Journal of Electronic Imaging

JElectronicImaging.org

Capturing method for integral three-dimensional imaging using multiviewpoint robotic cameras

Kensuke Ikeya
Jun Arai
Tomoyuki Mishina
Masahiro Yamaguchi

Kensuke Ikeya, Jun Arai, Tomoyuki Mishina, Masahiro Yamaguchi, "Capturing method for integral three-dimensional imaging using multiviewpoint robotic cameras," *J. Electron. Imaging* **27**(2), 023022 (2018), doi: 10.1117/1.JEI.27.2.023022.

Capturing method for integral three-dimensional imaging using multiviewpoint robotic cameras

Kensuke Ikeya,[a,b,*] **Jun Arai,**[a] **Tomoyuki Mishina,**[a] **and Masahiro Yamaguchi**[b]
[a]Japan Broadcasting Corporation (NHK), Science and Technology Research Laboratories, Tokyo, Japan
[b]Tokyo Institute of Technology, School of Engineering, Department of Information and Communications Engineering,
Yokohama, Kanagawa, Japan

Abstract. Integral three-dimensional (3-D) technology for next-generation 3-D television must be able to capture dynamic moving subjects with pan, tilt, and zoom camerawork as good as in current TV program production. We propose a capturing method for integral 3-D imaging using multiviewpoint robotic cameras. The cameras are controlled through a cooperative synchronous system composed of a master camera controlled by a camera operator and other reference cameras that are utilized for 3-D reconstruction. When the operator captures a subject using the master camera, the region reproduced by the integral 3-D display is regulated in real space according to the subject's position and view angle of the master camera. Using the cooperative control function, the reference cameras can capture images at the narrowest view angle that does not lose any part of the object region, thereby maximizing the resolution of the image. 3-D models are reconstructed by estimating the depth from complementary multiviewpoint images captured by robotic cameras arranged in a two-dimensional array. The model is converted into elemental images to generate the integral 3-D images. In experiments, we reconstructed integral 3-D images of karate players and confirmed that the proposed method satisfied the above requirements. [DOI: 10.1117/1.JEI.27.2.023022]

Keywords: integral imaging; three-dimensional; multicameras; robotic camera; image capturing; modeling.

Paper 170989 received Nov. 16, 2017; accepted for publication Mar. 13, 2018; published online Apr. 11, 2018.

1 Introduction

Much research is being done on technology to enable a greater sense of presence through three-dimensional (3-D) imaging that can be viewed with the naked eye. We are researching integral 3-D technology, an autostereoscopic image technology,[1] for next-generation 3-D television. The integral 3-D method requires no special glasses to view 3-D images and provides both vertical and horizontal disparity. Integral 3-D imaging is used in a wide range of applications, such as 3-D movie displays,[2,3] extraction and tracking of objects,[4,5] conversion of images into holograms,[6–8] refocusing,[9–13] and user interfaces.[14]

In the display stage of the integral 3-D method, a 3-D image is generated using a lens array to reconstruct light rays. These light rays are equivalent to the light rays emitted from the subject in the capturing stage. An image corresponding to a lens is called an elemental image. The integral 3-D display device consists of a lens array set on the front of a high-resolution display or projector.[15–17] Light rays from the display pixel pass through the lens array to reproduce the set of light rays emanating from the subject.

Integral 3-D imaging capturing methods can be categorized into two types, i.e., with[18–26] and without[27–31] a lens array. In the capturing method with a lens array, an integral 3-D camera combines a lens array and cameras and captures the set of light rays emanating from the subject. The light rays are captured in real space, and an integral 3-D display can be produced in real time. However, the camera has a problem in that it is difficult to capture a distant subject. That is, to capture distant subjects of an appropriate size, large lenses are needed to control the depth of the subject.

The capturing method without a lens array uses multiviewpoint cameras. With this method, 3-D models or interpolated images between viewpoints are generated from images taken with multiviewpoint cameras, and virtual light rays emanating from the subject in virtual space are calculated. This method requires no large lenses for controlling the depth of subject and can capture distant subjects by adjusting the baseline between the multiviewpoint cameras. However, the method has a problem in that it is difficult to capture dynamic moving subjects by the camera operator making panning and zoom-in shots because the pose and view angle of the multiviewpoint cameras are fixed.

To solve this problem, we are researching multiviewpoint robotic cameras.[32,33] The multiviewpoint robotic cameras are controlled through a cooperative synchronous system composed of a master camera operated by a camera operator and other reference cameras. Here, the camera operator instructs the master camera to capture subjects, and the reference cameras automatically follow its operations and capture multiviewpoint images of the subjects. It is possible to capture a dynamic moving subject by the camera operator making panning and zoom-in shots. However, this method also has problems wherein the quality of the generated integral 3-D images is low because the cooperative control and settings of the multiviewpoint robotic cameras have not yet been adapted to integral 3-D imaging. To generate high-quality integral 3-D images, one has to capture high-definition multiviewpoint images of the region to be reproduced by the integral 3-D display and to capture in both the vertical

[*]Address all correspondence to: Kensuke Ikeya, E-mail: ikeya.k-ec@nhk.or.jp

and horizontal directions within the field of view. It is difficult in the previous cooperative control method for multiviewpoint robotic cameras to capture the region to be reproduced by the integral 3-D display because they aim at a gaze point fixed on the subject's position and the view angles of all the cameras are kept the same as those of the master camera. As shown in Fig. 1(a), if the reproduced region does not completely fall within the view angle, the red parts in the region do not appear in the multiviewpoint images and cannot be shown in the integral 3-D images because the region cannot be reconstructed as a 3-D model. In contrast, as shown Fig. 1(b), if the view angle is too wide, despite that the reproduced region falls within the view angle, the resolution of the reproduced region will be low as will be the quality of the integral 3-D image. Furthermore, as shown in Figs. 1(a) and 1(b), if the multiviewpoint robotic cameras are arranged only linearly in the horizontal direction, it is difficult for them to capture rays in the vertical direction and the integral 3-D images may lack parts of the scene or subject in the field of view, as shown in the red part.

We propose a new integral 3-D capturing method using multiviewpoint robotic cameras. In this method, as shown in Fig. 1(c), a cooperative control function enables the multiviewpoint robotic cameras to capture images at the narrowest view angle that does not lose any part of the object region to be reproduced by the integral 3-D display, thereby maximizing the resolution of the image. The multiviewpoint cameras are arranged in the horizontal and vertical directions, and 3-D models are generated from the images they capture. Elemental images and integral 3-D images with pan, tilt, and zoom camerawork by a camera operator can be generated. Experiments in which integral 3-D images of a karate scene were generated confirmed the effectiveness of the method.

We explain the proposed method in Sec. 2. The experiments and results are presented in Secs. 3 and 4, respectively. Section 5 summarizes this paper and briefly touches on future work.

2 Proposed Method

The subjects are captured by multiviewpoint robotic cameras, and 3-D models are generated from the multiviewpoint images. The 3-D models are then converted into elemental images to generate integral 3-D images. Each step in this process is explained below.

2.1 *Capturing Images Using Multiviewpoint Robotic Cameras*

In the previous method, it is difficult for the robotic cameras to capture the region to be reproduced by the integral 3-D display because the direction of each camera is controlled such that the gaze point is fixed at the subject position and the view angles of all cameras are the same as that of the master camera. Furthermore, it is difficult to capture the vertical ray information because the cameras are arranged linearly in the horizontal direction. By contrast, in the proposed method, the cooperative control function enables the multiviewpoint robotic cameras to capture images at the narrowest view angle that does not lose any part of the object region, thereby maximizing the resolution of the image. The cameras are arranged horizontally and vertically so that rays in the vertical direction are captured as well.

The method to regulate the region reproduced by the integral 3-D display in real space, that is, to form a truncated pyramid region reproduced by the display in real space by composing the depth reproduction range of the display to the capturing region by the camera operator, is as follows. As shown in Fig. 2, the position and scale of the reproduced region change according to the subject's position and view angle during capture. The scale ratio k for the reproduced region in real space is defined as

$$k = \frac{d \cdot \tan\left(\frac{\theta}{2}\right)}{\left\{\frac{W}{2} - \Delta \tan\left(\frac{\theta}{2}\right)\right\}}, \tag{1}$$

where W is the size of the integral 3-D image that shows the subject, Δ is the distance that the 3-D image of the subject

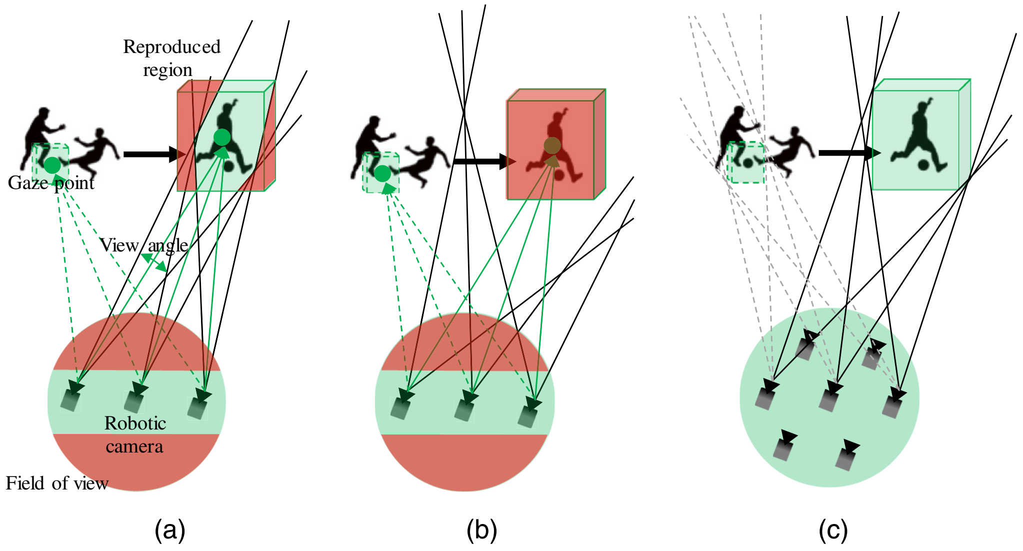

Fig. 1 Cooperative control and arrangement of multiviewpoint robotic cameras (a) previous method in the case of a narrow view angle, (b) previous method in the case of a wide view angle, and (c) proposed method.

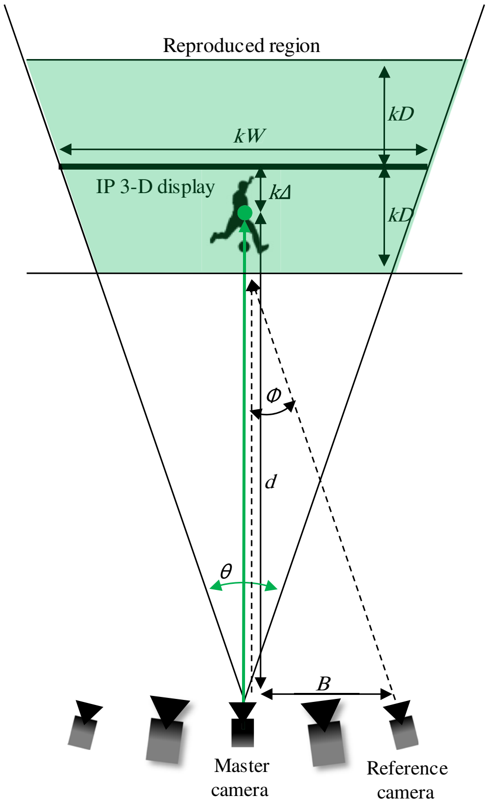

Fig. 2 Reproduced region of integral 3-D display and camera arrangement.

protrudes from the lens array attached to the display, d is the distance between the camera and subject, and θ is the view angle. These parameters multiplied by k show the reproduced region in real space (see Fig. 2).

The image capture process is as follows. First, six multiviewpoint robotic cameras are arranged in a regular hexagonal pattern and a seventh camera is set at the center of the hexagon, as shown in Fig. 3. The center camera is the master, and the other cameras are reference cameras. This camera arrangement enables rays in the vertical direction to be captured. In the proposed method, a 3-D model is generated by stereo matching using multiviewpoint images, which is explained in Sec. 2.2. Therefore, the baselines between neighboring cameras are optimized for stereo matching, and all of the baselines are the same distance. This camera arrangement enables us to capture within a wider field of view and with fewer cameras compared with other camera arrangements, e.g., a quadrangle, because the baselines between neighboring cameras can be set to the same distance. To decide the baseline, first, the subject's position and the view angle during capturing are guessed approximately. The distance between the camera and subject d_0 and view angle θ_0 at this time is substituted into Eq. (1), and the scale ratio k_0 is calculated. The convergence angle of the stereo cameras reaches a maximum at the nearest point from the master camera within the reproduced region. If the convergence angle is too wide, the corresponding point search fails in the stereo matching. The baseline is determined so that the convergence angle of the stereo cameras is below the limit of the convergence angle in stereo matching at the point. Denoting the limit of the convergence angle as Φ and the depth reproduction range as D, the base line B is defined as

$$B = \tan\Phi\{d_0 - k_0(D - \Delta)\}. \tag{2}$$

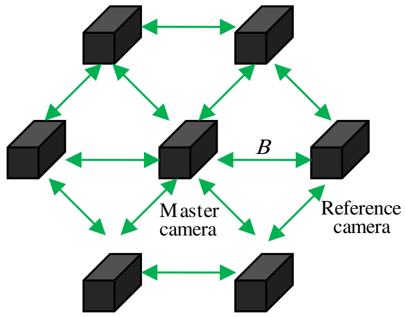

Fig. 3 Camera arrangement.

Next, the multiviewpoint robotic cameras are calibrated.[34] The conversion from world coordinates X to camera coordinates x is defined as

$$x_n = \begin{bmatrix} R_{0n} & t_{0n} \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}, \tag{3}$$

where the camera number of a robotic camera is denoted as n, the rotation matrix is denoted as R_0, and the translation vector is denoted as t_0. Figure 4 shows the cooperative control of multiviewpoint robotic cameras. The camera operator captures subjects by operating the pan, tilt, zoom, and depth of the master camera. The depth is the distance from the master camera to a point in 3-D space (gaze point) fixed

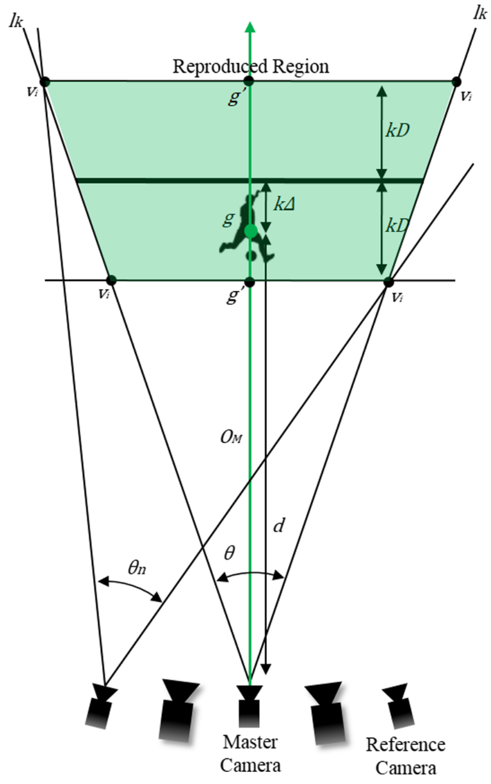

Fig. 4 Cooperative control of multiviewpoint robotic cameras.

by the pan, tilt, and depth operations. The operator uses the master camera to fix the gaze point on the subject position. Denoting the index for the master camera as M, the rotation matrix in operation as R, the pan and tilt values in camera calibration as P_0 and T_0, respectively, and the pan and tilt values in operation as P and T, respectively, the optical axes of the master camera o_M are defined as

$$R_M = R_{0M}^{-1} \begin{bmatrix} \cos(P_M - P_{0M}) & 0 & \sin(P_M - P_{0M}) \\ 0 & 1 & 0 \\ -\sin(P_M - P_{0M}) & 0 & \cos(P_M - P_{0M}) \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(T_M - T_{0M}) & -\sin(T_M - T_{0M}) \\ 0 & \sin(T_M - T_{0M}) & \cos(T_M - T_{0M}) \end{bmatrix}, \tag{4}$$

$$R_M = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \tag{5}$$

$$o_M = \begin{bmatrix} r_{13} \\ r_{23} \\ r_{33} \end{bmatrix}. \tag{6}$$

Denoting the depth as d, the gaze point g is defined as

$$g = -R_{0M}^{-1} t_{0M} + d \cdot o_M. \tag{7}$$

When the camera operator manipulates the depth, s/he needs to recognize whether the gaze point is adjusted to be on the subject's position. To do so, for example, there is a method in which another robotic camera is pointed toward the gaze point and the camera operator can recognize whether or not the gaze point is fixed on the subject's position from the subject's position in the video captured by the robotic camera.[32,33] Next, the reproduced region of the integral 3-D display is regulated in real space. Denoting the view angle as θ_M and the camera's horizontal and vertical aspect as w/h, the directions of the four lines constructing the capture region of the master camera $l_k (1 \leqq k \leqq 4)$ are defined as

$$l_k = R_{0M}^{-1} \begin{bmatrix} \cos(p_k) & 0 & \sin(p_k) \\ 0 & 1 & 0 \\ -\sin(p_k) & 0 & \cos(p_k) \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(t_k) & -\sin(t_k) \\ 0 & \sin(t_k) & \cos(t_k) \end{bmatrix} R_{0M} o_M,$$
$$p_k = \left\{ \frac{\theta_M}{2}, -\frac{\theta_M}{2}, -\frac{\theta_M}{2}, \frac{\theta_M}{2} \right\},$$
$$t_k = \left\{ \frac{\theta_M h}{2w}, \frac{\theta_M h}{2w}, -\frac{\theta_M h}{2w}, -\frac{\theta_M h}{2w} \right\}. \tag{8}$$

The near and far points at the intersection of the optical axis of the master camera $g'$are defined as

$$g' = \begin{cases} -R_{0M}^{-1} t_{0M} + (d - kD + k\Delta) \cdot o_M, & \text{near} \\ -R_{0M}^{-1} t_{0M} + (d + kD + k\Delta) \cdot o_M, & \text{far} \end{cases}. \tag{9}$$

The plane configuring the truncated pyramid region, which is perpendicular to o_M and includes g', is defined as

$$o_M X = o_M g'. \tag{10}$$

The four lines constructing the capture region of the master camera are defined as

$$X = -R_{0M}^{-1} t_{0M} + u \cdot l_k, \tag{11}$$

where the parametric is denoted as u. Accordingly, the eight vertices of the reproduced region in real space are regulated. The eight vertices $v_i (1 \leqq l \leqq 8)$ are defined as

$$v_i = -R_{0M}^{-1} t_{0M} + l_k \frac{o_M (g' + R_{0M}^{-1} t_{0M})}{o_M l_k}. \tag{12}$$

Then, using the cooperative control function of the pan, tilt, and zoom, the reference cameras can capture images at the narrowest view angle that does not lose any part of the reproduced region. Denoting the index for the reference camera as n, the vector o'_{ni} extending from the reference camera to v_i in camera coordinates is defined as

$$o'_{ni} = R_{0n} \frac{v_i + R_{0n}^{-1} t_{0n}}{\| v_i + R_{0n}^{-1} t_{0n} \|}. \tag{13}$$

The pan and tilt values for pointing toward v_i by the reference camera P_{ni} and T_{ni}, respectively, are defined as

$$o'_{ni} = [\, e_{ni1} \quad e_{ni2} \quad e_{ni3} \,]^{\mathrm{T}}, \tag{14}$$

$$P_{ni} = \tan^{-1}(e_{ni1}/e_{ni3}) + P_{0n}, \tag{15}$$

$$T_{ni} = \sin^{-1}(e_{ni2}) + T_{0n}. \tag{16}$$

Pan and tilt values for capturing of images at the narrowest view angle P_n and T_n are calculated using Eqs. (17) and (18), respectively, and the directions of the reference cameras are controlled with them

$$P_n = \frac{\max_i P_{ni} + \min_i P_{ni}}{2}, \tag{17}$$

$$T_n = \frac{\max_i T_{ni} + \min_i T_{ni}}{2}. \tag{18}$$

The narrowest view angle θ_n is calculated using Eq. (19), and the view angle of the reference cameras is controlled with it

$$\theta_n = \max\left[\max_i P_{ni} - \min_i P_{ni}, \frac{w}{h}(\max_i T_{ni} - \min_i T_{ni})\right]. \tag{19}$$

2.2 *Three-Dimensional Model Generation*

The 3-D model generation of the previous method has limited depth estimation accuracy because it utilizes only two multiviewpoint cameras arranged in the horizontal direction. Furthermore, the belief propagation method[35] used in the previous method for depth estimation taxes the memory resources of the PC, particularly in the case of high-definition systems. In contrast, the proposed method incorporates multibaseline stereo that utilizes information on seven multiviewpoint robotic cameras arranged in horizontal and vertical directions and cost volume filtering[36] to improve the accuracy of the depth estimation. These methods also enable high-definition multiviewpoint images to be used because they do not use up too much memory.

First, a camera calibration is done using the multiviewpoint images of the seven robotic cameras.[34] Then, the depth of each viewpoint image is estimated. As mentioned above, multibaseline stereo and cost volume filtering are utilized for the depth estimation. Six pairs of cameras, each pair being a camera targeted for depth estimation and another camera, are made. Rectification for parallel stereo is conducted for each camera pair. The cost E of assigning depth value labels to a pixel in the camera image targeted for depth estimation is defined as

$$E_{m,p,l} = \min\left[-\frac{\sum_{(i,j)\in R(p)}(I(i,j)-\overline{I})(I'_m(i+d(l),j)-\overline{I'_m})}{2\sqrt{(\sum_{(i,j)\in R(p)}(I(i,j)-\overline{I})^2)\left(\sum\limits_{(i,j)\in R(p)}(I'_m(i+d(l),j)-\overline{I'_m})^2\right)}}+0.5, T_E\right], \tag{20}$$

where m is the number of the camera pairs, p is the pixel being processed, l is the depth value label, $R_{(p)}$ is the set of pixels in a block around p, i and j are the indices of the pixel in the block, I is the pixel value of a camera targeted for depth estimation, I' is the pixel value of a camera paired to the target camera, $\overline{I}$ and $\overline{I}'$ are the corresponding averages of the pixel values in the block, $d(l)$ is the disparity corresponding to the depth value label, and T_E is the threshold of E. Equation (20) is based on zero-mean normalized cross correlation, which is robust to luminance differences between multiviewpoint images. A cost C that incorporates every pair's E is defined as

$$C_{p,l} = \frac{1}{6}\sum_m E_{m,p,l}. \tag{21}$$

A cost map is generated from the costs of all pixels at a depth value l that has been filtered. This cost is defined as

$$C'_{p,l} = \sum_q W_{p,q}(I)C_{q,l}, \tag{22}$$

where W is a filter and q is a pixel in the filter. A guided filter[37] is used for filtering, and I is the guidance image. The depth label l that has the minimum cost after filtering is assigned to pixel p. The depth f of pixel p is defined as

$$f_p = \mathrm{argmin}_l\, C'_{p,1}. \tag{23}$$

This depth estimation is done on the videos of all seven cameras.

Finally, the 3-D model is generated. Here, pixels that have high reliability in the depth estimation result are extracted; that is, pixels with a cost C' less than or equal to a certain threshold, T_C, are extracted. A 3-D point cloud model is generated by back projection of the pixels extracted from all cameras. Occlusions and regions of low reliability in the 3-D point cloud model of each viewpoint camera are interpolated complementarily using the 3-D point clouds of the other viewpoint cameras.

2.3 *Elemental Image Generation*

A camera operator captures scenes or subjects using pan, tilt, and zoom operations. Integral 3-D images reflecting this camerawork are generated by reproducing the camera, subject, and display in virtual space through elemental image generation. Here, the depth reproduction range is changed in proportion to the camera view angle in the camerawork, as shown in Fig. 2. For example, when a subject is captured by zooming in, both the size and sense of depth of the subject increase to the viewer.

The process of generating the elemental image is described below. First, the scale of the 3-D model is converted to the scale of the real subject, and the 3-D model in world coordinates is converted to the one in master camera coordinates using Eq. (3). Next, the display size is set to kW, and the display is set $d + k\Delta$ away from the master camera, where the optical axis of the master camera passes through the center of the display and the lens surface is vertical to the axis, as shown in Fig. 5. Then, the ray joining a pixel on the virtual display and the optical center of the lens corresponding to the pixel is traced. The color where the ray intersects the 3-D model is assigned to that pixel in the elemental image. The number of elemental images is the same as number of lenses in the array. In the implementation, rather than tracing each ray individually, the 3-D model is projected obliquely,[29] as shown in Fig. 5, to obtain all rays in a given direction at the same time. This obliquely projected

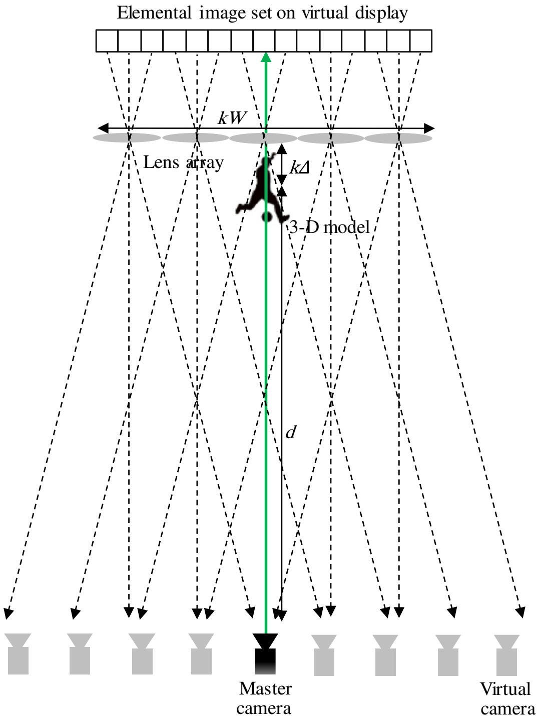

Fig. 5 Elemental image generation.

image only has some of the pixels in the elemental image, and these are reorganized into the elemental image set.

3 Experiments

The experiments were conducted to confirm that the cooperative control function enables the multiviewpoint robotic cameras to capture images at the narrowest view angle that does not lose any part of the object region and that integral 3-D images showing all of the object region can be generated from the captured multiviewpoint images.

3.1 *Verification of Cooperative Control of Multiviewpoint Robotic Cameras*

First, we verified the reproducibility of the cooperative control of the multiviewpoint robotic cameras in a simulation. Next, we developed a system in which the proposed method is implemented. Then, we verified the method's practicality in capturing experiments using the system.

3.1.1 *Simulation*

We developed a simulator to recognize how subjects and the reproduced region are captured by multiviewpoint robotic cameras controlled using the proposed method. The flowchart of the simulator is shown in Fig. 6.

Here, a newscaster was the subject and the multiviewpoint robotic cameras and 3-D models of the subject were arranged in virtual space, as shown in Fig. 7. We used the following parameters $\{W, D, \Delta, \Phi, d, \theta\} = \{243$ mm, 100 mm, 50 mm, 20 deg, 5000 mm, and 59 deg, respectively$\}$. The images of the reproduced region captured by the master camera and reference cameras are shown in Fig. 8. The center image was captured by the master camera, and the other images were captured by the reference cameras. The reproduced region enclosed by the blue lines cannot be observed in the image captured by the master camera because the reproduced region and view angle completely match. In contrast, the width or height of the reproduced region matches the width or height of the view angle in the reference images. This experimental result shows that the cooperative control function works as intended and thus can maximize the resolution of the image. The reproducibility of the proposed method is thus verified.

3.1.2 *System development*

We developed a system in which the proposed method is implemented. The system mainly consists of robotic cameras having a small HD camera mounted on an electrically powered dynamic pan–tilt platform with a board computer, an operation interface for the master camera to be operated by a camera operator, a recorder of multiviewpoint images, and a processor for 3-D model generation and elemental image generation. HD-SDI is utilized to transmit video data from the cameras to the recorder, and Ethernet is used for the multiviewpoint robotic cameras to communicate. The multiviewpoint robotic cameras are custom-made, and Fig. 9 shows the cameras. Figure 10 shows the operation interface, and Fig. 11 shows the system diagram.

3.1.3 *Capturing experiment*

We captured multiviewpoint images of subjects using the system based on the proposed method and synthesized the reproduced region from the captured images to verify the accuracy of direction and view angle control of the multiviewpoint robotic cameras in the presence of mechanical control errors. The subjects were karate players, and the multiviewpoint robotic cameras were arranged as shown in Fig. 12. We used the following parameters $\{W, D, \Delta, \Phi, d, \theta\} = \{293$ mm, 45 mm, 20 mm, 12 deg, 5000 mm, and 47 deg, respectively$\}$. We calibrated the

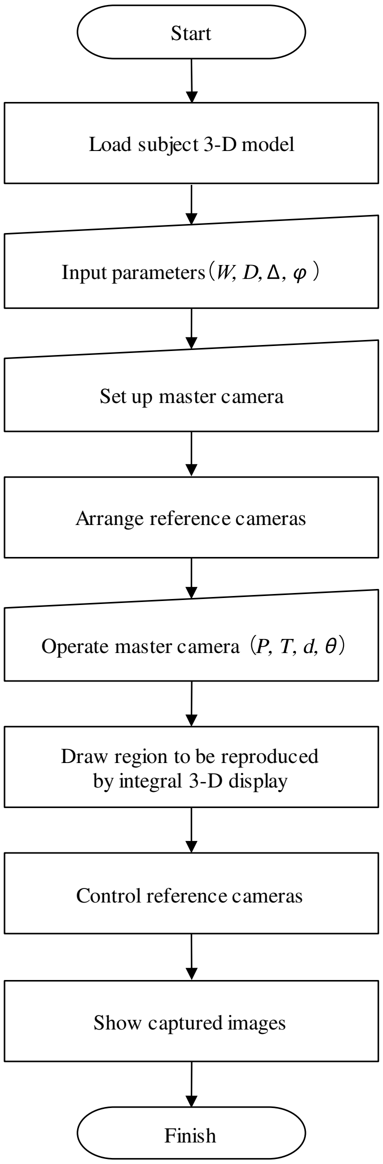
Start
Load subject 3-D model
Input parameters (W, D, Δ, φ)
Set up master camera
Arrange reference cameras
Operate master camera (P, T, d, θ)
Draw region to be reproduced by integral 3-D display
Control reference cameras
Show captured images
Finish

Fig. 6 Flowchart of simulator.

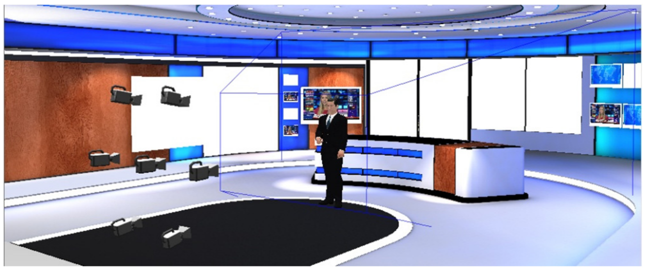

Fig. 7 Arrangement of multiviewpoint robotic cameras and subjects.

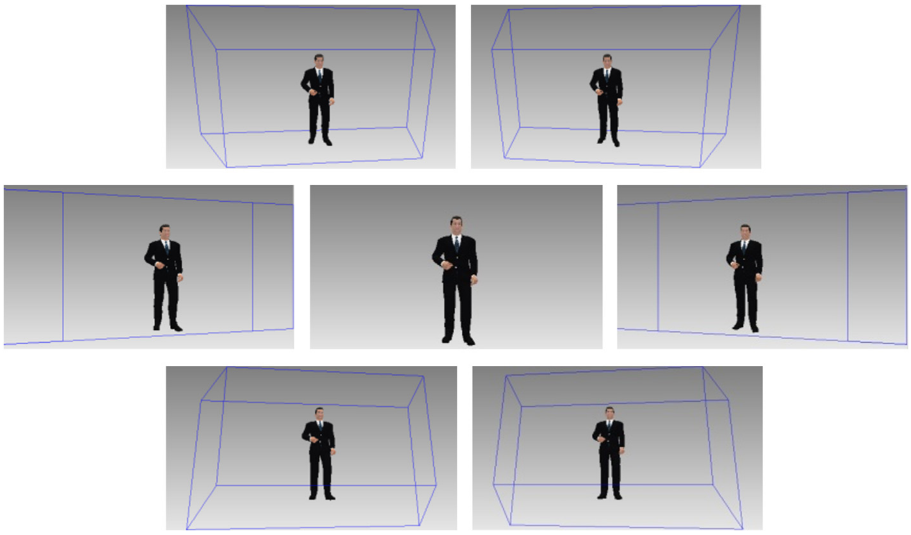

Fig. 8 Captured images of a reproduced region by integral 3-D display (the reproduced region is indicated by blue lines).

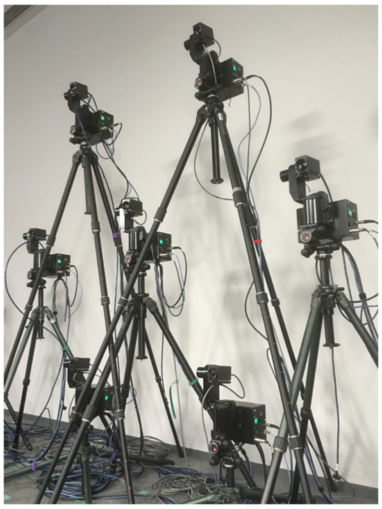

Fig. 9 Multiviewpoint robotic cameras.

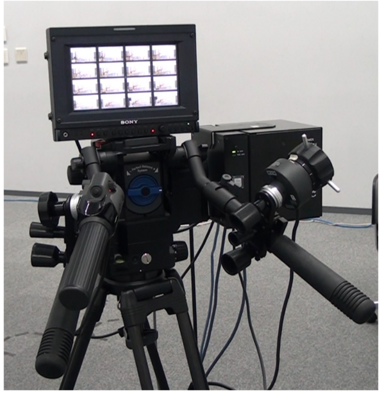

Fig. 10 Operation interface.

cameras using weak camera calibration.[34] The world coordinates of the reproduced region were calculated using the view angle of the master camera, the position of the gaze point, the size and depth reproduction range of the integral 3-D display, and the camera parameters obtained by the camera calibration. The resulting multiviewpoint images are shown in Fig. 13. The reproduced region and view angle did not match completely due to mechanical control errors, which is different from the simulated result in Sec. 3.1.1. We measured the position aberration in the images between the reproduced region and the view angle of each viewpoint camera and found that it was on average 44.2 pixels per image. In the previous method, the size of the reproduced region in these each image was vastly different in proportion to the distance between the cameras, and subjects especially when the distance was short. In contrast, the difference is much smaller in the current method, and this verifies its practicality.

3.2 *Integral Three-Dimensional Image Generation*

Integral 3-D images were generated from the multiviewpoint images captured in the experiment described in Sec. 3.1.3. In the depth estimation process for the 3-D modeling, we used the following parameters $\{R, T_E, T_C\} = \{9 \times 9,\ 0.2,\ 0.9,\ \text{respectively}\}$ and performed space sampling every 10 mm in the reproduced region. We used an integral 3-D display consisting of a high-resolution liquid-crystal panel and a lens array, as shown in Fig. 14, for this experiment. The parameters of the integral 3-D display are shown in Table 1. Images of the generated 3-D models captured from several viewpoints (front viewpoint, diagonally upward viewpoint, diagonally downward viewpoint) in virtual space are shown in Fig. 15. Elemental images are shown in Fig. 16, and two-dimensional images of integral 3-D images on the integral 3-D display taken by a digital camera are shown in Fig. 17. The integral 3-D images where motion parallax occurred on the display captured by a digital still camera from different viewpoints, i.e., a high angle, low angle, left angle, and right angle, are shown in Fig. 18. The experimental results show that the proposed method correctly generates 3-D models, elemental images, and integral 3-D images. We measured the processing time for 3-D model generation and elemental image generation. It took 2 min per frame for 3-D model generation and 30 s per frame for elemental image generation. Then, we verified whether

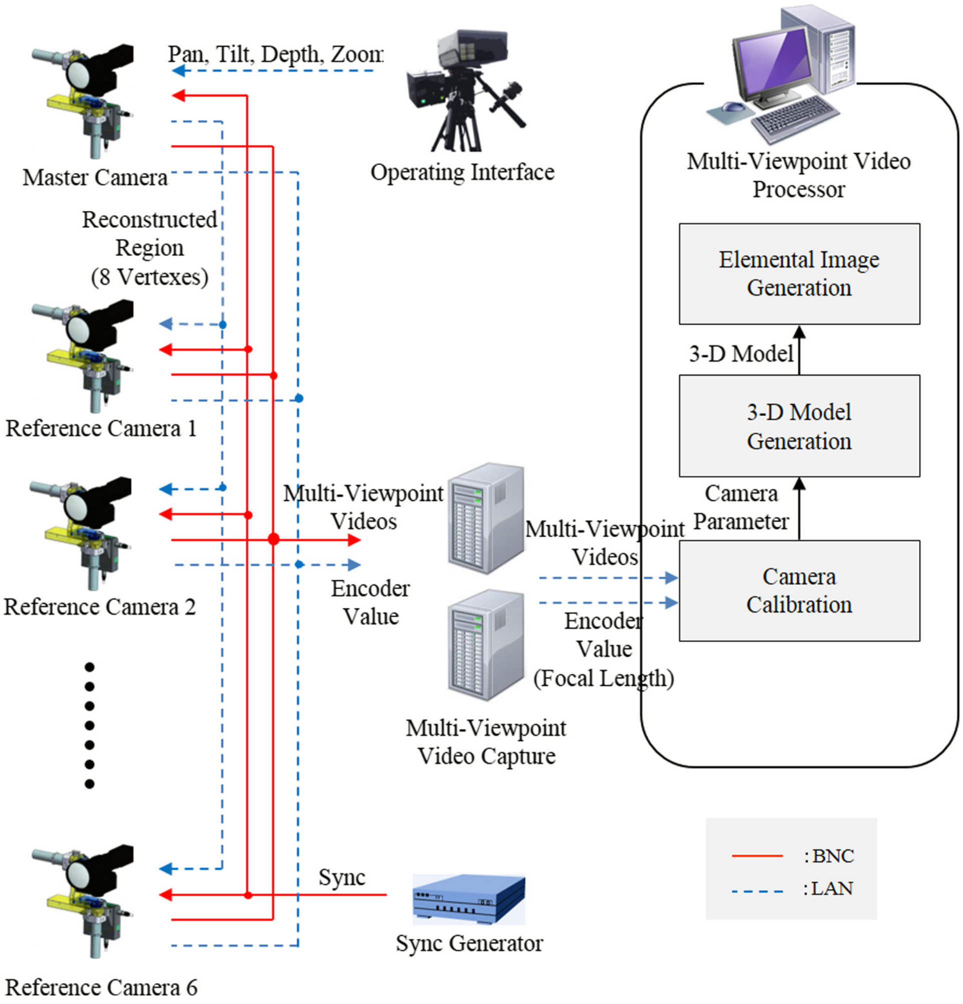

Fig. 11 System diagram.

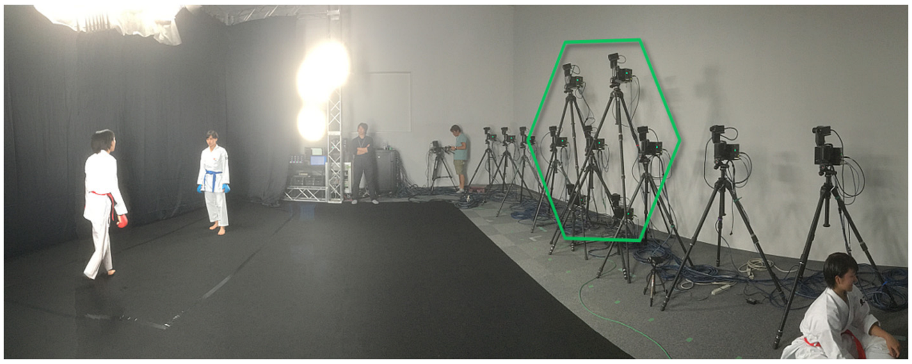

Fig. 12 Arrangement of multiviewpoint robotic cameras (cameras used in this experiment are within the green hexagon).

the positional relations of the cameras, subjects, and integral 3-D display in real space are reproduced in the elemental image generation process. In the scene depicted in the left row and second column of Fig. 17, the distance between the master camera and the subject is 5000 mm, the distance between the master camera and integral 3-D display is 5314 mm, and the view angle is 47 deg in real space. The height of the subject is 1239 mm, and the left knee of the subject, which is the most protruding part from the lens array surface, is positioned at x: −400 mm, y: −350 mm, and z: 139 mm when the origin is set at the center of the integral 3-D display. When the subject is shown on a display that has the parameters shown in Table 1, the size of the subject is 79 mm, and the left knee position is x: −27 mm, y: −25 mm, and z: 7mm. Accordingly, the calculated scale ratio of the real space to 3-D images is 15.73: 1. By comparison, the scale ratio of the measured size of the subject and the positions in real space and in the 3-D

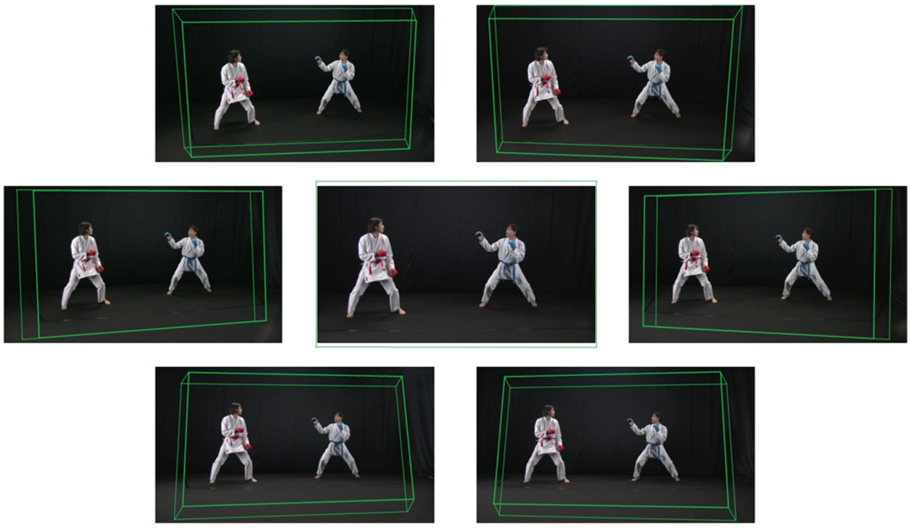

Fig. 13 Synthesis of region reproduced by the integral 3-D display from captured multiviewpoint images.

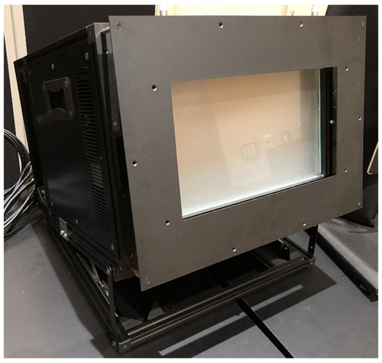

Fig. 14 Integral 3-D display.

Table 1 Parameters of integral 3-D display.

Number of lenses in array	294 × 191
Lens pitch	1.003 mm
Focal length of lens	1.74 mm
Display definition	7680 × 4320
Pixel pitch	0.03825
Display size	11 in. (horizontal 293 mm)

image were almost the same. Thus, the experiment proved that the correct positional relations of the cameras, subjects, and integral 3-D display in real space are reproduced in the 3-D images.

4 Discussion

Let us consider the performance of the cooperative control, which is a distinctive feature in our method. Figure 19 compares multiviewpoint images captured using common commercially available multiviewpoint cameras in which pan, tilt, and zoom are fixed and ones captured using cooperative control of multiviewpoint robotic cameras, together with the 3-D models constructed from them. Both multiviewpoint images were captured with seven HD cameras arranged in a hexagonal pattern, and the 3-D models were constructed using the 3-D model generation method explained in Sec. 2.2. When using common commercially available multiviewpoint cameras to generate integral 3-D images of dynamic moving shots such as of the subject's bust shown in Fig. 18, the multiview point images of the subject must be captured with a wide viewing angle so that the subject falls within that angle, as shown in Fig. 19(a). Figure 19(b) shows the 3-D model of the subject's bust shot constructed from those images, and it is clear that its surfaces are not completely formed because of insufficient 3-D point clouds. Although a 3-D model with all its surfaces could be formed by increasing the size of each point or by applying polygons, the resolution and sharpness of the resulting 3-D model and integral 3-D images would decrease. In contrast, as shown in the bust shot in Fig. 19(c), when using the cooperative control, the multiviewpoint images are captured in a way that maximizes the image resolution of the subject parts. Figure 19(d) shows the resulting 3-D model of the subject's bust shot constructed from these images, and it is clear that the surfaces of the 3-D model are completely formed because there are sufficient 3-D points in the point clouds. The 3-D model shown in Fig. 19(b) consists of 215,277 vertices, while the one in Fig. 19(d) has more than 10 times as many vertices, 2,166,035. Cooperative control of multiviewpoint robotic cameras is thus effective at generating high-quality 3-D models and integral 3-D images.

Now, let us consider the difference in performance of the proposed method and the previous one using cooperative control. When using cooperative control of multiviewpoint robotic cameras in the previous method, the subject might not always fall within the view angle of the reference cameras, as shown in Fig. 20(a). The subject parts that are not captured by the reference cameras cannot be reconstructed as 3-D models and cannot be shown in the integral 3-D images of the subject. In contrast, in the proposed method, integral

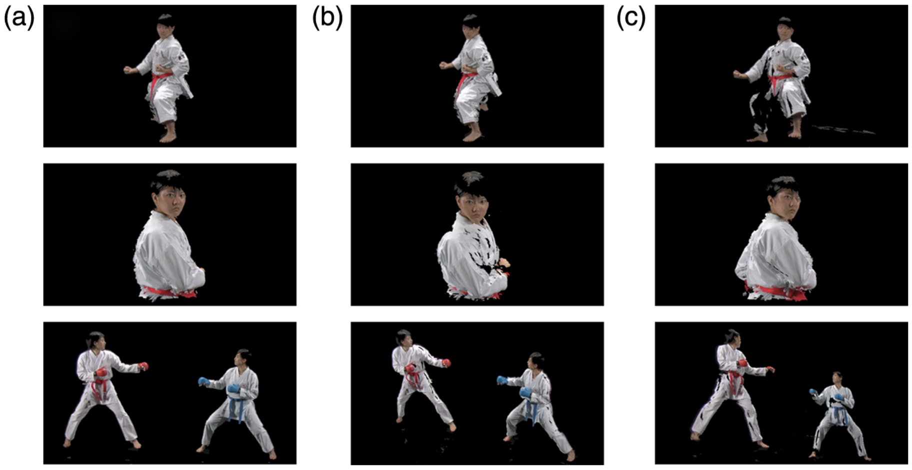

Fig. 15 Three-dimensional model (a) front viewpoint, (b) diagonally upward viewpoint, and (c) diagonally downward viewpoint.

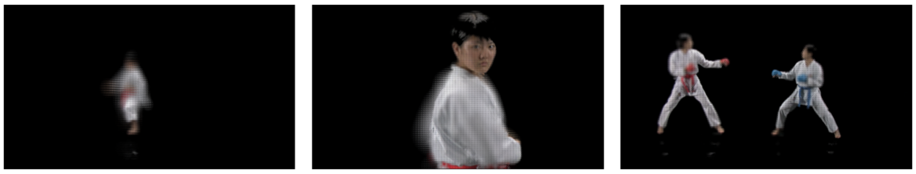

Fig. 16 Elemental images.

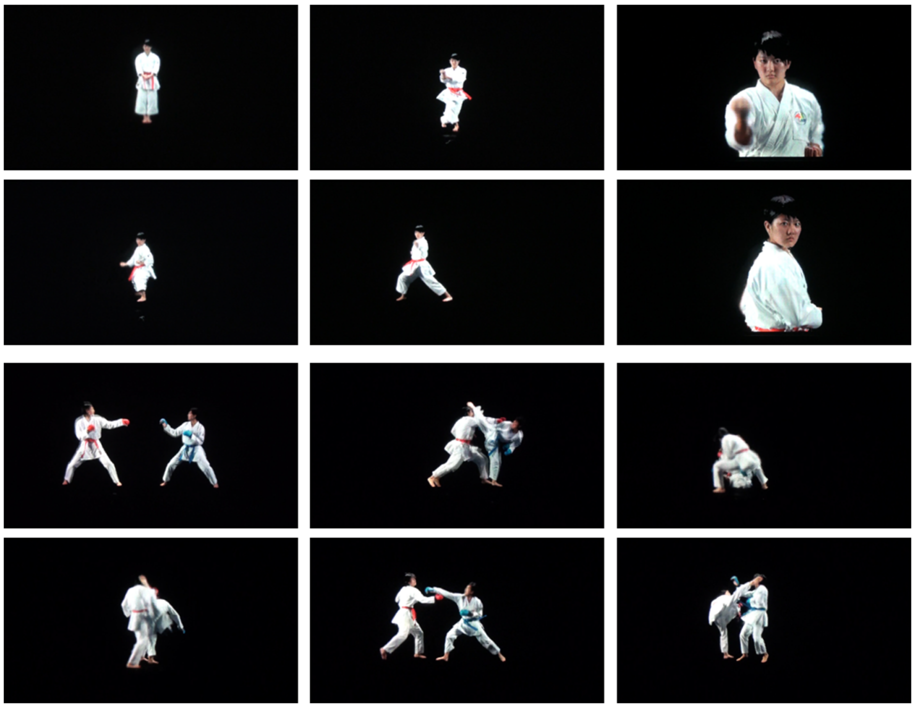

Fig. 17 Two-dimensional images of integral 3-D images on the integral 3-D display taken by a digital camera (Video 1, MPEG, 18.5 MB [URL: https://doi.org/10.1117/1.JEI.27.2.023022.1]).

3-D images are correctly produced because the subject falls within the view angle of the reference cameras by capturing the region to be reproduced by the integral 3-D display, as shown in Fig. 20(b).

Figure 21 compares the 3-D models constructed with the previous and proposed methods. In Fig. 21(a), a large hole appears in the player's body reconstructed by the previous method. Figure 21(b) shows the result of the proposed method, where the hole does not appear. The hole in the previous method is due to vertical rays not being able to be obtained because of the linear arrangement of multi-viewpoint robotic cameras in only the horizontal direction.

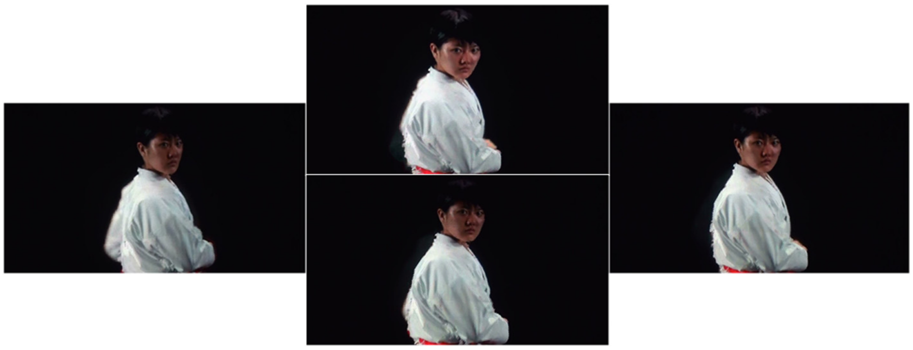

Fig. 18 Integral 3-D image where motion parallax occurred (viewpoints: left angle, high angle, low angle, and right angle).

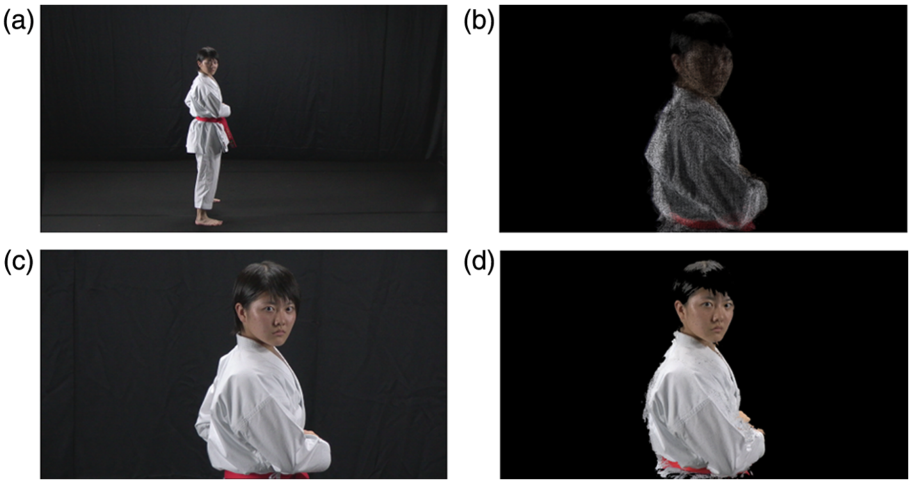

Fig. 19 Performance of cooperative control of multiviewpoint robotic cameras: (a) One of the captured multiviewpoint images using common commercially available multiviewpoint cameras, (b) 3-D models generated using common commercially available multiviewpoint cameras, (c) one of the captured multiviewpoint images using cooperative control of multiviewpoint robotic cameras, and (d) 3-D models generated using cooperative control of multiviewpoint robotic cameras.

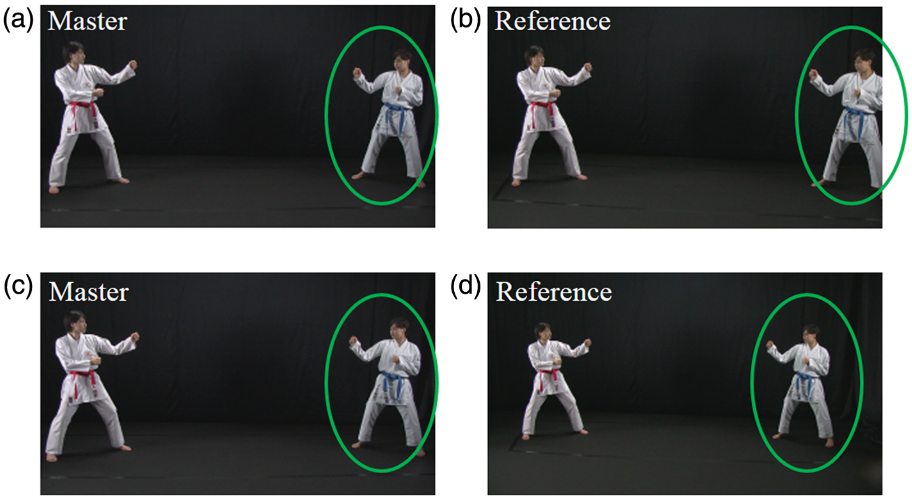

Fig. 20 Images captured using cooperative control of multiviewpoint robotic cameras: (a) When the previous method is used, the subject does not fall within the view angle of the reference cameras. (b) When the proposed method is used, the subject falls within the view angle of the reference cameras.

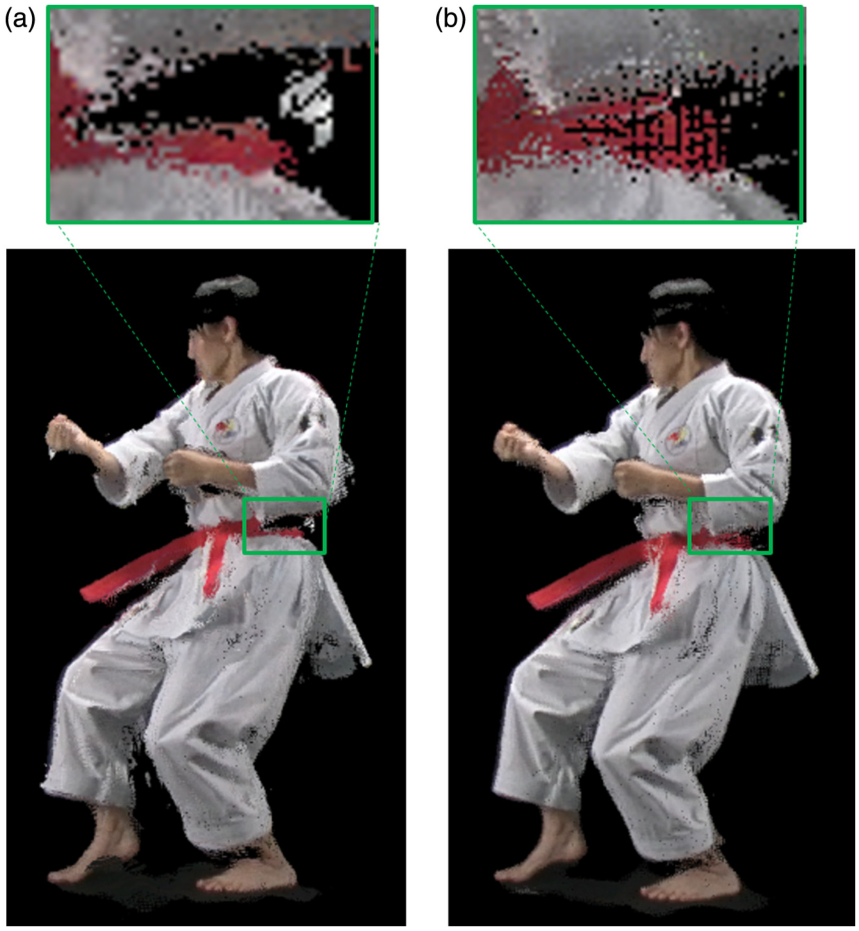

Fig. 21 3-D model generation (a) previous method and (b) proposed method.

In contrast, the proposed method inhibits occurrence of such a hole because the vertical and horizontal rays can be obtained with an array of cameras in both directions.

5 Conclusion

We proposed a capturing method for integral 3-D imaging using multiviewpoint robotic cameras. In this method, a cooperative control function enables the cameras to capture images at the narrowest view angle that does not lose any part of the object region to be reproduced by the integral 3-D display, thereby maximizing the resolution of the image. The multiviewpoint cameras are arranged in both the horizontal and vertical directions, and 3-D models are generated using the captured multiviewpoint images. Elemental images and integral 3-D images featuring pan, tilt, and zoom camerawork are generated. In simulations and actual capture experiments, we verified the reproducibility and practicality of the cooperative control. An examination of integral 3-D images of a karate scene confirmed the effectiveness of the method.

In future work, we will improve the method of generating the 3-D models. Inhibiting depth estimation errors is especially needed in texture-less regions and for the boundaries of the subject. In addition, interpolation[38–43] is required in occlusion regions that cannot be captured by all seven cameras. We will also try to reconstruct high-quality 3-D models by incorporating high-resolution cameras.

References

1. F. Okano et al., "Three-dimensional video system based on integral photography," *Opt. Eng.* **38**, 1072–1077 (1999).
2. X. Xiao et al., "Advances in three-dimensional integral imaging: sensing display and applications," *Appl. Opt.* **52**(4), 546–560 (2013).
3. B. Javidi, J. Sola-Pikabea, and M. Martinez-Corral, "Breakthroughs in photonics 2014: recent advances in 3-D integral imaging sensing and display," *IEEE Photonics J.* **7**(3), 1–7 (2015).
4. M. Cho and B. Javidi, "Three-dimensional tracking of occluded objects using integral imaging," *Opt. Lett.* **33**(23), 2737–2739 (2008).
5. Y. Zhao et al., "Tracking of multiple objects in unknown background using Bayesian estimation in 3D space," *J. Opt. Soc. Am. A* **28**(9), 1935–1940 (2011).
6. R. V. Pole, "3-D Imagery and holograms of objects illuminated in white light," *Appl. Phys. Lett.* **10**(1), 20–22 (1967).
7. T. Mishina, M. Okui, and F. Okano, "Calculation of holograms from elemental images captured by integral photography," *Appl. Opt.* **45**(17), 4026–4036 (2006).
8. K. Wakunami, M. Yamaguchi, and B. Javidi, "High-resolution three-dimensional holographic display using dense ray sampling from integral imaging," *Opt. Lett.* **37**(24), 5103–5105 (2012).
9. R. Ng et al., "Light field photography with a hand-held plenoptic camera," Technical Report CSTR, Vol. **2**, Stanford University, Stanford, California (2005).
10. Lytro, https://www.lytro.com.
11. M. Levoy et al., "Light field microscopy," in *Proc. ACM SIGGRAPH*, pp. 924–934 (2006).
12. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE Int. Conf. on Computational Photography (ICCP '09)*, pp. 1–8 (2009).
13. C. Hahne et al., "Refocusing distance of a standard plenoptic camera," *Opt. Express* **24**(9), 21521–21540 (2016).
14. V. J. Traver et al., "Human gesture recognition using three-dimensional integral imaging," *J. Opt. Soc. Am. A* **31**(10), 2312–2320 (2014).
15. J. Arai et al., "Integral imaging system with 33 mega-pixel imaging devices using the pixel-offset method," *Proc. SPIE* **8167**, 81670X (2011).

16. N. Okaichi et al., "Integral 3D display using multiple LCD panels and multi-image combining optical system," *Opt. Express* **25**(3), 2805–2817 (2017).
17. H. Watanabe et al., "Wide viewing angle projection-type integral 3D display system with multiple UHD projectors," in *IS&T Int. Symp. on Electronic Imaging 2017*, pp. 67–73 (2017).
18. F. Okano et al., "Real-time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.* **36**(7), 1598–1603 (1997).
19. J. Arai et al., "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Appl. Opt.* **37**(11), 2034–2045 (1998).
20. J.-S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," *Opt. Lett.* **27**(13), 1144–1146 (2002).
21. J.-S. Jang and B. Javidi, "Formation of orthoscopic three-dimensional real images in direct pickup one-step integral imaging," *Opt. Eng.* **42**(7), 1869–1870 (2003).
22. J.-S. Jang and B. Javidi, "Three-dimensional projection integral imaging using micro-convex-mirror arrays," *Opt. Express* **12**(6), 1077–1083 (2004).
23. M. Martínez-Corral et al., "Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays," *Appl. Opt.* **43**(31), 5806–5813 (2004).
24. J. Arai et al., "Compact integral three-dimensional imaging device," *Proc. SPIE* **9495**, 94950I (2015).
25. M. Miura et al., "Integral three-dimensional capture system with enhanced viewing angle by using camera array," *Proc. SPIE* **9391**, 939106 (2015).
26. J. Arai et al., "Progress overview of capturing method for integral 3-D imaging displays," *Proc. IEEE* **105**(5), 837–849 (2017).
27. Y. Taguchi et al., "TransCAIP: a live 3D TV system using a camera array and an integral photography display with interactive control of viewing parameters," in *Proc. ACM SIGGRAPH ASIA Emerging Technologies*, pp. 47–50 (2008).
28. M. Katayama and Y. Iwadate, "A method for converting three-dimensional models into auto-stereoscopic images based on integral photography," *Proc. SPIE* **6805**, 68050Z (2008).
29. Y. Iwadate and M. Katayama, "Generating integral image from 3D object by using oblique projection," in *Proc. IDW '11*, pp. 269–272, ITE, Nagoya (2011).
30. K. Hisatomi et al., "3D archive system for traditional performing arts," *Int. J. Comput. Vision* **94**(1), 78–88 (2011).
31. K. Ikeya et al., "Depth estimation from three cameras using belief propagation: 3D modelling of sumo wrestling," in *Conf. for Visual Media Production (CVMP '11)*, pp. 118–125 (2011).
32. K. Ikeya and Y. Iwadate, "Bullet time using multi-viewpoint robotic camera system," in *11th European Conf. on Visual Media Production* (2014).
33. K. Ikeya and Y. Iwadate, "Multi-viewpoint robotic cameras and their applications," *ITE Trans. Media Technol. Appl.* **4**(4), 349–362 (2016).
34. N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring image collections in 3D," in *ACM Transactions on Graphics (Proc. of SIGGRAPH 2006)*, Vol. 25, p. 3 (2006).
35. P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," in *Proc. of IEEE CVPR*, Vol. 1, pp. 261–268 (2004).
36. A. Hosni et al., "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(2), 504–511 (2013).
37. K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc of the 11th European Conf. on Computer Vision*, Vol. 1, pp. 1–14 (2010).
38. A. Sharf, M. Alexa, and D. Cohen-Or, "Context-based surface completion," in *Proc. ACM SIGGRAPH*, pp. 878–887 (2004).
39. R. Schnabel, P. Degener, and R. Klein, "Completion and re-construction with primitive shapes," *Comput. Graphics Forum* **28**(2), 503–512 (2009).
40. S. Salamanca et al., "A robust method for filling holes in 3D meshes based on image restoration," in *Int. Conf. on Advanced Concepts for Intelligent Vision Systems*, pp. 742–751 (2008).
41. J. Becker, C. Stewart, and R. J. Radke, "LiDAR inpainting from a single image," in *IEEE 12th Int. Conf. on Computer Vision Workshops (ICCV Workshops)*, pp. 1441–1448 (2009).
42. S. Park et al., "Shape and appearance repair for incomplete point surfaces," in *Tenth IEEE Int. Conf. on Computer Vision (ICCV '05)*, Vol. 2, pp. 1260–1267 (2005).
43. N. Kawai et al., "Surface completion of shape and texture based on energy minimization," in *18th IEEE Int. Conf. on Image Processing (ICIP '11)*, pp. 913–916 (2011).

Kensuke Ikeya received his BS degree in engineering from Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2004 and his MS degree in engineering from the University of Electro-Communications, Tokyo, Japan, in 2006. He joined Japan Broadcasting Corporation (NHK), Tokyo, Japan, in 2006. Since 2009, he has been engaged in research on multi-viewpoint robotic cameras and their applications at NHK Science and Technology Research Laboratories. His current research interests include computer vision and 3-D image processing.

Jun Arai received his BS, MS, and PhD degrees in applied physics from Waseda University, Tokyo, Japan, in 1993, 1995, and 2005, respectively. In 1995, he joined the Science and Technology Research Laboratories, Japan Broadcasting Corporation (NHK), Tokyo, Japan. Since then, he has been working on 3-D imaging systems. He is a senior manager at the Planning and Coordination Division.

Tomoyuki Mishina received his BS and MS degrees from Tokyo University of Science, Japan, in 1987 and 1989, respectively, and his PhD from Tokyo Institute of Technology, Japan, in 2007. In 1989, he joined Japan Broadcasting Corporation (NHK), Tokyo, Japan. Since 1992, he has been working on three-dimensional imaging systems in the Science and Technology Research Laboratories, NHK.

Masahiro Yamaguchi is a professor in the School of Engineering, Tokyo Institute of Technology. He received his DEng from Tokyo Institute of Technology in 1994. Since 1989, he has been a faculty member of the same institute, from 2011 to 2015 as a full-professor at Global Scientific Information and Computing Center. His research includes color and multispectral imaging, holography, and pathology image analysis. From 2011 to 2017, he was the chair of CIE TC8-07 "Multispectral Imaging."