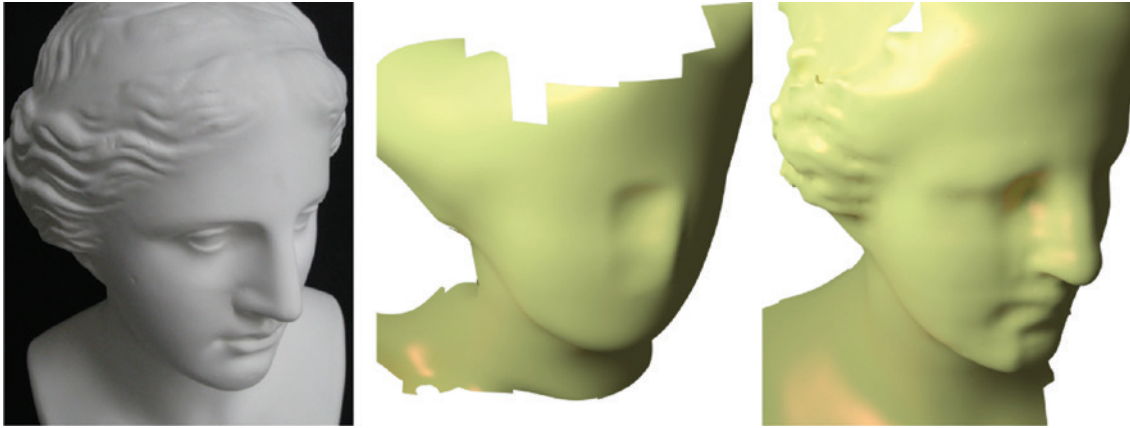
Figure 3: High-resolution surface reconstruction from multiple depth images [35]. (Left) Picture of original surface of the model. (Center) Surface reconstructed using a single depth image. (Right) Surface reconstructed by proposed method using 30 depth images.

shown in Figure 4. In this system, the color texture of the target also can be integrated [41].

3.2 High-speed 3D user interaction

High-speed 3D sensing is expected to be a key technology for allowing intuitive and natural operation of human-machine interfaces because it can acquire information that is much more useful than the information that can be input via a keyboard, mouse, or a touch panel display and can provide low-latency response, allowing natural and stress-free interactions. In particular, recognition of in-air hand gestures is one promising application. Example high-speed systems have achieved recognition at around 200 fps [42] and 500 fps [43], which is high enough for games, computer operation, and 3D displays and virtual reality.

High-speed 3D sensing is useful not only for observing inputs but also for controlling the display of objects on arbitrary surfaces. This point is very important for shape changing tangible interfaces [44–51]. In these interfaces, the display surface can be deformed, and this deformation can be detected and used as an input operation.

Although this type of display is expected to open up new paradigms, it is difficult to obtain rich 3D information at high speed. A system that shows promise in solving this performance problem is called the Deformable Workspace [52]. This display is designed to allow 3D manipulation of a virtual object. A deformable screen is used for both the display and the input, and the images are displayed based on rear projection. 3D sensing at about 1000 fps is used to obtain 3D deformation of the screen. The image distortion caused by the screen deformation is adaptively

compensated in real time based on the obtained 3D surface information. As a result, the user feels that the boundary between the real and virtual worlds disappears and that he or she can naturally manipulate the virtual object because the visual perception through the display and the somatic sensation through the input are virtually consistent. Demonstrations including 3D motion control of a virtual object, 3D modeling, and volume slicing of a 3D object are shown in Ref. [53]. Figure 5 shows the developed system and the demonstration of 3D motion control.

3.3 High-speed digital archiving

High-speed 3D sensing can also contribute to digital archiving. In particular, book digitization technologies are in high demand in a wide range of fields, from professional use in public work and business to personal use. The needs include archiving of library collections, document conservation in companies like publishers, disaster control in the event that books are lost due to natural hazards other than aging deterioration, and simple and convenient copying of personal notes. However, the book scanning speed has not been high enough, limiting the progress of those applications.

Book Flipping Scanning [55] is a new technological concept designed to improve the speed of book digitization, allowing any user in any field to digitize books 10 times faster than the conventional approaches. Conventional book digitization is basically based on opening a book and scanning it while keeping the pages stationary. High-speed 3D sensing is a key technology that allows a significant departure from this traditional style. Book Flipping Scanning allows scanning of a book while the pages
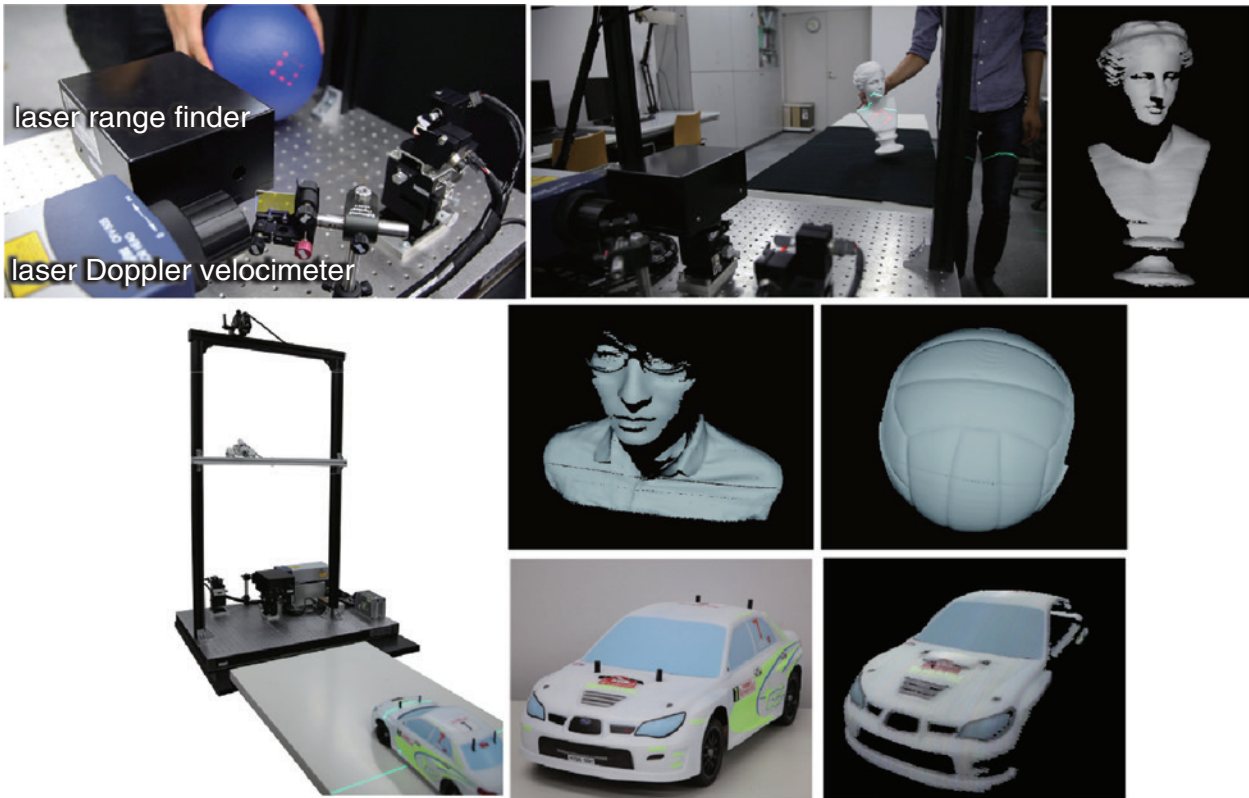
Figure 4: High-resolution shape and color integration of dynamic rigid body using 3D motion sensing system [40, 41].
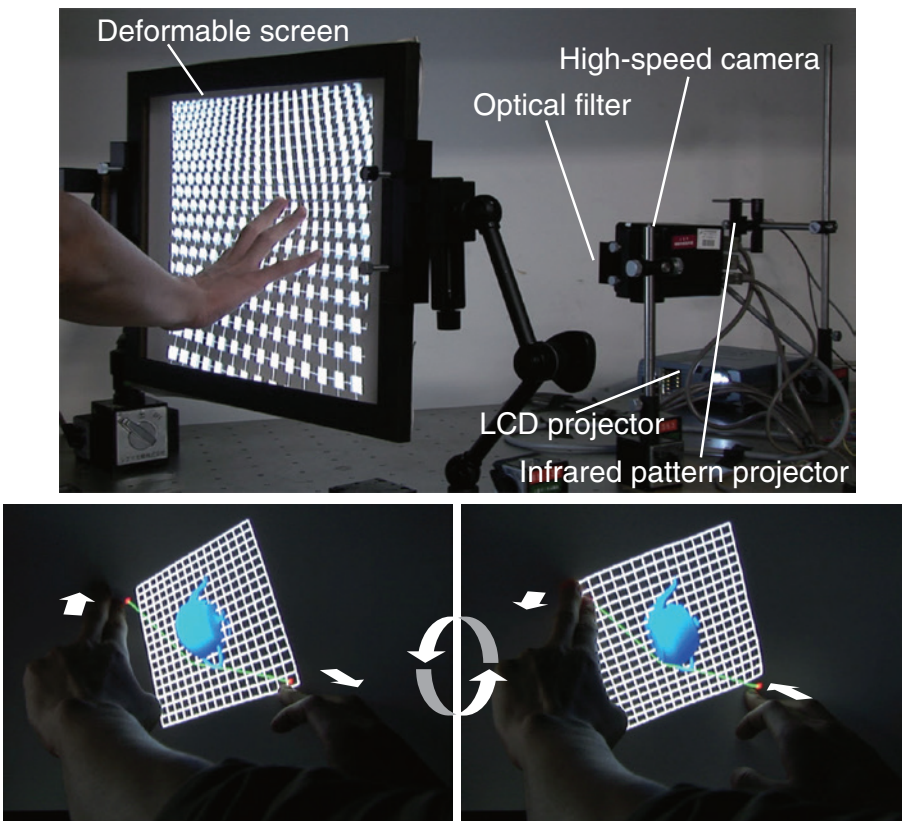

Figure 5: The Deformable Workspace [54]. (Top) Developed system. (Bottom) 3D motion control of a virtual object.

are being flipped, that is, while the pages are dynamically moving. This style of scanning enables high-speed, simple, convenient, and nondestructive book digitization. The speed that can potentially be reached is limited only by the speed of the page flipping.

The technical challenges to be overcome in realizing this concept include high-speed automatic page flipping, high-speed, high-resolution image capturing, and techniques for rectifying the deformed document images. High-speed 3D sensing offers effective solutions to the last two challenges, in particular.

There have been various systems based on Book Flipping Scanning, according to a diverse range of needs. One system is BFS-Auto [56], which is capable of improving three aspects of performance: speed, resolution, and automation. Figure 6 shows the developed system. A demonstration video is shown in Ref. [58]. BFS-Auto includes a high-speed page-turning machine for turning the pages in a contactless manner by utilizing the elastic force of the paper and an air blast [59]. The system can digitize 250 pages per minute at around 500 dpi. BFS-Auto also includes a novel high-speed sensing module consisting of high-speed 3D sensing and high-resolution cameras. High-speed 3D sensing observes the 3D deformation of a page at 500 fps while the page is being flipped. BFS-Auto recognizes the best timing at which each page can be imaged with the highest quality by a high-resolution camera. At this timing, the high-resolution camera captures the document image adaptively. The obtained 3D shape is segmented into a page region, which is then tracked time sequentially. Therefore, each page is captured only once

by the high-resolution camera. This adaptive imaging configuration enables high speed and high resolution in book digitization applications. This type of system architecture in which high-speed 3D sensing controls the sampling of other sensing devices and realizes flexible and high performance is considered to be effective for other applications, too.

As the captured document image is distorted because of the page deformation, it needs to be rectified. Typical rectification methods utilize lines of text. The rectification can be described using the assumption that the lines of text are horizontal in a flat document. However, the extraction of lines of text is still challenging in various page layouts in which text and figures are mixed [60–62]. Instead, using the 3D deformation of the document, the rectification problem can be described without depending on the content of the pages. BFS-Auto corrects the distorted document image by estimating a developable surface [63, 64]. The developable surface can be developed to a unique plane because the Gaussian curvature equals identically zero over the surface. Paper is a special non-rigid surface and meets this surface condition.

With the aim of achieving higher resolution, a system including a multi-camera array has been proposed [64, 65]. A page is divided into multiple ranges that are captured by multiple cameras, respectively. High-accuracy integration with the non-uniform rectification required for these input images has also been proposed. There are two types of systems: one in which the camera array is formed of high-speed cameras [64], and one in which the camera array is formed of high-resolution cameras with adaptive

Figure 6: BFS-Auto: High speed and high definition book scanner [57].

capturing [65]. Also, a rectification method for more complicated deformation has been proposed in Ref. [66]. In addition, as a personal-use system, BFS-Solo [67, 68] has been proposed. BFS-Solo uses video captured by a single camera for Book Flipping Scanning. Moreover, the real purpose of digital archiving is to store real-world information in a digital form that is sufficiently complete to allow reconstruction of the original physical object in the future. In this aspect, present book digitization technology is not good enough because it focuses mainly on conservation of the readable characters and illustrations. To scan all information about the book, such as the texture and glaze of the paper and ink, high-speed and efficient measurement of the reflectance distribution on the target surface by an algebraic solution based on adaptive illumination has been proposed [69, 70].

4 Conclusion

This paper described high-speed 3D sensing technologies, with a focus on contactless, active and real-time types. Three promising applications opened up by this type of sensing were introduced, including surface reconstruction from time-sequential depth images, high-speed 3D user interaction, and high-speed digital archiving. This paper also shows that, with the rise of these emerging applications, new high-speed sensing techniques for sensing 3D motion and reflectance distributions are being developed.

References

[1]   F. Blais, *J. Electron. Imaging* 13, 231–243 (2004).
[2]   R. Sagawa, H. Kawasaki, S. Kiyota and R. Furukawa, in 'Proceedings of IEEE International Conference on Computer Vision', (2011), pp. 1911–1918.
[3]   Microsoft Xbox 360 Kinect. Available at: http://www.xbox.com/.
[4]   Y. Watanabe, H. Oku and M. Ishikawa, Opt. Rev. 21, 875–882 (2014).
[5]   S. Yoshimura, T. Sugiyama, K. Yonemoto and K. Ueda, in 'International Solid-State Circuits Conference', (2001), pp. 94–95.
[6]   Y. Oike, M. Ikeda and K. Asada, IEEE Trans. Electron Devices 50, 152–158 (2003).
[7]   H. Gao, T. Takaki and I. Ishii, in 'SPIE', (2012), pp. 84370J.
[8]   S. J. Koppal, S. Yamazaki and S. G. Narasimhan, Int. J. Comput. Vis. 96, 125–144 (2012).
[9]   M. Schaffer, M. Grosse and R. Kowarschik, Appl. Opt. 49, 3622–3629 (2010).
[10] Y. Watanabe, G. Narita, S. Tatsuno, T. Yuasa, K. Sumino and M. Ishikawa, in 'The International Display Workshops', (2015), pp. 1064–1065.
[11] S. Heista, P. Lutzkea, I. Schmidta, P. Dietricha, P. Kuhmstedta, et al. Opt. Lasers Eng. 87, 90–96 (2016).
[12] S. Rusinkiewicz, O. Hall-Holt and M. Levoy, in 'Proceedings of Computer Graphics and Interactive Techniques', (2002), pp. 438–446.
[13] J. Takei, S. Kagami and K. Hashimoto, in 'Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems', (2007), pp. 3211–3216.
[14] Y. Liu, H. Gao, Q. Gu, T. Aoyama, T. Takaki, et al. J. Robotic. Mech. 26, 311–320 (2014).
[15] J. Chen, Q. Gu, H. Gao, T. Aoyama, T. Takaki, et al. in 'Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems', (2013), pp. 2683–2688.
[16] T. Weise, B. Leibe and L. V. Gool, in 'Proceedings of IEEE Conference on Computer Vision and Pattern Recognition', (2007).
[17] J. Salvi, J. Pages and J. Batlle, Pattern Recognit. 37, 827–849 (2004).
[18] M. Takeda and K. Mutoh, Appl. Opt. 22, 3977–3983 (1983).
[19] X. Su and W. Chen, Opt. Lasers Eng. 35(5), 263–284 (2001).
[20] Y. Gong and S. Zhang, Opt. Express 18, 19743–19754 (2010).
[21] M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C. Zach, et al. ACM Trans. Graph. 33, 156 (2014).
[22] S. R. Fanello, C. Rhemann, V. Tankovich, A. Kowdle, S. O. Escolano, et al. in 'Proceedings of IEEE Conference on Computer Vision and Pattern Recognition', (2016).
[23] Y. Watanabe, T. Komuro and M. Ishikawa, in 'Proceedings of IEEE International Conference on Robotics and Automation', (2007), pp. 3192–3197.
[24] Real-time Shape Measurement. Available at: http://www.youtube.com/watch?v=JAochkh-52s.
[25] M. Tateishi, H. Ishiyama and K. Umeda, in 'Proceedings of IEEE International Conference on Robotics and Automation', (2008), pp. 3022–3027.
[26] S. Tabata, S. Noguchi, Y. Watanabe and M. Ishikawa, in 'Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems', (2015), pp. 3900–3907.
[27] High-speed 3D Sensing with Three-view Geometry using a Segmented Pattern. Available at: https://www.youtube.com/watch?v=WQKMVAO4O48.
[28] High-speed 3D Sensing with Three-view Geometry using a Segmented Pattern. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/SegmentedPattern/.
[29] M. Yasui, Y. Watanabe and M. Ishikawa, in 'Proceedings of IEEE International Conference on Computational Photography', (2016).
[30] Occlusion-Robust 3D Sensing Using Aerial Imaging. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/Aerial3D/.
[31] Occlusion-Robust 3D Sensing Using Aerial Imaging. Available at: https://www.youtube.com/watch?v=vWN7ltKhlTA.
[32] BLUETECHNIX. Available at: http://www.bluetechnix.com/.
[33] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, et al. in 'Proceedings of the 24th annual ACM symposium on User interface software and technology', (2011), pp. 559–568.
[34] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, et al. in 'Proceedings of IEEE International Conference on Robotics and Automation', (2012).
[35] S. Noguchi, Y. Watanabe and M. Ishikawa, IPSJ Trans. CVA. 5, 143–152 (2013).
[36] Y. Watanabe, T. Komuro and M. Ishikawa, in 'IEEE International Conference on Computer Vision', (2009), pp. 1787–1794.

[37] B. Curless and M. Levoy, in 'Proceedings of ACM SIGGRAPH', (1996), pp. 303–312.

[38] L. Miyashita, R. Yonezawa, Y. Watanabe and M. Ishikawa, ACM Trans. Graph. 34, 218:1–218:11 (2015).

[39] 3D Motion Sensing of any Object without Prior Knowledge. Available at: https://www.youtube.com/watch?v=rrvAHh3-4qU.

[40] 3D Motion Sensing of any Object without Prior Knowledge. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/3D_Motion/.

[41] High-resolution Shape and Color Integration of Dynamic Rigid Body Using 3D Motion Sensing System. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/3D_Integration/.

[42] Leap Motion. Available at: https://www.leapmotion.com/.

[43] M. S. Alvissalim, M. Yasui, C. Watanabe and M. Ishikawa, in 'Proceedings of International Conference on Advanced Computer Science and Information Systems', (2014), pp. 198–203.

[44] A. Cassinelli and M. Ishikawa, in 'Proceedings of ACM SIGGRAPH 2005 Emerging technologies', (2005).

[45] M. Coelho and J. Zigelbaum, Pers. Ubiquitous Comput. 15, 161–173 (2011).

[46] H. Ishii, C. Ratti, B. Piper, Y. Wang, A. Biderman, et al. BT Technol. J. 22, 287–299 (2004).

[47] T. Gründer, D. Kammer, M. Brade and R. Groh, in 'Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems – Workshop: Displays Take New Shape: An Agenda for Future Interactive Surfaces', (2013).

[48] G. M. Troiano, E. W. Pedersen and K. Hornbæk, in 'Proceedings of International Working Conference on Advanced Visual Interfaces', (2014).

[49] P. Punpongsanon, D. Iwai and K. Sato, Virtual Real. 19, 45–56 (2014).

[50] Y. Fujimoto, R. T. Smith, T. Taketomi, G. Yamamoto, J. Miyazaki, et al. IEEE Trans. Vis. Comput. Graph. 20, 540–549 (2014).

[51] J. Steimle, A. Jordt and P. Maes, in 'Proceedings of SIGCHI Conference on Human Factors in Computing Systems', (2013), pp. 237–246.

[52] Y. Watanabe, A. Cassinelli, T. Komuro and M. Ishikawa, in 'Proceedings of IEEE International Workshop on Horizontal Interactive Human Computer Systems', (2008), pp. 145–152.

[53] The Deformable Workspace. Available at: http://www.youtube.com/watch?v=X5ZDvlYOZiM.

[54] The Deformable Workspace: a Membrane between Real and Virtual Space. Available at: http://www.k2.t.u-tokyo.ac.jp/perception/DeformableWorkspace/.

[55] T. Nakashima, Y. Watanabe, T. Komuro and M. Ishikawa, in 'Adjunct Proceedings of 22nd Symposium on User Interface Software and Technology', (2009), pp. 79–80.

[56] S. Noguchi, M. Yamada, Y. Watanabe and M. Ishikawa, in 'Proceedings of IEEE Winter Conference on Applications of Computer Vision', (2014), pp. 137–144.

[57] BFS-Auto: High Speed & High Definition Book Scanner. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/BFS-Auto/.

[58] BFS-Auto: High Speed Book Scanner at over 250 pages/min. Available at: https://www.youtube.com/watch?v=03ccxwNssmo.

[59] Y. Watanabe, M. Tamei, M. Yamada and M. Ishikawa, in 'Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems', (2013), pp. 272–279.

[60] H. Cao, X. Ding and C. Liu, in 'Proceedings of International Conference on Document Analysis and Recognition', (2003), pp. 71–75.

[61] J. Liang, D. DeMenthon and D. Doermann, IEEE Trans. Pattern Anal. Mach. Intell. 30, 591–605 (2008).

[62] Y. Tian and S. G. Narasimhan, in 'Proceedings of IEEE Conference on Computer Vision and Pattern Recognition', (2011), pp. 377–384.

[63] Y. Watanabe, T. Nakashima, T. Komuro and M. Ishikawa, in 'Proceedings of International Conference on Pattern Recognition', pp. 197–200.

[64] Y. Watanabe, K. Itoyama, M. Yamada and M. Ishikawa, in 'Proceedings of Asian Conference on Computer Vision', (2013), pp. 394–407.

[65] Document Digitization and its Quality Improvement using a Multi-camera Array. Available at: http://www.k2.t.u-tokyo.ac.jp/vision/MultiBFS_boundary/.

[66] M. Hirano, Y. Watanabe and M. Ishikawa, in 'Proceedings of IEEE International Conference on Image Processing', (2014), pp. 2604–2608.

[67] BFS-Solo: High Speed Book Digitization using Monocular Video. Available at: https://www.youtube.com/watch?v=tCq32jhWz1Q.

[68] H. Shibayama, Y. Watanabe and M. Ishikawa, in 'Proceedings of Asian Conference on Computer Vision', (2013), pp. 350–364 (2013).

[69] L. Miyashita, Y. Watanabe and M. Ishikawa, in 'Proceedings of International Conference on 3D Vision', (2014), pp. 232–239.

[70] Rapid SVBRDF Measurement by Algebraic Solution Based on Adaptive Illumination. Available at: https://www.youtube.com/watch?v=gGv6SSrmJ-A.

Yoshihiro Watanabe
Graduate School of Information Science and Technology, University of Tokyo, Tokyo Japan
Yoshihiro_Watanabe@ipc.i.u-tokyo.ac.jp

Yoshihiro Watanabe received his BE, ME, and PhD (Information Science and Technology) degrees from the University of Tokyo, Japan, in 2002, 2004, and 2007, respectively. He is currently a lecturer at the Graduate School of Information Science and Technology of the University of Tokyo. His research interests include high-speed vision, computer vision, high-speed projection display, virtual/augmented reality, computer-human interaction, and digital archiving.