3D Interaction with Virtual Objects in a Precisely-aligned View using a See-through Mobile AR System

Yuko Unuma[†], Takashi Komuro[†] (member)

Abstract In this paper, we propose a system that enables users to interact with virtual objects that are displayed on a mobile display in a precisely-aligned view using their hands. By projecting a 3D scene obtained by a depth camera according to the user's viewpoint position, the scene including the user's hand displayed on the screen appears seamlessly connected to the actual scene outside the screen, which enables natural interaction with virtual objects through the screen. We conducted an experiment to evaluate the positional accuracy in the presented images. The maximum mean error was 8.60 mm, and the maximum standard deviation was 1.69 mm, which could be improved by further refinement. We also conducted an experiment to evaluate the usability of the system. We asked the participants to perform tasks using the proposed system in the aligned and non-aligned see-through modes. Despite some restrictions in our prototype system, 9 out of 14 participants completed the task faster in the aligned see-through mode. This result shows the future potential of the proposed system in interaction with virtual objects.

Keywords: augmented reality, mobile device, direct manipulation.

1. Introduction

There have been studies on interaction with virtual objects that are overlaid on real scene images using augmented reality (AR) technology. AR systems using a head-mounted display (HMD) ¹⁾⁻⁴⁾ offer immersive user experiences, but users have to wear the device on their head.

On the other hand, there have been studies on interaction using a mobile device ⁵⁾⁻⁷⁾ which does not require users to wear a special device. These studies demonstrated interaction with virtual objects overlaid on a real scene displayed on the screen of the mobile device using the user's hand. However, the images of the scene and the user's hand that are displayed on the screen have different positions and sizes from the real ones, and the appearances of the scene inside and outside the screen are inconsistent. This makes users experience a sense of incongruity since the hand displayed on the screen is inconsistent with their somatic sensation. There have been studies that make images displayed on the screen of a mobile device consistent with the

appearance of a real scene ⁸⁾⁻¹¹⁾. In these studies, the scene on the screen seems connected to that outside the screen, realizing a so-called see-through display. However, only image transformation has been performed, and interaction with virtual objects has not yet been implemented.

In this study, we propose a system that enables users to interact with virtual objects that are overlaid on see-through images displayed on a mobile display in a precisely-aligned view. In the proposed system, a depth camera is attached to the back of the mobile display, and a camera for face tracking is attached to the front of the display. The depth camera is used for obtaining 3D information of a real scene and the user's hand. Face tracking is performed to generate images that change depending on the user's viewpoint position. The 3D scene obtained by the depth camera is projected so as to be centered at the user's viewpoint position, so that an image whose appearance is consistent with that outside the screen is obtained. Thereby, the scene including the user's hand displayed on the screen appears seamlessly connected to the actual scene outside the screen. This enables intuitive interaction in which the user's hand displayed on the screen is consistent with the user's somatic sensation.

The purpose of our study is to investigate how such an

Received August 19, 2016; Revised November 28, 2016; Accepted December 30, 2016

[†]Saitama University
(Saitama, Japan)

aligned view affects users' manipulation performance. Especially, this paper focuses on the issue of the user's recognition of his or her own hand on an aligned-view see-through mobile display. We had a hypothesis that our system allows a user recognize his or her own hand on the screen more easily. We therefore conducted an experiment to verify the hypothesis and asked the participants questions such as how much they felt that the hand that was displayed on the screen was your own hand, how easy it was to recognize the positional relationship between your hand and objects, and how much they had mental fatigue. This is the first work on performance evaluation of manipulating a virtual object by one's own hand in a precisely aligned view on a mobile display.

2. Related Work

There have been studies on mobile AR systems that enable users to interact with virtual objects that are displayed on the screen of a mobile device using their hand. Caballero et al.'s Behand⁵⁾ allows users to perform direct manipulation of virtual objects, such as grasping, by moving their own hand behind a smartphone. Baldauf et al.'s system⁶⁾ and Hürst et al.'s system⁷⁾ also have enabled interaction with virtual objects using the user's hand behind a smartphone based on the fingertip positions. These systems have realized interaction with virtual objects that are overlaid on real scene images displayed on a smartphone; however, only 2D image processing is performed, and 3D positions of the hands and fingers are not obtained. Therefore, accurate 3D interaction has not been realized. There have been studies that realize 3D interaction with virtual objects using a smartphone. Bai et al. realized 3D gesture interaction using the positions and motions of users' fingers obtained from a depth camera¹²⁾¹³⁾. Their experiments showed that 3D gesture input is more natural and intuitive than 2D touch input.

The above studies have enabled interaction with virtual objects using mobile devices. However, the image of a user's hand displayed on the screen is different from the real one, in position and size. Therefore, the user's field of view becomes narrow since the user tends to gaze only at the screen, and also the user experiences a sense of incongruity since the hand displayed on the screen is inconsistent with the user's somatic sensation.

To overcome this problem, there have been studies on mobile AR systems that make images on the screen consistent with the appearance of a real scene. For

instance, Yoshida et al. have developed ARScope⁸⁾. When a user looks at a handheld device that has the shape of a magnifying glass, the user can see the scene behind the device as if the device were transparent; one drawback, however, is that the user has to wear a projector and a camera on his or her head. Uchida et al. have developed a system that generates images that are consistent with a real space by applying a perspective transformation to the images⁹⁾. However, only a part of the scene that is on a plane with a marker is made consistent with the real space, whereas the rest of the scene is not made consistent. Tomioka et al. have also developed a system that generates images that are consistent with a real space by estimating the 3D positions of feature points obtained by a camera and the poses of the camera¹⁰⁾. This system does not require markers, but the whole scene is approximated by a plane and it is difficult to apply the method to non-planar scenes. Moreover, in these systems, only image transformation has been performed, and overlaying virtual objects has not been realized.

Baričević et al. have measured the time from when a virtual object appeared on the screen to when a user selected it, and compared these times for cases where the images displayed on a virtual tablet were consistent or inconsistent with the appearance of a virtual space, in a VR environment presented by an HMD¹¹⁾. On a tablet-size screen, the case where the images were consistent with the virtual space took a shorter time than the other case. This result showed that making the images on a display consistent with the appearance of a real space is effective for achieving natural interaction. They have also developed a system that presents 3D scenes on a tablet display so that the appearances inside and outside the screen become consistent by tracking infrared LEDs attached to the user's head. Since the 3D scenes are constructed using the KinectFusion algorithm¹⁴⁾ in advance, the system can only present static 3D scenes. Therefore, it is impossible to use the system for interaction with virtual objects using the users' hand.

Pucihar et al. have conducted an experiment in which the participants touch targets that appear on images captured by the camera of a smartphone in two rendering methods: device-perspective rendering and user-perspective rendering¹⁵⁾. They confirmed that user perspective rendering allowed much more effective spatial recognition and was preferred by users. It is expected that user-perspective rendering would also be effective in 3D interaction with virtual objects.

3. See-through Mobile AR System

The configuration of the system is shown in Fig. 1. A camera for face tracking is attached to the front of a mobile display, and a depth camera for capturing images of a hand and the background is attached to the back of the display. The mobile display has a screen area of $151.44 \text{ mm} \times 90.576 \text{ mm}$. The camera for face tracking has a resolution of 640×480 and a frame rate of 60 fps. The depth camera has a resolution of 1280×720 , a frame rate of 30 fps, and an angle of view of $63.2^\circ \times 49.3^\circ$ for color images, and a resolution of 320×240 , a frame rate of 60 fps, an angle of view of $74^\circ \times 58^\circ$, and a detectable depth range of 15 cm to 1m for depth images. The asmlibrary¹⁶⁾ is used for face tracking to detect the viewpoint position of a user. The mobile display is used only as a screen, and processing is performed on a PC.

The depth camera obtains color and depth images in every frame. The depth image is associated with the color image, and 3D point cloud data is generated. If the points in the 3D data are projected as they are, the projection image would become grainy. Therefore, surfaces are generated by connecting the points in adjacent pixels in the depth image, and these are projected.

To present an image of the scene as viewed from the user's viewpoint position on the mobile display, face tracking is performed to obtain the viewpoint position. The position of the user's dominant eye is used as the viewpoint position. The user's face is detected from an image obtained by a camera that is attached to the front of the mobile display, and the 2D position of the user's dominant eye and the distance between both eyes in the image are obtained. Then, the 3D position of the user's dominant eye is calculated using the real distance between both eyes and the camera's focal length.

Even when the user is not moving, the detected viewpoint position moves slightly in frames due to noise in images. To reduce this fluctuation, the viewpoint position is smoothed using exponential smoothing.

The 3D scene obtained by the depth camera is displayed after changing the center of projection to the user's viewpoint position, as shown in Fig. 2.

Figure 3 shows the see-through images of a real scene using the proposed system. They were captured by fixing the viewpoint position and moving or tilting the display. The black regions in the images are the regions which are not in the field of view of the depth camera.

The projection images in which virtual objects are

Fig. 1 System configuration.

Fig. 2 Projection depending on the viewpoint position.

Fig. 3 See-through images of a real scene.

overlaid on the scene are generated by placing the virtual objects in the same space as that of the 3D scene. Interaction between virtual objects and the user's hand is realized using the depth information obtained from the depth camera. To simplify the calculation, the shapes of virtual objects are approximated by spheres. Grasping an object as shown in Fig. 4 is realized by the following algorithm. First, the upper half of the sphere that approximates the object is divided into front and back regions. Here, up means the y-axis positive direction in the camera coordinate system and

Fig. 4 Grasping a virtual object.

front/back mean the z-axis negative/positive directions. The centroids of all points in the 3D scene data in the front and back regions are calculated, and the object is grasped when there are points inside the both regions. While grasped, the object is moved to the midpoint of the centroids. The object is released when there are no points inside either of the regions. The algorithm does not recognize a hand, but simply use all points in the scene for realizing interaction.

4. Evaluation of Positional Accuracy

We conducted an experiment to evaluate the errors in the observed 2D positions of objects between two cases: one where the objects are viewed through the screen using the proposed system and one where they are viewed directly without using the proposed system.

4.1 Methodology

Using the centroid position of a marker that was directly observed by a camera as a base position, the centroid positions of the marker were observed ten times from the viewpoint position, and the mean and standard deviation of the errors (in distances) were calculated. Figure 5 (a) shows the setup of equipment. A sheet of paper on which a marker with a diameter of 4 cm was printed was attached to a wall. We used a camera that was mounted on a tablet device (resolution: 1920×1080) as a viewpoint camera, which captures the images of the scene that was viewed from the user's viewpoint. A printed image of a human face with a hole in the left eye position was attached to the tablet device so that the hole was overlapped with the camera. The viewpoint was obtained by face tracking. Figure 5(b) shows a snapshot of the mobile display screen, where both the real and estimated marker positions are drawn in the screen.

The distances between the marker and the mobile display were set to four different combinations, as shown in Fig. 6. With the positions of the marker and the camera fixed, the mobile display was moved so that the

Fig. 5 Experimental environment:.

Fig. 6 Experimental conditions.

marker was displayed at the center, right, left, top, and bottom of the screen of the mobile device, and the images of the marker were captured 10 times for each position. When the displayed marker position was at the center of the screen, the camera, the marker, and the mobile display were placed so that their centers were at the same height. When the position was at the left or right of the screen, the mobile device was moved by 6 cm horizontally. When the position was at the top or bottom of the display, the mobile display was moved by 2.5 cm vertically.

4.2 Results

Figure 7 shows a scatter plot of the centroid positions of the marker that was displayed on the mobile display. The lengths in the table are those on the mobile display. The base position of the marker that was directly measured without the mobile display is indicated with a

Fig. 7 Scatter plot of the centroid positions.

Table 1 Errors of face tracking and viewpoint estimation.

distance	Δx [pixel]	Δy [pixel]	ΔZ [cm]
30 cm	-1.13 (0.93)	0.39 (0.65)	1.93 (0.52)
50 cm	-0.32 (0.32)	-0.72 (0.38)	2.16 (0.51)

circle. The minimum mean error was 1.71 mm, and the maximum was 8.60 mm.

Since the errors in (b) were smaller than those in (a), and those in (d) were smaller than those in (c), the accuracy was better when the distance from the mobile display to the viewpoint camera was 50 cm. Especially in (a) and (c), the centroids were generally shifted to the right. This was caused by the differences between the internal and external camera parameters used in the calculation and the actual ones. This could be improved by performing camera calibration more carefully.

The standard deviations were 0.14 mm at minimum and 1.69 mm at maximum. There were some points whose positions were largely deviated. These deviations were caused by the failure to perform face tracking at the time of image capturing. If these two points are removed, the standard deviation was 0.81 mm at most, which would not affect interaction.

To investigate how the accuracy of 3D viewpoint estimation affects the positional consistency, we measured the accuracy of face tracking and 3D viewpoint estimation. Since the errors of viewpoint estimation vary depending on the real viewpoint positions, we fixed the position so that the distance from the camera to the printed image of the face becomes 30cm and 50 cm, which are the same condition as in Fig. 6, and that the left eye in the camera image is at the center of the image. Table 1 shows the means and standard deviations (in parentheses) of the errors in the left eye positions (x , y) in 2D camera images and the estimated distance Z from the camera to the eye.

Since the 3D positions (X , Y , Z) are calculated using the relation $x = fX/Z$ and $y = fY/Z$, where $f = 308$ [pixel] is the focal length of the camera for face tracking, errors in 3D positions for any viewpoint positions can be estimated as $\Delta X = (Z/f) \Delta x + (X/Z) \Delta Z$, $\Delta Y = (Z/f) \Delta y + (Y/Z) \Delta Z$.

In the condition of Fig. 6(a), for example, the standard deviation of the marker's movement due to the errors in 3D viewpoint estimation becomes 0.55 [mm], which is similar to the measured standard deviations in the experiment and that shows the main cause of the positional fluctuation of the marker was the errors in 3D viewpoint estimation.

5. Evaluation of Usability

We conducted an experiment to evaluate the errors in the observed 2D positions of objects between two cases: one where the objects are viewed through the screen using the proposed system and one where they are viewed directly without using the proposed system.

We conducted an experiment to verify if the proposed system is effective for interaction with virtual objects overlaid on a real scene displayed on a mobile display.

5.1 Methodology

We recruited 14 participants (male 11, female 3) from outside our laboratory. Their ages ranged from 20 to 24 with a mean of 21.86. Eleven of them were right-handed, and three were left-handed. Eleven of them were right-eye dominant, and three were left-eye dominant. Eight of them had some knowledge of AR, three only knew the term AR, and three had no knowledge of AR. Five of them had experience of using applications or games using AR.

Figure 8 shows the task used in the experiment. The task was to carry virtual objects (cylinders) that were stacked in three layers on a red marker, which can be seen on the screen of the mobile display, to the position of a green marker outside the screen, one by one, and to stack them in three layers again.

Red and green circular markers with diameters of 4 cm were used to determine the initial and goal positions of the virtual objects, respectively. The markers were put on a desk with a separation of 45 cm. We asked the participants to be seated on a chair in front of the desk, to hold the mobile display with one hand, and to grasp a virtual object with the other hand. We gave them detailed instructions for grasping.

When a participant held a mobile device so that the red marker was displayed on the screen, three stacked cylinders appeared on the marker. The participant grasped one of the cylinders from the top, carried it to the position of the green marker, and put it on the marker. One task was finished when all the three

Fig. 8 Task used in the experiment.

Fig. 9 A user performing the task.

cylinders were stacked on the green marker. While moving a virtual object, the participant had to also move the mobile display so that the object was kept displayed on the screen.

The same algorithm that was described in Section 3 was used for grasping recognition. The color of a cylinder was yellow while it was not grasped. When it was grasped, the outline was drawn in orange lines. When it was put on the green marker, it turned white. Only the object that had to be carried, and an object that was already put on the green marker did not move. The participants were told in advance that it is no problem if their hand touched the other objects while carrying an object. Figure 9 shows an example in which a user performs the task.

The participants performed the task under two conditions. In one condition, the appearances of the scene inside and outside the screen were made consistent. We call this condition the "aligned see-through mode". In the other condition, the center of projection for projecting the 3D scene was fixed at a position 60 cm in front of the mobile display screen, towards the user. This is the position at which the entire image obtained by the depth camera can be displayed on the screen. We call this condition the "non-aligned see-through mode". Figure 10 shows the presented images in both modes.

We asked the participants to perform the task three times in each of the aligned and non-aligned see-through modes, for a total of six tasks. We gave a three-minute training session before the first time they performed the task in both the aligned and non-aligned see-through modes, and gave a one-minute training session before the second and third times. Half of the participants performed the task in the aligned see-through mode first, and the other participants performed the task in the non-aligned see-through mode first. Participants rested for one minute after each task.

We recorded the time for completing each task (from

Fig. 10 Aligned (left) and non-aligned (right) see-through modes.

Fig. 11 Task completion time of each participant.

when a participant grasped a virtual object to when he or she finished moving all three objects. We did not tell the participants that we measured the time, and asked them to perform the task as carefully as possible. After a participant finished all the tasks, we asked him or her to answer the following questions on a 5-point Likert scale.

- Q1. Were you able to manipulate objects comfortably? (comfort)
- Q2. Did you feel that the hand that was displayed on the screen was your own hand? (reality of appearance)
- Q3. Was it easy to recognize the positional relationship between your hand and objects? (spatial recognition)
- Q4. Did you feel like grasping real objects? (reality of manipulation)
- Q5. Did you have mental fatigue? (fatigue)
- Q6. Was it fun to manipulation the objects? (fun)
- Q7. Do you want to use this system in the future? (preference)

In addition, we asked the participants to write freely what they felt in the experiment, their opinions about the system, and their impressions.

5.2 Results

Figure 11 shows the mean task completion time of each participant. Nine participants completed the task faster in the aligned see-through mode, and five participants completed the task faster in the non-aligned see-through mode.

In the aligned see-through mode, the mean task completion time of all the participants was 127.5 s, and

Fig. 12 Results of questionnaire.

the standard deviation was 82.3 s, and in the non-aligned see-through mode, the mean task completion time of all the participants was 141.7 s, and the standard deviation was 123.89 s. The result of the Shapiro-Wilk test showed that the task completion times in both modes were not normally distributed. We therefore conducted the Wilcoxon signed-rank test, and the result showed that there was no significant difference between the two modes ($p = 0.28$).

Figure 12 shows the results of the questionnaire. We allocated scores from 5 to 1 to the answers, where 5 is the best and 1 is the worst, and the scores of 14 participants were averaged for each question. The result of the Wilcoxon signed-rank test showed that there were no significant differences for all questions ($p > 0.05$).

Some of the comments we received from the participants are: "It was difficult to recognize the object's positions", "It was fun to grasp the object", "I felt as if I was grasping a soft object like a marshmallow", "I got tired holding the device", "I was excited that I can move a virtual object with my hand", "There was a delay in the motions on the screen when I moved the display".

5.3 Discussion

In the results of the questionnaire, the mean scores in the aligned see-through mode were worse than those in the non-aligned see-through mode for all questions except Q4, though there were no significant differences. Possible causes include the following problems that existed in our prototype system.

The first problem is that the user's whole face has to be in the field of view of the front camera, which restricts the way to hold the mobile display. This restriction can be reduced by using a lens having a wider field of view to make the front camera always view the whole face.

The second problem is that there can be regions that are outside the depth camera's field of view and that cannot be drawn on the display. When a user looks at a

Fig. 13 Drawable ranges of the system.

vertical surface, the scene is displayed on the entire screen as shown in Fig. 13(a). However, when the user looks at a horizontal surface, such as a desk, the scene is not displayed in the bottom part of the screen, as shown in Fig. 13(b), which made it hard to see the scene when the participants grasped a virtual object. This problem can be solved by using a depth camera having a wider field of view, or by attaching the depth camera at a lower position.

The third problem is that the scene that is displayed on the screen sometimes moves by a large amount. As shown in Section 4, the standard deviation of position errors was less than 1 mm, which was mainly caused by errors in 3D viewpoint estimation as discussed in Section 4.2. This positional fluctuation is small enough for stable interaction and may not have affected the results of the usability evaluation. However, the viewpoint moves largely in a frame when the system fails to track a face and then detects the face again. This problem can be solved by improving the face tracking algorithm, or by predicting the viewpoint position using the positions in the previous frames even if face tracking fails.

Despite these problems, there was no significant difference in the mean task completion time. If we can solve the problems with the current prototype system described above, the completion time would perhaps become significantly faster in the aligned see-through mode.

6. Conclusion

In this paper, we proposed a system that enables a user to interact with virtual objects displayed on a mobile display using his or her hand. By projecting a 3D scene that is obtained by a depth camera according to the user's viewpoint position, the scene inside the screen looks seamlessly connected to that outside the screen, which enables natural interaction with virtual objects through the screen.

In usability evaluation, we confirmed that, despite some restrictions in our prototype system, more than half of the participants completed the task faster in the aligned see-through mode, where the appearances of the scene inside and outside the screen were made consistent. This result showed the future potential of the proposed system.

Future work includes overcoming the problems that became apparent in the experiment, and performing the experiment again to show the effectiveness of a precisely-aligned view for a variety of manipulations and various shapes of virtual objects. In addition, it will be necessary to show the possibility of practical use by creating applications using the proposed system.

References

- 1) Minkyung Lee, Richard Green and Mark Billinghurst: "3D Natural Hand Interaction for AR Applications", Proceedings of the 23rd International Conference Image and Vision Computing New Zealand (IVCNZ 2008), pp.1-6 (2008)
- 2) Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt and Vivek K Goyal: "Mime: Compact, Low-Power 3D Gesture Sensing for Interaction with Head-Mounted Displays", Proceedings of the 26th annual ACM Symposium on User Interface Software and Technology (UIST 2013), pp.227-236 (2013)
- 3) Usman Sargaana, Hossein S. Farahani, Jong Weon Lee, Jeha Ryu and Woontack Woo: "Collaborative billiARds: Towards the Ultimate Gaming Experience", Proceedings of the 4th International Conference on Entertainment Computing (ICEC 2005), pp.357-367 (2005)
- 4) Benjamin Knoerlein, Gábor Székely and Matthias Harders: "Visuo-Haptic Collaborative Augmented Reality Ping-Pong", Proceedings of the International Conference on Advances in Computer Entertainment Technology (ACE 2007), pp.91-94 (2007)
- 5) María Luz Caballero, Ting-Ray Chang, María Menéndez and Valentina Occhialini: "Behand: Augmented Virtuality Gestural Interaction for Mobile Phones", Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2010), pp.451-454 (2010)
- 6) Matthias Baldauf, Sebastian Zambanini, Peter Fröhlich and Peter Reichl: "Markerless Visual Fingertip Detection for Natural Mobile Device Interaction", Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2011), pp.539-544 (2011)
- 7) Wolfgang Hürst and Casper van Wezel: "Gesture-based Interaction via Finger Tracking for Mobile Augmented Reality", Multimedia Tools and Applications, Volume 62, Issue 1, pp.233-258 (2013)
- 8) Takumi Yoshida, Shinobu Kuroki, Hideaki Nii, Naoki Kawakami and Susumu Tachi: "ARScope", Proceedings of the 35th International Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH 2008), Article No. 4 (2008)
- 9) Hikari Uchida and Takashi Komuro: "Geometrically Consistent Mobile AR for 3D Interaction", Proceedings of the 4th Augmented Human Interaction Conference (AH 2013), pp.229-230 (2013)
- 10) Makoto Tomioka, Sei Ikeda and Kosuke Sato: "Approximated User-Perspective Rendering in Tablet-Based Augmented Reality", Proceedings of the 12th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2013), pp.21-28 (2013)
- 11) Domagoj Baričević, Cha Lee, Matthew Turk, Tobias Höllerer and Doug A. Bowman: "A Hand-Held AR Magic Lens with User-Perspective Rendering", Proceedings of the 11th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2012), pp.197-206 (2012)
- 12) Huidong Bai, Lei Gao, Jihad El-Sana and Mark Billinghurst: "Markerless 3D Gesture-based Interaction for Handheld Augmented Reality Interfaces", Proceedings of the 12th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2013), pp.1-6 (2013)
- 13) Huidong Bai, Gun A. Lee, Mukundan Ramakrishnan and Mark Billinghurst: "3D Gesture Interaction for Handheld Augmented Reality", Proceedings of the 7th SIGGRAPH Asia 2014 Symposium on Mobile Graphics and Interactive Applications (SA 2014), Article No. 7 (2014)
- 14) Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison and Andrew Fitzgibbon: "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera", Proceedings of the 24th annual ACM Symposium on User Interface Software and Technology (UIST 2011), pp.559-568 (2011)
- 15) Klen Čopič Pucihar, Paul Coulton and Jason Alexander: "Evaluating Dual-view Perceptual Issues in Handheld Augmented Reality: Device vs. User Perspective Rendering", Proceedings of the 15th ACM on International Conference on Multimodal Interaction (ICMI 2013), pp.381-388 (2013)
- 16) asmlibrary - Library of Active Shape Model, <https://code.google.com/p/asmlibrary/>

Yuko Unuma received the B.E. and M.E. degrees in information and computer sciences from Saitama University in 2014 and 2016, respectively. At present she works for Marvelous Inc.

Takashi Komuro received the B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo in 1996, 1998, and 2001, respectively. At present he is an associate professor of mathematics, electronics and informatics at Saitama University. His current research interests include image sensing, computer vision, user interfaces, and augmented reality.