

Direct Finger Manipulation of 3D Object Image with Ultrasound Haptic Feedback

Atsushi Matsubayashi

The University of Tokyo
Bunkyo-ku, Tokyo

matsubayashi@hapis.k.u-tokyo.ac.jp

Yasutoshi Makino

The University of Tokyo
Kashiwa-shi, Chiba

yasutoshi_makino@k.u-tokyo.ac.jp

Hiroyuki Shinoda

The University of Tokyo
Kashiwa-shi, Chiba

hiroyuki_shinoda@k.u-tokyo.ac.jp

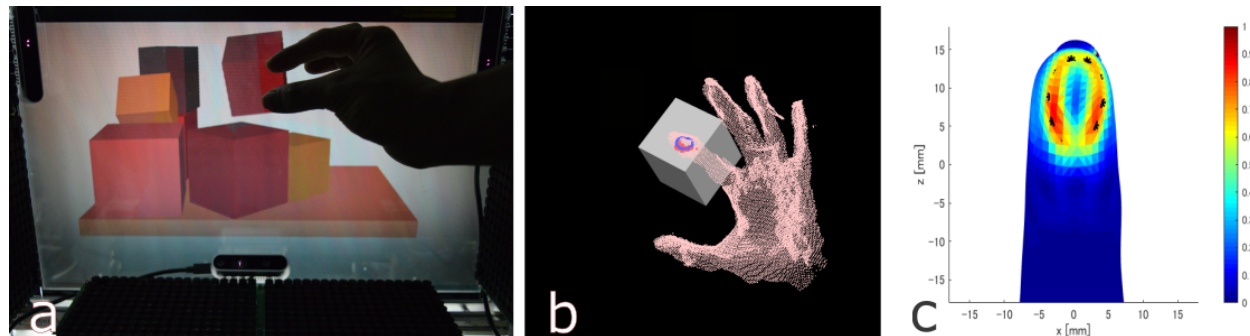


Figure 1: (a) A user can grasp and move a 3D object image. (b) The cross section of the finger is estimated. (c) The ultrasound focus moves around the cross section and generates a pressure distribution on the finger surface.

ABSTRACT

In this study, we prototype and examine a system that allows a user to manipulate a 3D virtual object with multiple fingers without wearing any device. An autostereoscopic display produces a 3D image and a depth sensor measures the movement of the fingers. When a user touches a virtual object, haptic feedback is provided by ultrasound phased arrays. By estimating the cross section of the finger in contact with the virtual object and by creating a force pattern around it, it is possible for the user to recognize the position of the surface relative to the finger. To evaluate our system, we conducted two experiments to show that the proposed feedback method is effective in recognizing the object surface and thereby enables the user to grasp the object quickly without seeing it.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Interactive systems and tools**.

KEYWORDS

Touch/Haptic/Pointing/Gesture; Virtual/Augmented Reality

ACM Reference Format:

Atsushi Matsubayashi, Yasutoshi Makino, and Hiroyuki Shinoda. 2019. Direct Finger Manipulation of 3D Object Image with Ultrasound Haptic Feedback. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3290605.3300317>

1 INTRODUCTION

Recently, with the development of 3D display and sensing technology, interaction with stereoscopic images has attracted much attention. A VR system that enables users to touch and manipulate a 3D object image is not only used for visualizing images in simulation, telexistence, or gaming, but also for supporting a user as an intuitive interface system. An interactive system that a user can operate naturally and intuitively, as if manipulating a real object, can make various operations easy and efficient. In particular, such a system has a great potential to improve the operability of a task in 3D modeling, which is unintuitive with a conventional



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5970-2/19/05.

<https://doi.org/10.1145/3290605.3300317>

2D display and pointing device. Haptic feedback is an indispensable element in the system that fully utilizes the human inherent ability for object handling.

In this study, we prototype and examine a system that allows a user to manipulate a 3D virtual object using multiple fingers without wearing any device. For example, as shown in Figure 1 (a), a user can hold a 3D virtual object with two fingers and move it freely with haptic feedback. The system performs a physical simulation according to the hand posture measured by depth sensors and makes the result reflect on the image in the 3D display, which creates a visual interaction. Simultaneously, haptic feedback is provided by ultrasound phased arrays. By controlling the phase of each transducer, an ultrasound focus can be created at an arbitrary position in the air, so that the user can feel the tactile sensation at any contact point in the workspace without wearing any device. Furthermore, this system can generate a pressure distribution on the finger depending on its contact state with the surface by moving the focal point of the ultrasound at a high speed. In our algorithm, the cross section of the finger penetrating the virtual object (Figure 1 (b)) is calculated and the focus moves along the intersection line so that the user can perceive a time-averaged pressure distribution spreading on the finger surface (Figure 1 (c)). The pressure distribution that spreads as the finger enters the object surface can make the user feel an expansion of the contact area and recognize the penetrating depth of the object surface. Accurate position recognition of the virtual object surface can prevent excessive penetration of the finger that causes unintended object reactions in the physics simulation. As the result, it enables the user to perform certain operations without seeing the object, which is often necessary for an environment where visual occlusion is an issue.

To evaluate the effectiveness of ultrasound haptic feedback quantitatively, we conducted experiments using the prototype system. We compared the variable contact area method, as explained above, with a nonvariable one. The result of the first experiment shows that the former feedback method improves the accuracy of the angle recognition of the object surface. The result of the second experiment indicates that it is possible to grab a virtual object quickly with only tactile feedback, and that the former method enriches the reality of objects.

This is the first system that enables a user to manipulate a 3D object image with haptic feedback without wearing any device. In addition, it is the first trial of a mid-air haptics system that provides the pressure distribution on finger surfaces according to the contact condition.

2 RELATED WORKS

Interaction with 3D virtual objects

Many research efforts have been made to develop a system that enables interaction with 3D images. Kim et al. proposed a system for fingertip interaction [17]. In this system, the fingertip is captured by two cameras and its position is determined by stereo geometry. A user can touch the 3D image, reproduced by an integral 3D display, with fingertips. Butler et al. presented Vermeer, which enables the user to interact with a 360° viewable 3D display [4]. In this system, a 3D image is reproduced by a spinning directional diffuser and a high frame rate projector, and the touchable image is reconstructed by a pair of parabolic mirrors facing each other. Hilliges et al. proposed Holodesk, which is an interactive system combining an optical see through display and a Kinect camera [11]. In this system, a virtual image of a 3D scene is rendered in accordance with the user's head position, so that the user can directly interact with the virtual object floating in midair. Recently, a system in which a sensing device and display are integrated has been proposed by Wang et al. [30] [29]. Because an optical sensor array is embedded in an LCD, this system does not require a registration between them.

These systems provide only visual information for users to recognize the spatial information of a 3D image. However, haptic information is also important for spatial recognition. Yoshida et al. proposed RePro 3D, which is a system that enables haptic interaction with stereoscopic images displayed by multiple projectors and retroreflective materials [31]. In this system, haptic feedback is presented from hand-mounted devices. Han et al. also proposed a system for interacting with 3D images, projected in the same way as HoloDesk, with haptic devices attached to the finger [8]. Generally, the method of presenting haptic feedback by attaching a haptic device to the human body has a problem that the device affects the 3D image and the attachment impairs the convenience of the interface.

Recently, several methods of presenting mid-air haptic feedback have been proposed. Sodhi et al. proposed a tactile display using a ring-shaped air flow called a vortex ring [26]. Although this display has a simple mechanism and is easy to create, the range and resolution of the tactile sensation that can be presented is limited both spatially and temporally. Iwamoto et al. proposed a device with a higher spatial and temporal resolution using acoustic radiation pressure of ultrasound waves called airborne ultrasound tactile display [15]. Carter et al. also presented a method, named Ultrahaptics, using ultrasound transducers [5].

Several previous works enable ultrasound haptic interaction with a floating image. Haptomime [22] enables haptic interaction with a floating 2D image reconstructed in the

air using a special mirror device. Ultrasound haptic feedback is presented at the contact point, but an interaction can be performed with only one finger because the image is two-dimensional. Sand et al. proposed a system combining a head-mounted display and an ultrasound transducer array [25]. Although this system enables 3D tactile interaction, in the experiment, only 2D interaction using a virtual keyboard was performed. HaptoClone [21] realizes real-time interaction with a floating 3D image. However, because the image is an optical copy of a real object, it is impossible to freely create 3D images and interact with them. In these systems, haptic feedback is provided by creating an ultrasound focus at the contact position of the finger, but a pressure distribution on the finger is not produced. Though users can understand that their hand is touching the object, it is difficult to perceive the relative positional relationship between their fingers and the object surface. Therefore, a handling task with multiple fingers has not been performed.

Displaying haptic objects

Haptic feedback is an important research field in interacting with virtual objects, and many methods to display touchable objects using ultrasound waves have been proposed. These methods can be broadly classified into two categories.

The first one is to create a haptic image in midair. Gavrilov proposed a method for generating a two-dimensional haptic image in water by generating multiple focal points [6]. A method of using multiple focal points was also used by Korres et al. to generate three-dimensional haptic objects in midair [18]. In contrast, several methods for generating spatial sound pressure distributions have been proposed. Hasegawa et al. proposed a method to generate a three-dimensional distribution with transducers located in a three-dimensional disposition [10]. Inoue et al. also proposed a method of creating a three-dimensional haptic shape by controlling the ultrasound pressure distribution [12]. In these methods, because a haptic image is created in advance, the influence of the delay of hand sensing is less. However, the resolution of the haptic image is low and it is difficult to present the sharp pressure change the moment the finger touches the object.

The other strategy for displaying haptic objects is to control the pressure of a finger based on prompt feedback from the finger motion sensing. The method to generate a focus at the contact point is categorized under this. Long et al. proposed a method of controlling the sound pressure at the intersection of the object with the hand [20]. This method enables to recognize the three-dimensional shape with the palm, but it is difficult to apply this to a small area such as a fingertip. Inoue et al. proposed a method of dividing the finger surface with a mesh and controlling the sound pressure distribution on it [12]. Although this method can reproduce

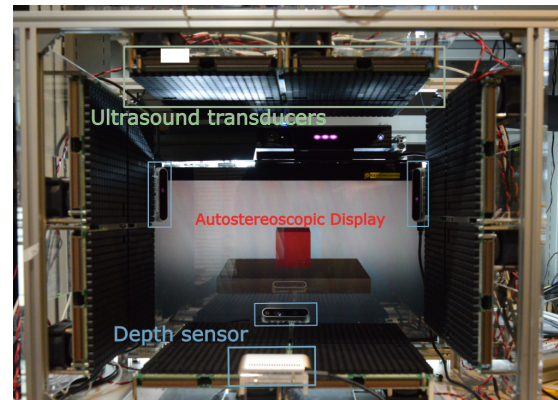


Figure 2: Configuration of proposed system

the pressure distribution of the finger surface that touches the object, it is difficult to generate an accurate mesh model, and the cost of numerical calculation is high. We propose a feedback method with a low computational cost, which can generate the pressure distribution of the fingertip corresponding to the contact surface.

3 PROTOTYPE SYSTEM

Figure 2 shows the configuration of the prototype system. An autostereoscopic display created by Kakeya et al. [16] [32] generates a floating image. Depth sensors capture the hand shape, and when the user touches the 3D object image, ultrasound transducers provide haptic feedback at the contact point.

Hand sensing

To interact with a virtual object, it is required to measure the user's hand shape. The prototype system uses RealSense D415 and D435 (Intel Corporation) [2] to sense the user's hands moving. As shown in Figure 2, we set four RealSense cameras to acquire the depth information of all sides of the hand. These sensors use stereo vision to calculate the depth, so that they do not interfere with each other. As shown in Figure 3, the background is cut out from the raw depth image acquired by RealSense, and subsequently smoothed by a bilateral filter [23]. From the processed depth image, a point cloud representing the shape of the hand is obtained by reprojecting each pixel to the three-dimensional space using the camera parameters. The point cloud is down-sampled by a voxel grid filter to speed up subsequent processing and equalize the point cloud density. We set the interval of the grid to 0.5 mm in the prototype system. From the down-sampled point cloud, the system performs a physics simulation and reflects the result on the 3D image. In the prototype system, the resolution of the depth image captured by each depth sensor is 848×480 , and the refresh rate is 30

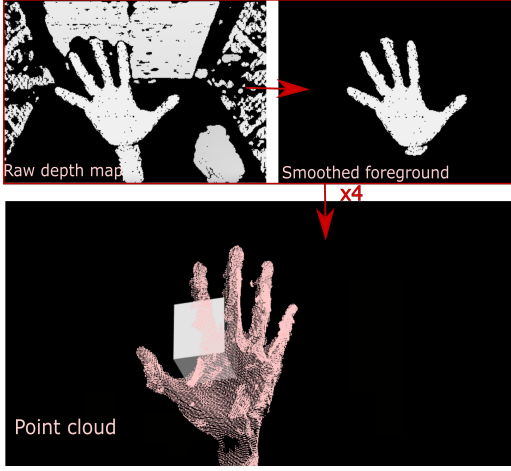


Figure 3: Excluding the background from raw depth images, smoothing them, and reprojecting them to the three-dimensional space

Hz. The voxel grid filter is implemented with point cloud library [24], which is an open source library for point cloud processing.

Physics simulation

When the point cloud exists inside the virtual object, the system regards that the hand is touching the object, and the position and angle of the object are changed accordingly. To enable operation with multiple fingers, the point cloud inside the object is clustered based on the distance between points. The centroid of each cluster is regarded as a fingertip position, and a physical simulation is performed according to its movement. The force applied to the object is determined based on a method that uses a proxy, called a god-object [33], which is mainly used for 3D pointing devices such as Phantom [3]. As shown in Figure 4 (a), the god-object is placed at the fingertip position when the finger penetrates into the object for the first time. The position of the god-object $\mathbf{x}_g \in \mathbb{R}^3$ represents the contact position between the finger and the object. Thereafter, until the finger leaves the object, a force is applied to the object at \mathbf{x}_g as if it is connected with a spring and damper to the fingertip. Let $\mathbf{x} \in \mathbb{R}^3$ be the current fingertip representative position. The applied force $\mathbf{f} \in \mathbb{R}^3$ is calculated as

$$\mathbf{f} = K (\mathbf{x} - \mathbf{x}_g) + D (\dot{\mathbf{x}} - \dot{\mathbf{x}}_g),$$

where K is a proportionality constant to the distance, and D is a proportionality constant to the relative velocity. The component horizontal to the plane of this force represents the frictional force. When it exceeds the maximum static frictional force, the fingertip is considered to be sliding on

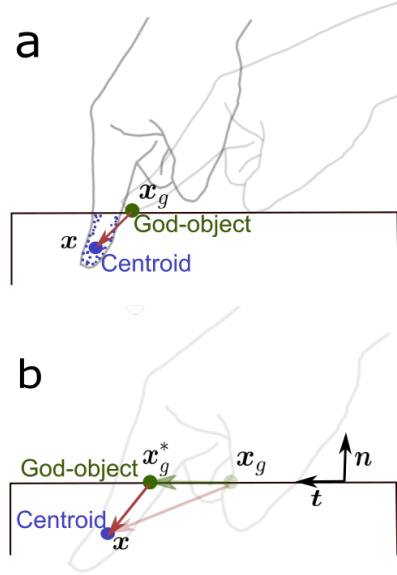


Figure 4: (a) A god-object is generated at the position touched by the finger, and then a force is added to the object according to the fingertip position. (b) When the friction force exceeds the maximum static friction, the god-object moves.

the surface of the object and the god-object moves so that the applied force becomes dynamic friction [9] as shown in Figure 4 (a). Let μ_s and μ_d be the coefficients of static and dynamic friction, respectively, and $\mathbf{n} \in \mathbb{R}^3$ and $\mathbf{t} \in \mathbb{R}^3$ be the normal and tangential unit vectors, respectively. If

$$(\mathbf{x} - \mathbf{x}_g) \cdot \mathbf{t} > \mu_s (\mathbf{x} - \mathbf{x}_g) \cdot \mathbf{n},$$

then the position of the god-object is updated. In the above stationary determination, the force that is proportional to the velocity is ignored for the stability of the simulation. The updated position of the god-object $\mathbf{x}_g^* \in \mathbb{R}^3$ is calculated as

$$\mathbf{x}_g^* = \mathbf{x}_g + ((\mathbf{x} - \mathbf{x}_g) \cdot \mathbf{t} - \mu_d (\mathbf{x} - \mathbf{x}_g) \cdot \mathbf{n}) \mathbf{t}.$$

Similar to the method proposed by Talvas et al. [28], by putting multiple god-objects applying torque to the rotation around the fingertip, the system enables the user to stably grasp an object.

In the prototype system, all physical operations including applying gravity, collision processing with other objects, etc. are implemented with bullet physics [1], which is an open source physics engine. The physics simulation is done at 60 Hz, while the hand tracking refresh rate is 30 Hz. The position of the virtual object is updated twice by a set of point clouds.

Haptic feedback

Haptic feedback is added to the contact points of fingers with virtual objects. As shown in Figure 2, ultrasound transducers are arranged around the operation area, and the phase of each transducer is controlled to create the focal points of ultrasound waves. When a focal point is blocked by the human body surface, vibrotactile sensation is generated based on the acoustic radiation pressure. We use a simple summation method to generate multiple focal points. Let $\mathbf{x}_n \in \mathbb{R}^3, n = \{1, \dots, N\}$ be the position of each focal point and $\mathbf{x}_m \in \mathbb{R}^3, m = \{1, \dots, M\}$ be the position of each transducer. The complex amplitude of each transducer $q_m \in \mathbb{C}$ can be calculated as

$$q_m = \sum_{n=1}^N \|\mathbf{x}_n - \mathbf{x}_m\| e^{2\pi i \frac{f}{c} \|\mathbf{x}_n - \mathbf{x}_m\|}$$

where c is the speed of sound and f is the frequency of the ultrasound waves, and for this system $f = 40000$ Hz. Humans can only feel the vibrations of frequencies up to 1 kHz [19], not those of ultrasound waves. In previous systems [22] [25] [21], a perceptible vibrotactile sensation was generated by applying an amplitude modulation of about 100–250 Hz because this method has a low acoustic power efficiency. Recently, it has been reported that a stronger stimulation is perceived by vibrating the focus position along the skin while keeping the acoustic power constant [27]. By moving the focal points around the contact surfaces of the finger and the object, our system efficiently generates a perceptible vibrotactile sensation and pressure pattern on the finger surface according to the shape of the contact surface. As shown in Figure 5 (a), the points whose distance to the object plane is less than a certain threshold value are extracted and approximated by an ellipse to estimate the sectional shape of the finger and the object. For an ellipse regression, we use the method proposed by Halir et al. [7]. By moving the focal points along the ellipses as shown in Figure 5 (b), the user can perceive a time-averaged pressure distribution spreading on the finger surface and thereby recognize the position of the object surface relative to the finger.

As described in the previous subsection, the prototype system clusters a point cloud within the virtual object by gathering points at distances within a certain threshold and regards each cluster as a finger. Because an ellipse regression is performed for each cluster, it is possible to generate pressure distributions on multiple fingers. If two nearby fingers touch a virtual object, they may be recognized as one, and one pressure ellipse is generated on their contact surface.

The architecture of the ultrasound transducer array used in the system is that proposed by Inoue et al. [14]. Because the phase and amplitude of the ultrasound transducers are switched at a constant period, the focal point moves over

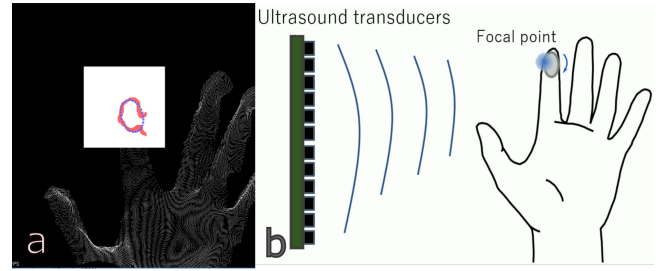


Figure 5: (a) The points close to the object surface (red) is approximated by an ellipse (blue) and the cross section shape of the finger is estimated. (b) The ultrasound focus revolves around the cross section of the finger.

several points on the ellipse discretely and sequentially. The focal position update rate is limited (up to 1000 Hz), so that there is a trade-off between the spatial smoothness of the pressure distribution generated on a fingertip and the simultaneity of the stimuli. The larger the number of points at which the focus is generated, the smoother the distribution, but the frequency of revolving around the ellipse decreases and the stimuli become less simultaneous. The current system provides 10 pressure points at 100 Hz to maintain a balance between the two.

Simulation of pressure pattern

We conducted a numerical simulation of the acoustic radiation pressure distribution of the finger surface based on the boundary element method (BEM). We used a mesh model 5 cm from the tip of the finger, which consists of 4322 meshes. Figure 6 shows the coordinate system and arrangement of the ultrasound transducers in the prototype system. The vertices of the mesh model were regarded as a point cloud obtained from a depth sensor, and our feedback method determined 10 focus positions and the output of each transducer when the finger touched the surface. Then, based on the BEM, the acoustic radiation pressure distribution on the finger surface is calculated. The numerical calculation method is the same as that used by Inoue et al. [13]. Figure 7 shows an example of the average of the pressure distributions, each of which occurs when one of the ten focal points is created. A pressure pattern conforming to the sectional shape of the finger is generated. Although the transducers are arranged surrounding the finger, no significant pressure is generated on the back of the finger because of interference. The spatial resolution (approximately 5 mm) seems higher than when a finger is irradiated only by a single panel of the phased array facing the finger.

The top row of Figure 8 shows the transition of the pressure distribution when the position of the surface changes. As the finger touches the surface, the pressure distribution

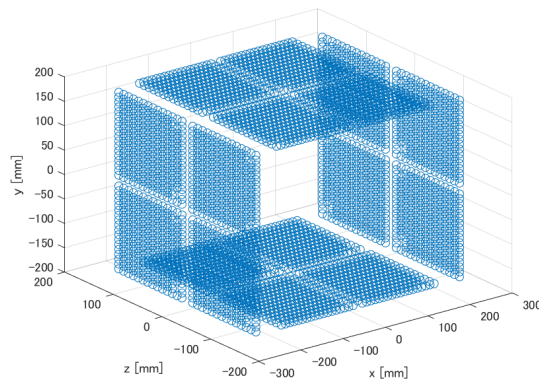


Figure 6: The coordinate system and the arrangement of the ultrasound transducers in the prototype system

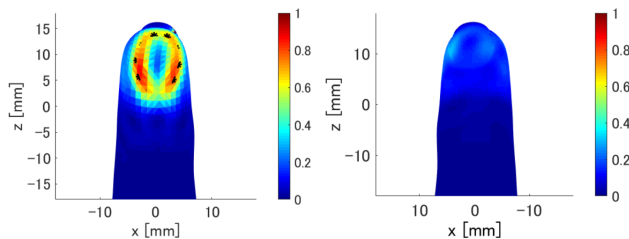


Figure 7: An example of a time-averaged acoustic radiation pressure distribution on the finger (left: finger pad, right: back). The asterisks represent positions of 10 focal points on the estimated ellipse. The distribution is normalized by the maximum value.

of the finger surface spreads (Figure 8 (1a)-(1c)). It is considered that the expansion of the pressure pattern makes the user feel an increase of contact surface than when touching the actual object, as a result, the reality of the object is enhanced. Furthermore, from the expansion of the pressure pattern according to the sectional shape of the finger, the user can recognize the position of the face relative to the finger. It is important to determine the position of the surface if the finger penetrates deep (Figure 8 (1d)), which can cause unintended object reactions in the physics simulation. We conducted experiments to examine the effect of ultrasound haptic feedback on the reality of an object and the position recognition of its surface. Furthermore, in the experiments, we compared our method of focus revolving around the estimated ellipse (hereinafter called dynamic focus method) to the method of generating only one focal point at the center of the ellipse (hereinafter called static focus method). The bottom row of Figure 8 shows the transition of the pressure distribution in the static focus method. The pressure distribution does not spread according to the entry of the finger,

and when the finger penetrates deep, a high-pressure region is generated off the virtual object surface.

4 USER STUDY

To evaluate the effectiveness of ultrasound haptic feedback and compare feedback methods quantitatively, we conducted two experiments. In these experiments, participants repeatedly perform tasks sitting in front of the prototype system as shown in Figure 9.

Experiment 1

We conducted an experiment in which participants were asked to estimate the angle and position of the object surface only from the haptic sensation. The purpose of this experiment is to find out how effective ultrasound haptic presentation is in recognizing the angle and position of the object surface.

Procedure. First, the face was displayed on the 3D display as shown in Figure 10 (a) and the participants were given time to get accustomed to the haptic feedback generated while touching the face for about 2 min. During this time, the participants could freely change the angle and position of the face using the keyboard, and perceive a change in the corresponding haptic feedback. Then, the subject repeatedly estimated the position and angle of the face using the tactile information only. As shown in Figure 10 (b), the surface that could be touched (haptic surface) and the surface that could be viewed (visual surface) were presented at different positions. The participants changed the position and angle of the visual surface to match the haptic feedback by keyboard operations. The participants continued this operation until the positions and angles of the two surfaces matched. The difference between the two surfaces after the operation was used to evaluate the extent to which the surfaces were recognized. The participants wore headphones during the experiment to prevent the influence of noise generated during the haptic feedback presentation. The visual surface was a square of side 5 cm. The tactile surface was a circle with radius 1 cm, and haptic feedback was presented only when the center of the ellipse was inside this circle. The above procedure was performed 20 times each during the dynamic and static methods. Six of the eleven participants performed the experiment with the dynamic focus method first and five performed the experiment with the static method first. Changes in the angle of the surface were limited to the roll direction with respect to the line of sight, and the angle of the haptic surface was randomly varied between -45° - 45° degree in each task. The participants could change the angle by 0.5° by one keyboard operation. The position of the haptic surface was limited to the movement in the vertical direction, and it was randomly changed in the range of 2 cm

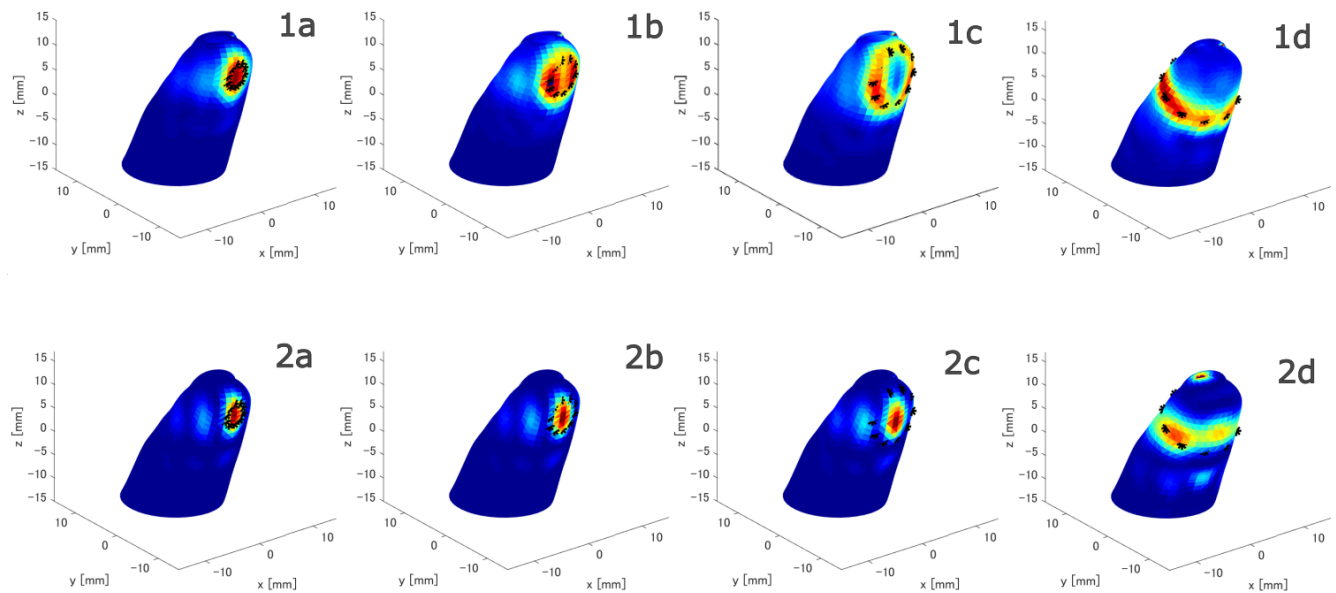


Figure 8: The transition of pressure distribution when the position of the surface changes. In the proposed method (top row), the value is the average of the distributions generated by 10 focal points at the positions marked by asterisks. In the static focus method (bottom row), the distribution is generated by a single focal point at the centroid of the asterisks.

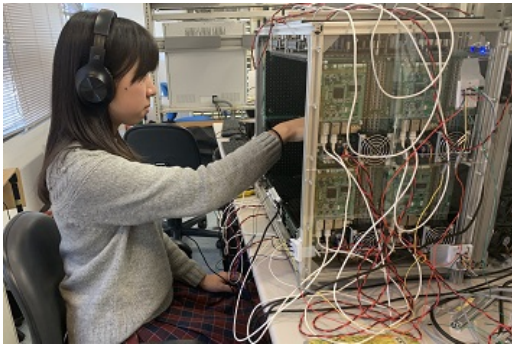


Figure 9: The participants perform the experiment sitting in front of the prototype system

in each task. The participants could change the position by 0.5 mm with one keyboard operation.

We used only one depth sensor (RealSense SR300) to detect the hand shape in this experiment as shown in Figure 10 (a), whose setup is different from that described in the previous section. Because the participants touched the virtual surface only from above, it was possible to measure the shape of the finger with only one depth sensor placed at the bottom of the virtual surface.

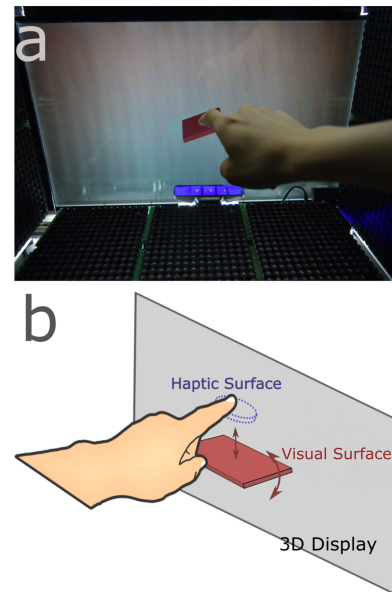


Figure 10: (a) The participants touched the surface on the 3D display and got accustomed to the haptic feedback. (b) The participants estimated the position and angle of the face from the haptic feedback and manipulated the face on the 3D display to match it.

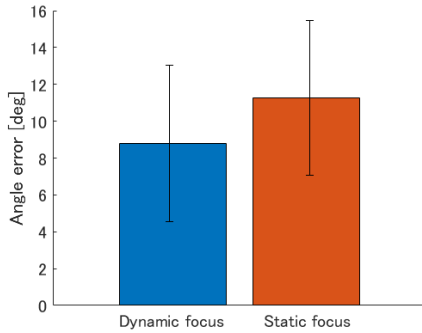


Figure 11: Mean and standard deviation of angle estimation error

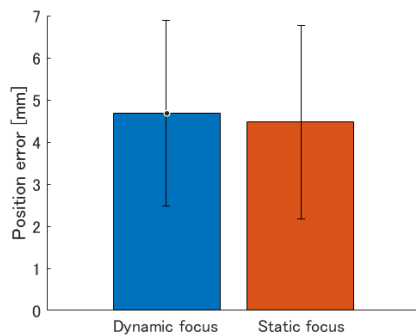


Figure 12: Mean and standard deviation of position estimation error

Result. In this experiment, there were 11 participants (six male and five female) aged between 21 and 26 and all right-handed. The mean and standard deviation of the estimation error are shown in Figure 11 and 12. For angle estimation, the average value using the dynamic focus method is smaller than that using the static focus method, and a paired t-test yielded a significant difference between them ($p < 0.05$). However, for position estimation, the difference between the dynamic and static focus methods is small, and there is no significant difference in the paired t-test ($p > 0.05$).

Discussion. The result of the angle estimation shows that ultrasound haptic feedback is effective in recognizing the angle of the surface, and in particular, the dynamic focus method gives the angle information to the user with a higher accuracy (error of about 9 degrees). In the position estimation, no significant difference was obtained between the two methods, but both estimation errors were about 4.5 mm, indicating that the position of the surface can be estimated using any of the feedback methods. An accurate recognition of the position of the surface only from tactile sensation makes it possible to grasp the object without looking at it.

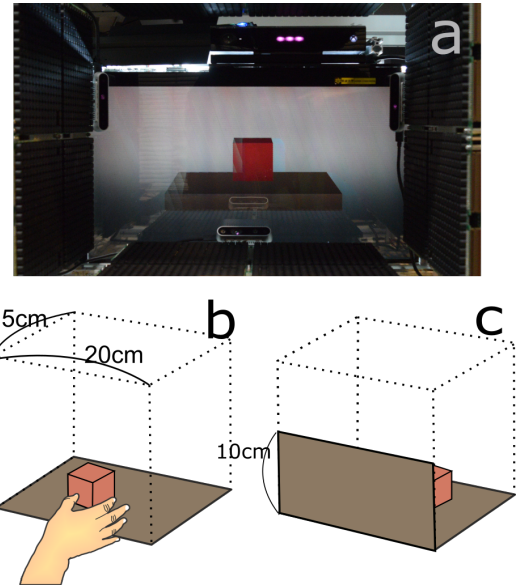


Figure 13: (a) A cube is displayed. The participants repeatedly performed the task of grasping and lifting it. (b) The operation area was covered with invisible walls. (c) The participants cannot see the cube behind the wall.

In this experiment, a change of only one degree of freedom is estimated for both the angle and position. However, even if we can recognize a change of only one degree of freedom, we can grab the cubes that are hidden from our view. For example, if the user knows the plane on which the bottom of the cube is placed, it is possible to estimate the state of the cube by recognizing the position and angle of the face with the first touch.

Experiment 2

For the second experiment, the participants performed the task of lifting the cubes with the condition that they can or cannot see it. We evaluated the operability of each feedback method from the time required for the operation. In addition, the participants answered a questionnaire and evaluated the subjective operability and reality of the object in each feedback method.

Procedure 1. A cube is displayed, as shown in the Figure 13 (a), and the participants repeatedly performed the task of grasping and lifting it under one of the three feedback conditions (dynamic focus, static focus, and no haptic feedback). When the cube was lifted by 5 cm or more for 1.5 s, one task was completed. We evaluated the operability of the system from the time of the first touch on the cube to the completion of the task. Three tasks under each of the three feedback conditions were regarded as one set and the participants performed twenty-five sets of tasks (total seventy-five tasks).

The order of the feedback conditions in one set was random. The participants wore headphones while performing the tasks to prevent the influence of noise generated during the haptic feedback presentation. The operation area was covered with invisible walls as shown in Figure 13 (b), so the cube never went out of that area. The initial position and angle of the cube were randomly determined in the operation area for each task. The side length of the cube was also random between 3 and 5 cm. After completing 75 tasks, the participants answered a questionnaire with the following questions:

(Q1) Were you able to lift the cube as you desired?

(Q2) Did you feel that the actual cube was present?

The participants answered these questions for each feedback condition by assigning a number from 1 to 7, where 1 meant "definitely no" and 7 meant "definitely yes."

Procedure 2. The participants performed the same tasks as in Procedure 1 with the wall in front of the operation area, as shown in Figure 13 (c); hence, they did not know the actual location of the cube. Although the wall is visible, haptic feedback is not generated while touching it. In the prototype system, because the height of the wall was 10 cm, the participants could not see the cube from above the wall. Lifting the cube made it possible to see it above the wall. Two tasks under each of the two feedback conditions (our method and static focus) were regarded as one set and the participants performed twenty-five sets of tasks (total fifty tasks). Then, they answered the same questionnaire, as in Procedure 1, for each of the two feedback methods.

Result. In this experiment, 10 people (three male and seven female), aged between 21 and 25, and all right-handed, participated. Figures 14 and 15 show the results of Procedure 1. Regarding the time taken to perform the task, a one-way repeated measures ANOVA yielded no significant differences between the three feedback conditions. However, for the questionnaire score, a Friedman test yielded significant differences between the three feedback conditions ($p < 0.05$). A Wilcoxon signed-rank test (Bonferroni corrected), performed as a post hoc test, showed a significant difference between all combinations of feedback conditions ($p < 0.05$).

The results of procedure 2 are shown in Figure 16 and 17. Regarding the time taken to perform the task, a paired t-test yielded a significant difference between the two conditions ($p < 0.05$). Regarding the questionnaire score, a Wilcoxon signed-rank test also showed a significant difference ($p < 0.05$).

Discussion. Considering that the mean value contains 1.5 s to keep the cube, the result of Procedure 1 suggests that in the system, a user can grasp and lift a cube easily using only visual information. Irrespective of whether haptic feedback was generated or not, the time taken for the task was

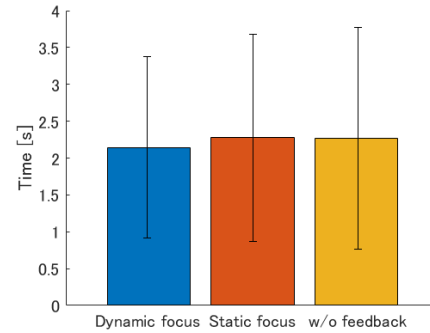


Figure 14: Mean and standard deviation of the time taken for performing a task with visual information

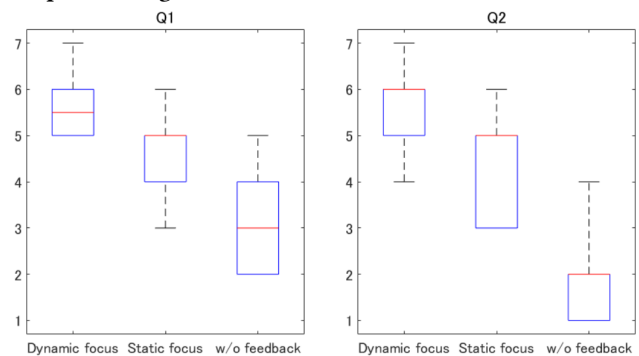


Figure 15: Box-and-whisker plot of question scores under the condition when the cube is visible

almost the same. The question scores show that the haptic feedback increased the ease of the subjective operation of the user. In particular, the dynamic focus method led to high operability. The same is true for subjective reality, therefore, it is suggested that the spread of the pressure distribution according to the increased contact area enhances the reality of the object.

The result of Procedure 2 shows that dynamic focus method is effective in grasping and lifting cubes that are not visible. In the system, users cannot lift a cube unless they grab the two faces on the opposite sides of the cube. Considering the results of Experiment 1, it is suggested that the angle recognition of the surface using dynamic focus method enables a fast operation. When performing operations such as 3D modeling, the occlusion problem is inevitable, but if the user can move the object using only tactile sensation, this problem can be solved without operating the view point. The question scores of Procedure 2 show that the user can feel high operability and reality of the object from haptic sensation, even if the object is not visible.

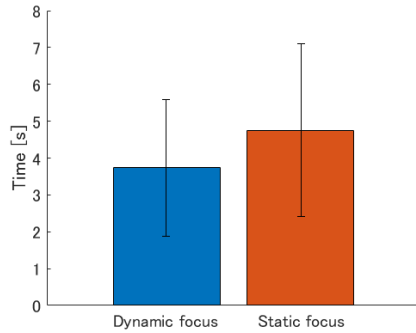


Figure 16: Mean and standard deviation of the time taken for performing a task without visual information

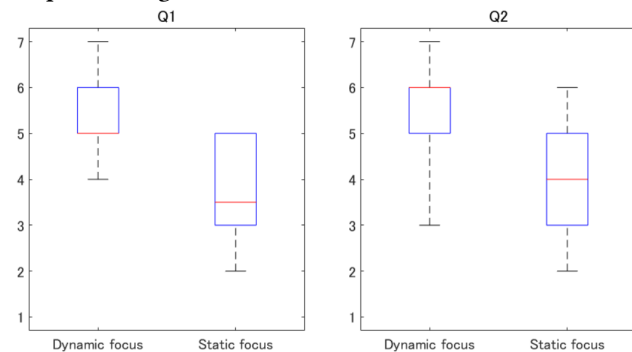


Figure 17: Box-and-whisker plot of question scores under the condition when the cube is hidden

5 LIMITATIONS

Ultrasound tactile presentation in the current system has several limitations. The displayed object seems very elastic owing to the upper limit of the force presented to the fingertip, and it is difficult to express the edges and corners of objects because of the limited resolution of the generated ultrasound distribution. Although the total force limit may be theoretical, there remains much room for improvement of haptic experiences. Our next step is to improve the spatial resolution of the tactile display, which can increase the perceived intensity under the total force limitation and widen the variety of tactile feeling. Hand sensing also faces a limitation. The depth sensors are arranged so that the fingers are always visible at least for typical hand postures. However, there still exist dead zones for certain postures (e.g., grasping a small object). This problem may be solved by estimating the shape of the hand from the obtained point cloud and complementing the hidden part. In that case, more advanced acoustic rendering will be necessary considering the interference of the hand with the finger surface.

6 CONCLUSIONS

In this study, we proposed and examined a new interactive system that enables a user to manipulate 3D object images with multiple bare fingers receiving haptic feedback. By estimating the cross section of the finger and the object, and moving the ultrasound focus around it, the system generates a pressure distribution on the finger surface according to the positional relationship between the finger and the surface. This pressure distribution increases the reality of the 3D object image and consequently helps the user to recognize the position of the object surface.

To evaluate the effectiveness of the ultrasound haptic feedback, we conducted two experiments comparing the dynamic and static focus methods. The result of the first experiment shows that the dynamic focus method improves the accuracy of the angle recognition of the object surface. The result of the second experiment shows that the dynamic focus method enables the user to hold and move the object easily even if it is not visible, and enhances the reality of the 3D object.

ACKNOWLEDGMENTS

The authors would like to thank Professor Hideki Kakeya for lending the stereoscopic display used in the research. This work is supported in part by the JSPS Grant-in-Aid for Scientific Research 16H06303 and JST CREST JPMJCR18A2.

REFERENCES

- [1] 2018. Bullet Real-Time Physics Simulation. <https://pybullet.org>
- [2] 2018. Intel Corporation. <https://www.intel.com/>
- [3] Gerald Bianchi, Benjamin Knoerlein, Gabor Szekely, and Matthias Harders. 2006. High precision augmented reality haptics. In *Proceedings of EuroHaptics*, Vol. 6. 169–178.
- [4] Alex Butler, Otmar Hilliges, Shahram Izadi, Steve Hodges, David Molyneux, David Kim, and Danny Kong. 2011. Vermeer: Direct Interaction with a 360° Viewable 3D Display. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. ACM, 569–576.
- [5] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 505–514.
- [6] L. R. Gavrilov. 2008. The possibility of generating focal regions of complex configurations in application to the problems of stimulation of human receptor structures by focused ultrasound. *Acoustical Physics* 54, 2 (2008), 269–278.
- [7] Radim HalÁs and Jan Flusser. 1998. Numerically stable direct least squares fitting of ellipses. In *Proceedings of 6th International Conference in Central Europe on Computer Graphics and Visualization*, Vol. 98. Citeseer, 125–132.
- [8] Seungju Han and Joonah Park. 2014. Holo-haptics: Haptic interaction with a see-through 3d display. In *Proceedings of the IEEE International Conference on Consumer Electronics*. IEEE, 512–513.
- [9] William S. Harwin and Nicholas Melder. 2002. Improved haptic rendering for multi-finger manipulation using friction cone based god-objects. In *Proceedings of the Eurohaptics conference*. Citeseer, 82–85.

- [10] Keisuke Hasegawa and Hiroyuki Shinoda. 2013. A method for distribution control of aerial ultrasound radiation pressure for remote vibrotactile display. In *SICE Annual Conference, 2013 Proceedings of*. IEEE, 223–228.
- [11] Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. 2012. HoloDesk: Direct 3D Interactions with a Situated See-through Display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2421–2430.
- [12] Seki Inoue, Yasutoshi Makino, and Hiroyuki Shinoda. 2015. Active touch perception produced by airborne ultrasonic haptic hologram. In *Proceedings of the World Haptics Conference*. IEEE, 362–367.
- [13] Seki Inoue, Yasutoshi Makino, and Hiroyuki Shinoda. 2016. Mid-Air Ultrasonic Pressure Control on Skin by Adaptive Focusing. In *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 68–77.
- [14] Seki Inoue, Yasutoshi Makino, and Hiroyuki Shinoda. 2016. Scalable Architecture for Airborne Ultrasound Tactile Display. In *Proceedings of the International AsiaHaptics conference*. Springer, 99–103.
- [15] Takayuki Iwamoto, Mari Tatezono, and Hiroyuki Shinoda. 2008. Non-contact method for producing tactile sensation using airborne ultrasound. In *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 504–513.
- [16] Hideki Kakeya, Ken Okada, and Hayato Takahashi. 2018. [Papers] Time-Division Quadplexing Parallax Barrier with Subpixel-Based Slit Control. *ITE Transactions on Media Technology and Applications* 6, 3 (2018), 237–246.
- [17] Nam-Woo Kim, Dong-Hak Shin, Dong-Jin Kim, Byung-Gook Lee, and Eun-Soo Kim. 2009. Vision-Based 3D Fingertip Interface for Spatial Interaction in 3D Integral Imaging System. In *Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems*. 1006–1011.
- [18] Georgios Korres and Mohamad Eid. 2016. Haptogram: ultrasonic point-cloud tactile stimulation. *IEEE Access* 4 (2016), 7758–7769.
- [19] P. J. J. Lamore, H. Muijsers, and C. J. Keemink. 1986. Envelope detection of amplitude-modulated high-frequency sinusoidal signals by skin mechanoreceptors. *The Journal of the Acoustical Society of America* 79, 4 (1986), 1082–1085.
- [20] Benjamin Long, Sue Ann Seah, Tom Carter, and Sriram Subramanian. 2014–11. Rendering Volumetric Haptic Shapes in Mid-air Using Ultrasound. *ACM Trans. Graph.* 33, 6 (2014–11), 181:1–181:10.
- [21] Yasutoshi Makino, Yoshikazu Furuyama, Seki Inoue, and Hiroyuki Shinoda. 2016. HaptoClone (Haptic-Optical Clone) for mutual tele-environment by real-time 3D image transfer with midair force Feedback. In *Proceedings of the Conference on Human Factors in Computing Systems*. ACM, 1980–1990.
- [22] Yasuaki Monnai, Keisuke Hasegawa, Masahiro Fujiwara, Kazuma Yoshino, Seki Inoue, and Hiroyuki Shinoda. 2014. HaptoMime: Mid-air Haptic Interaction with a Floating Virtual Screen. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. ACM, 663–667.
- [23] Sylvain Paris and Fr  do Durand. 2006. A fast approximation of the bilateral filter using a signal processing approach. In *Proceedings of the European conference on computer vision*. Springer, 568–580.
- [24] Radu Bogdan Rusu and Steve Cousins. 2011. 3d is here: Point cloud library (pcl). In *Proceedings of the IEEE International Conference on Robotics and automation*. IEEE, 1–4.
- [25] Antti Sand, Ismo Rakkolainen, Poika Isokoski, Jari Kangas, Roope Raisamo, and Karri Palovuori. 2015. Head-mounted Display with Mid-air Tactile Feedback. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. ACM, 51–58.
- [26] Rajinder Sodhi, Ivan Poupyrev, Matthew Glisson, and Ali Israr. 2013. AIREAL: interactive tactile experiences in free air. *ACM Transactions on Graphics* 32, 4 (2013), 134.
- [27] Ryoko Takahashi, Keisuke Hasegawa, and Hiroyuki Shinoda. 2018. Lateral Modulation of Midair Ultrasound Focus for Intensified Vibrotactile Stimuli. In *Proceedings of the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 276–288.
- [28] A. Talvas, M. Marchal, and A. L  cuyer. 2013. The god-finger method for improving 3D interaction with virtual objects through simulation of contact area. In *Proceedings of 2013 IEEE Symposium on 3D User Interfaces*. 111–114.
- [29] C. Ting, T. Jen, C. Chen, H. D. Shieh, and Y. Huang. 2016. 3D Air-Touch User Interface With High Touch Accuracy on Stereoscopic Displays. *Journal of Display Technology* 12, 5 (2016), 429–434.
- [30] Guo-Zhen Wang, Yi-Pai Huang, Tian-Sheuan Chang, and Tsu-Han Chen. 2014. Bare finger 3D air-touch system using an embedded optical sensor array for mobile displays. *Journal of Display Technology* 10, 1 (2014), 13–18.
- [31] Takumi Yoshida, Keitaro Shimizu, Tadatashi Kurogi, Sho Kamuro, Kouta Minamizawa, Hideaki Nii, and Susumu Tachi. 2011. RePro3D: full-parallax 3D display with haptic feedback using retro-reflective projection technology. In *Proceedings of the IEEE International Symposium on VR Innovation*. IEEE, 49–54.
- [32] Qu Zhang and Hideki Kakeya. 2014. A High Quality Autostereoscopy System Based on Time-Division Quadplexing Parallax Barrier. *E97-C*, 11 (2014), 1074–1080.
- [33] Craig B Zilles and J Kenneth Salisbury. 1995. A constraint-based god-object method for haptic display. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 3. IEEE, 146–151.