

Received 22 January 2024, accepted 4 March 2024, date of publication 7 March 2024, date of current version 19 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374782

RESEARCH ARTICLE

An Optical Design for Interaction With Mid-Air Images Using the Shape of Real Objects

SHOHEI ANDO¹ AND NAOYA KOIZUMI¹, (Member, IEEE)

Department of Informatics, The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan

Corresponding author: Naoya Koizumi (koizumi.naoya@uec.ac.jp)

This work was supported in part by JSPS KAKENHI Grant Number JP21K19821, and in part by Canon Foundation.

ABSTRACT Mid-air images, which are augmented reality (AR) technologies, enable computer graphics (CG) images to be superimposed on a physical space. The mid-air image can be placed side-by-side with real objects, allowing various interactions, such as directly manipulating them to contact the mid-air image on the same plane. In this case, the measurement of the shape of real objects is necessary to realize geometric consistency between the mid-air image and real objects. However, in mid-air image optics, real objects cannot be placed behind the mid-air image (i.e., at a position where they interrupt the light rays that form the mid-air image). This limits the placement of the sensor and may prevent accurate measurement of the shape of the real objects. Consequently, we proposed an optical system for interaction with mid-air images that virtually measures the shape of real objects from behind the mid-air image. In our system, a virtual infrared (IR) sensor is formed behind the mid-air image using a hot mirror that reflects only IR light. The optical system considers the visible area of the mid-air image and the measurable area of the sensor. We evaluated the sharpness, luminance, and chromaticity to assess whether the hot mirror had changed the appearance of the mid-air image. The results confirmed that there was little impact on user perception. Furthermore, we developed four supporting applications for our system to show its efficacy.

INDEX TERMS Mid-air interaction, augmented reality, mid-air image.

I. INTRODUCTION

Augmented reality (AR) technology, known for superimposing digital information onto real space, is used in various fields, including information visualization [1], [2], rehabilitation for disabilities [3], and entertainment [4], [5]. AR offers users a digital experience in their physical environment, through head-mounted displays (HMDs), handheld devices like tablets, or projection mapping. However, these traditional methods have limitations, including user inconvenience due to equipment mounting and the need for specific projection surfaces.

Mid-air image is an AR technology, which allows for the display of computer-graphics (CG) images next to real objects without physical equipment, visible to multiple viewers. The mid-air image is a real image formed in the air by light from a light source using retro-transmissive

optics, allowing the user to directly reach out the image. However, the intangibility of the mid-air images causes no interaction between mid-air images and real objects, neither visually or physically. Therefore, we must adapt the mid-air images based on the real-time positioning and movement of real objects to merge virtual and real objects seamlessly.

Geometric consistency is important for displaying the mid-air images alongside real objects, requiring accurate measurement of their shapes. Sato et al. [6] cited the “geometric consistency” as the key element for seamless integration of virtual objects and real space. Especially, because of the optical principle of the mid-air images, if there are opaque real objects that block the light rays towards the image formation plane, parts of the mid-air image may not be rendered. To avoid this, we need to measure their contact point (i.e., the position where they make contact) and display the mid-air images such that they do not overlap the real objects.

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi¹.

However, conventional mid-air image optics have limitations that the shape of the real objects may not be accurately measured or they may be obstructed. The principle of mid-air images makes it impossible to place the sensor behind the mid-air images. In this case, the position or shape of real objects required for interaction with mid-air images may not be sufficiently acquired, resulting in no geometric consistency.

The core of our research is an optical system that virtually measures the shape of real objects from behind the mid-air image using an IR sensor and a “hot mirror”, a transparent plate that reflects only IR light. An optical element for forming the mid-air image and the hot mirror are superimposed on each other, and visible and IR lights are used for image presentation and user measurement, respectively. This ensures that the optical axis of the sensor is orthogonal to the mid-air image plane, such that the shape of the real objects in this plane can be accurately measured, realizing the interaction with geometric consistency.

The following are the requirements that our system must meet.

- 1) The sensor’s measurable area must cover the mid-air image display area.
- 2) The mid-air image should not be degraded compared to conventional optical systems.
- 3) Achieve interaction that ensures geometric consistency between mid-air images and real objects.

First, the sensor must sense the entire mid-air image rendering area to control the mid-air image based on real objects, addressing requirement (1). We designed our optics considering the sensor and display positions to meet this requirement. Second, in the interactive experience with mid-air images, it is the mid-air image that the user sees the most, and a degradation in its sharpness and luminance diminishes the overall quality of the experience. Requirement (2) ensures that the mid-air image quality is not compromised by the hot mirror. We evaluated sharpness, luminance, and chromaticity to address the uncertainty of degradation. Third, consistent geometric alignment between real objects and mid-air images is vital for user comfort in interaction. Requirement (3) allows mid-air images and real objects to visually interact with geometric consistency. We implemented and demonstrated several applications to verify that our method satisfies requirement (3) and operates in practice.

This paper is organized as follows. Section I presents the background of the mid-air image system and the limitation of the conventional system. Section II discusses the research regarding mid-air images and interaction with AR objects. Section III presents the principle, optical design, and implementation of our system. Section IV explains the experimental setups and the results in evaluation of mid-air image quality. Section V demonstrates several supporting applications with our system. Section VI presents the discussion of the results along with limitations of our system and its potential improvements. Finally, Section VII summarizes the major conclusions drawn from this study.

II. RELATED WORK

A. MID-AIR IMAGING TECHNOLOGY

Several methods have been proposed to display images in the air and superimpose digital information on real space. They are similar to those in science fiction (SF) films and attract the interest of those imaginative, as reported by Norasikin et al. [7]. Tokuda et al. [8] and Rakkolainen et al. [9] proposed methods for projecting images onto fog displays. Although fog displays enable free-form aerial displays, they are susceptible to environmental factors, such as wind and humidity. LeviProps [10] employs a transparent lightweight piece of fabric that floats on ultrasonic waves as an aerial display. SonicSpray [7] uses ultrasonic vessel beams to generate laminar aerosol flows for precise control of aerial displays. In these methods, the speed of objects floating in the air is limited by their control mechanisms. Ochiai et al. [11] realized the rendering of volumetric graphics in the air using femtosecond lasers. However, laser-based methods require adequate safety precautions.

There are several optics that can be used to form CG images in the air. Retro-transmissive optics include micro-mirror array plates (MMAP) and dihedral corner reflector array (DCRA) [12]. In contrast, dual-axis retro-reflective optics include a roof mirror array (RMA) [13] and a retro-reflective mirror array [14]. Additionally, aerial imaging by retro-reflection (AIRR) [15] is one combination of the mid-air imaging system. In MMAP and DCRA, the mirror array reflects the light from the light source to form a mid-air image in a plane-symmetrical position relative to the element, regarding the light source. In RMA, the reflection of light through numerous small grooves forms a mid-air image. These optical elements function on their own and are easy to handle. AIRR is an optical system that forms mid-air images with a wide viewing area using a retro-reflective material and a beam splitter. The light transfer mechanism of these systems allows the appearance of the mid-air image to be instantly changed by redrawing the images in the display. We used MMAP because of its simple structure, easy installation, and ability to display high-sharpness mid-air images.

B. INTERACTION WITH OBJECTS IN THE AIR

The experience of touching and manipulating virtual objects in mid-air contributes to the immersive and realistic feel of XR. Several touch interactions with objects and displays floating in the air have been proposed. HoverPad [16] is an aerial display that can be operated hands-free using a crane to control the display and maintain it in the air. HapticSphere [17] uses an HMD connected to a finger by strings to realize precise touch interaction in a virtual space. Digitouch [18] uses a glove-like input device to support keyboard input in the air in a virtual space. In these systems, there are entities, such as display surfaces or gloves between the user and images, supporting suitable manipulations in the interactive systems. However, the display of CG images is limited to the frame of the physical display, and we cannot directly manipulate the images with real objects.

The following systems have achieved touch interaction with mid-air images. HoloDesk [19] combines an optical see-through display and a depth camera to provide the experience of touching and moving 3D objects floating in the air. The above system requires a mirror between the user and the image because it is an imaginary image. In contrast, the user can directly reach out to the formation position of the image and interact with it using the mid-air image of a real image. Vermeer [20] enables touch interaction with mid-air images omnidirectionally. Hunter et al. [21] and Takazaki et al. [22] solved the occlusion problem during touch interaction with mid-air images. Touching the void [23] projects the shadow of the user's finger detected by an IR camera onto the mid-air images to clarify where the user is touching. Furthermore, HaptoMime [24] and HaptoClone [25] are systems that use ultrasound to add tactile feedback to mid-air images, compensating for the disadvantage of mid-air images that cannot be touched because there are no physical substances.

In addition, various interactions are realized by using sensors installed in the optical system to measure areas other than the fingertips. MARIO [26], an interactive system with mid-air images, uses MMAP to measure blocks in real space and moves the mid-air image character to the highest block. SkyAnchor [27] tracks real objects with capacitive markers for mid-air images to follow the real objects. FairLift [28] uses ultrasonic distance sensors to measure the height of the water surface and provides the experience of scooping up the mid-air image. These studies aim to move mid-air images by manipulating physical objects because it is impossible to directly touch the mid-air images.

However, in the aforementioned mid-air image studies [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], the sensing directions are limited to from the oblique direction to the mid-air image formation plane or from above it. Therefore, only a portion of the geometry information in the mid-air image plane can be measured, and interaction with geometric consistency cannot be obtained. In addition, it is difficult to use active IR sensors that employ IR reflection, for example, when the sensor is via diffuser [20] or half-mirror [19].

III. DESIGN

We used the following optical system we proposed before [29]. Our system uses IR reflection that employs a hot mirror to realize the interaction with mid-air images, where the sensor measures real objects virtually from directly behind the mid-air image and it remains normally observable. A hot mirror is an optical element that only reflects IR light and only transmits visible light. Some hot mirrors are used for eye measurement using IR light in HMDs [30], [31]. A hot mirror is used to separate wavelengths, such as IR, for user measurement and visible light for image formation in this study.

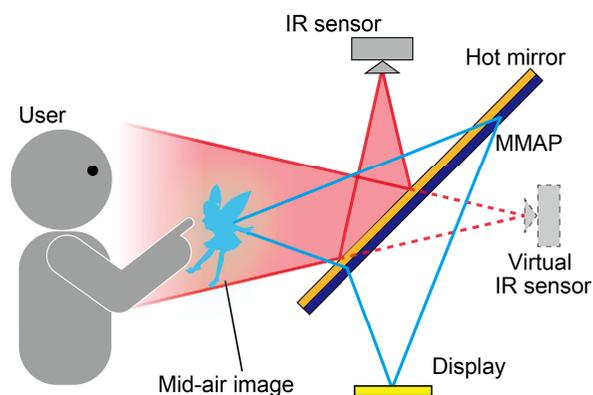


FIGURE 1. Optical design of our system. The hot mirror reflects IR light, allowing the IR sensor installed at the top to virtually measure the real objects from behind the mid-air image, that is from the position of “virtual IR sensor”.

A. SYSTEM DESIGN

Figure 1 illustrates our proposed optical system consisting of an IR sensor, display, MMAP, hot mirror, and light shields. The MMAP, which is installed at a 45-degree tilt, reflects the light emitted from the display several times and forms a mid-air image at a plane-symmetrical position to the MMAP. Meanwhile, an IR sensor installed at the top of the system receives the IR light reflected by the hot mirror. When an active IR sensor is used, the irradiated IR light is reflected in the order of the hot mirror, target, and hot mirror, and then reflected back to the sensor to enable measurement of the target.

In conventional mid-air image optics, the sensing method is inadequate for accurately capturing the shapes of real objects. The presence of an opaque object between the viewer and MMAP results in occlusion due to the inherent properties of mid-air images, as illustrated in Figure 2. Consequently, in previous systems, sensor placement is confined to specific locations, as depicted in Figure 3. This limitation leads to areas of obscurity, notably within regions, such as the diagonal line area. A method of capturing mid-air image positions by optically transferring the camera viewpoint has also been proposed [32]; however, the reduction of sharpness and insufficient distance to the target in the image formation position are problems. Fujii et al. [33] has proposed a mid-air image video calling system that uses AIRR to achieve optical eye-matching. This system uses polarized light reflections to observe the shape of real objects from behind the mid-air image. However, this system targets visible light, making it difficult to use depth sensors, which use reflection of IR light. In addition, implementation with an MMAP, which form mid-air images with higher sharpness than AIRR, has not been considered. In this study, we demonstrated the optical system that measures the shape of the real objects at the mid-air image formation position while maintaining the high sharpness.

In this study, our system uses the IR light reflection from a hot mirror to measure the real object shapes from the

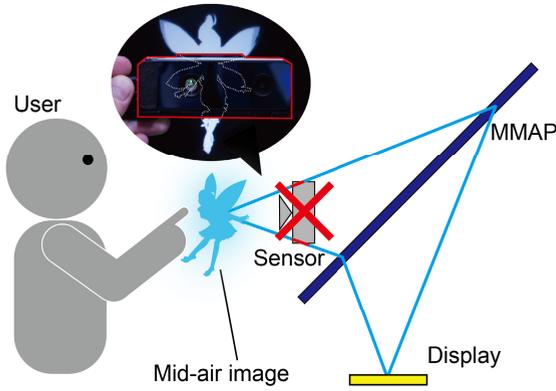


FIGURE 2. Sensors cannot be placed behind the mid-air image because it would be obscured.

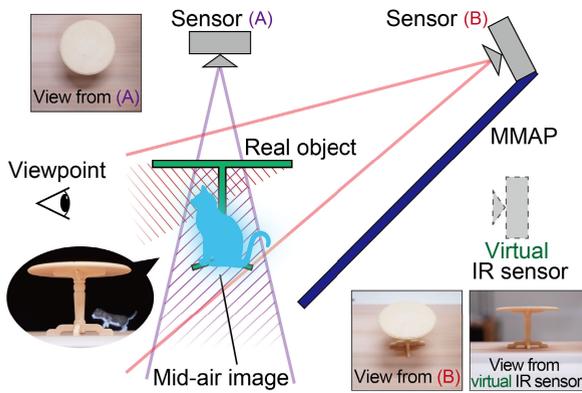


FIGURE 3. Problems with sensor placement in the optics for interaction with mid-air images. At the setup of sensors (A) and (B), there are areas shaded by the real objects. Picture-in-pictures show the sample views from sensors at each position.

position of the “virtual IR sensor”, as shown in Figure 1. This means that it can virtually measure them from behind the mid-air image, even though the actual IR sensor is installed at the top of the optics. The shape information of the real objects acquired like in this manner realizes an interaction with geometric consistency between the mid-air images and the real objects.

Because the visible light that forms the mid-air image passes through the hot mirror, the hot mirror theoretically causes no change in the visibility of the mid-air image. Although a half-mirror can be used rather than a hot mirror for measurement using reflection, the luminance of the mid-air image is reduced by half. Conversely, the luminance of the mid-air image is reduced by only approximately 95% because a hot mirror transmits visible light, as shown in Section IV.

In this study, a time-of-flight (ToF) sensor is used as an IR sensor. The ToF sensor is a depth sensor that estimates the distance to targets by measuring the time it takes for IR light to reflect off the targets. Compared to other types of depth sensors, the ToF sensor requires less computation and operates at higher speeds. This makes it possible to measure the shape, including depth information, of real objects in the mid-air image plane for interaction.

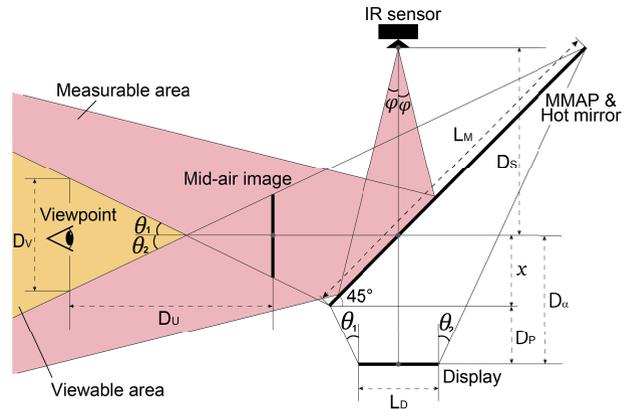


FIGURE 4. A “viewable area” where the user can observe the fully visible mid-air image using our method, and a “measurable area” where the IR sensor can measure the real objects.

B. VIEWABLE AREA & MEASURABLE AREA

We derived the “viewable area” in which the user can view the mid-air image and the “measurable area” that the IR sensor can measure to satisfy the requirement (1) described in Section I. It is important to consider the viewable area when implementing an application because mid-air images have a limited viewable area to observe depending on their size and location. In addition, we defined the measurable area needed to measure the real objects within the mid-air image formation area and derived the sensor placement and viewing angle that realizes that area. This enables interaction between the mid-air image and real objects controlled within the measurable area.

The derivation of the viewable area, where the entire mid-air image is visible without any part being unrenderable, is shown below. First, the position of the display relative to the MMAP is derived to determine the appropriate mid-air image position in this system. Herein, MMAP and hot mirror are considered to be on the same plane to simplify the calculation. As shown in Figure 4, in the case of viewing the mid-air image while moving in the vertical direction at a viewpoint, θ_1 and θ_2 should be equal to observing the mid-air image with a wider view area. $\theta (= \theta_1 = \theta_2)$ is obtained from Equation (1) using the popping up distance of the mid-air image D_P , size of the mid-air image L_D , and sizes of MMAP and hot mirror L_M .

$$\theta = \tan^{-1} \frac{L_M - \sqrt{2}L_D}{2\sqrt{2}D_P + L_M} \quad (1)$$

By deriving x from θ , the position of the display can be determined.

$$x = D_P \tan \theta + \frac{L_D}{2} \quad (2)$$

Furthermore, the distance D_V that the user can vertically move at the viewpoint can be obtained using Equation (3).

$$D_V = 2D_U \tan \theta - L_D \quad (3)$$

Subsequently, we determined the measurable area required for user interaction with the mid-air images. We assumed

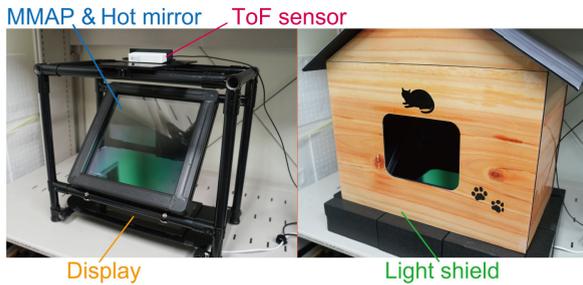


FIGURE 5. (Left) Implemented system. (Right) Light shields conceal the interior system from the users.

that the sensor is used to measure the real objects between the user's viewpoint and the hot mirror. Furthermore, the measurable area must cover the size of the mid-air image L_D to measure the real objects to be used to control the mid-air image. The minimum view angle of the sensor φ_{min} to realize such a measurable area is obtained using Equation (4).

$$\varphi_{min} = \tan^{-1} \frac{L_D}{2(D_S + D_\alpha)} \quad (4)$$

Based on the above, the optical system will satisfy the measurable area required for interaction if a sensor is installed with a view angle larger than φ_{min} for its optical axis to pass through the center of the display.

C. IMPLEMENTATION

Figure 5 shows the implementation of the our system. The display is a Magedok PI-X3 (3840 × 2160 px), the MMAP is ASKA3D-310 from ASUKANET (310 mm × 310 mm, 5 mm thick), the hot mirror is from Tokai Corporation (310 mm × 310 mm, 3.3 mm thick), and DCAM710 from Vzense is used for ToF sensor. The IR wavelength of the ToF sensor was selected to be 940 nm to avoid interference with visible light, and the hot mirror was custom-made to reflect the 940 nm IR wavelength. In this implementation, the L_D , D_P , and D_S are 50 mm, 100 mm, and 200 mm, respectively, and from Equations (1) and (2), θ and x are obtained as 22 degrees and 65 mm, respectively. Furthermore, the user can move within a D_V range of 83 mm at a viewpoint 500 mm away from the mid-air image. The view angle of the sensor $2\varphi_{min}$ needed to capture this range is 20.2 degrees, and the vertical view angle of the sensor used in this implementation is approximately 50 degrees, satisfying the condition. In addition, the frame of the enclosure and other parts are not preferred to be visible to the user to improve the appearance of the overall mid-air image device. Therefore, to the extent possible, styrene boards were used as light shields to conceal everything but the mid-air image.

Based on the depth information obtained from the ToF sensor, the shape of the real objects can be mapped to the CG space. Figure 6 shows the visualization of the obtained depth information, showing that the orange objects is in the back and the green objects is in the front, as seen from the user. Unity (version: 2019.1.8f1), a game development engine, was used to render the CG image displayed as a

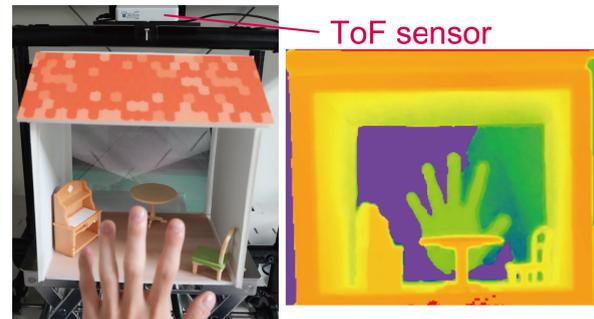


FIGURE 6. Depth data acquired by the ToF sensor.

mid-air image. We approximate the captured real objects with proxy cubes corresponding to each pixel of the sensor in the Unity space, and only the cubes of the pixel whose depth value exceeded the threshold at the mid-air image formation position were made active state. This makes it possible to implement interactions using shape information of the real objects in the Unity space.

IV. EVALUATION OF MID-AIR IMAGE QUALITY

We evaluated the sharpness, luminance, and chromaticity of the mid-air image to confirm that the requirement (2) described in Section I is satisfied. This examination determines whether placing the MMAP and hot mirror on top of each other causes a difference in the appearance of the mid-air image from the conventional mid-air image. Sharpness and luminance are reduced in conventional mid-air image optics when light rays are reflected or refracted by the MMAP. In addition, the hot mirror, integrated to our system, exhibit varying reflectance across different light wavelengths, potentially influencing the luminance and chromaticity of the mid-air image. These are the important factors in visual display. For instance, reduced luminance may result in diminished visibility of the image, while alterations in chromaticity could lead to the perception of an inaccurate color scheme for the user. Furthermore, we considered that the image sharpness is a vital visual metric in user perception because it defines the fineness of the image that can be displayed, and if it is reduced, for example, detailed text cannot be displayed. This aspect is crucial in assessing the performance of mid-air image engineering systems. Therefore, we measured these three factors in this system.

A. PROCEDURE

The modulation transfer function (MTF) was measured to investigate the sharpness of the mid-air image. MTF is a measure of image sharpness and is used to evaluate the sharpness of mid-air images [34], [35]. Figure 7 illustrates the setup of MTF measurement. The ISO 12233 Resolution Test Chart from Edmund Optics Japan Co. Ltd. was displayed as a mid-air image, and the edges were photographed from the front with a Sony α 7R V (image sensor 35.7 mm × 23.8 mm). The shutter speed was 1/20 s, ISO sensitivity was 100, F-number was 4.0, and the focus was adjusted to the mid-air

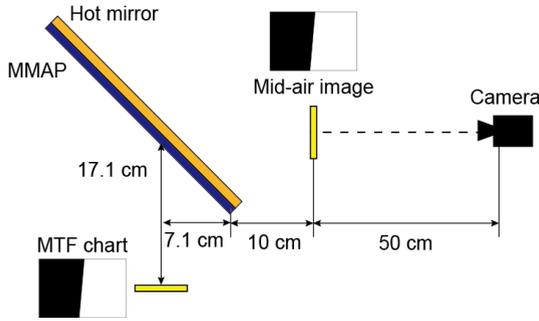


FIGURE 7. Setup of MTF measurement.

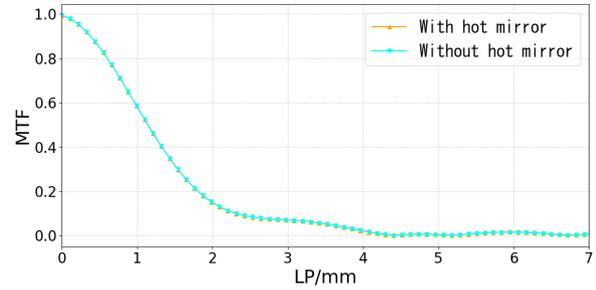


FIGURE 9. MTF measurement results. The two curves almost overlap, indicating no change in sharpness.

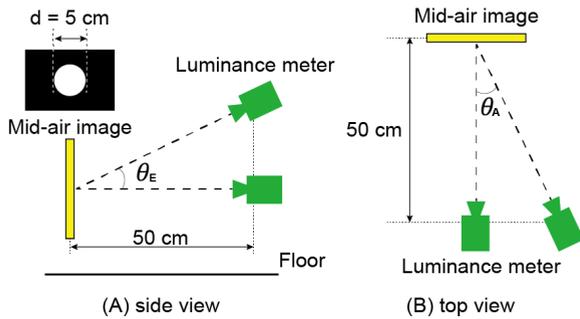


FIGURE 8. Setup of luminance and chromaticity measurement.

images. The test chart was captured with and without the hot mirror, and each MTF was calculated.

Additionally, the luminance and chromaticity of mid-air images with and without the hot mirror were investigated. A white circle with a diameter of 5 cm was displayed as a mid-air image, and the luminance and chromaticity were measured using a Konica Minolta CS-150 luminance meter. The elevation angle of the luminance meter and mid-air image were set to θ_E and the azimuth angle to θ_A , as shown in Figure 8. $\theta_E = \{-15^\circ, -10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ, 15^\circ\}$ and $\theta_A = \{0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ\}$ were used for measurements. These angles were determined based on the range where the mid-air image could be seen without missing. Negative values of θ_E indicate a tilt toward the floor, rather than the horizontal state, and the azimuth angle was measured only on one side due to the symmetry of the optical system.

B. RESULT

Figure 9 shows the sharpness measurement’s results. The MTF curve of the mid-air image with a hot mirror is shown in orange, and the one without is shown in cyan. The two curves almost overlap.

Table 1 summarizes the luminance measurement’s results. The values in the table represent the luminance ratios for each angle, or luminance of the mid-air image with the hot mirror relative to the one without it. This demonstrates how the luminance ratio tends to decrease as the elevation angle decreases and the azimuth angle increases. Note that, $(\theta_E, \theta_A) = (15, 20)$ is not correctly measured because the mid-air image is half covered by stray light (*1).

TABLE 1. Luminance measurement results. The values show the luminance ratio with and without the hot mirror.

Luminance ratio [%]		Azimuth angle (θ_A)				
		0°	5°	10°	15°	20°
Elevation angle (θ_E)	-15°	83.6	84.2	82.7	82.4	78.0
	-10°	91.4	91.8	89.5	90.3	88.2
	-5°	93.2	93.2	93.0	92.4	91.4
	0°	94.8	94.5	94.2	94.2	93.4
	5°	94.9	95.1	95.4	94.4	94.4
	10°	94.9	94.8	94.8	94.8	94.7
	15°	94.5	94.0	94.3	95.3	97.4(*1)

Table 2 lists the chromaticity measurement’s results. The values in the table represent the color difference ΔE for each angle. The luminance meter measures $X, Y,$ and Z values in the XYZ color system, which were converted to the values in $L^*a^*b^*$ color system using Equations (5).

$$\begin{aligned}
 L^* &= 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 \\
 a^* &= 500\left\{\left(\frac{X}{X_n}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}}\right\} \\
 b^* &= 200\left\{\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n}\right)^{\frac{1}{3}}\right\}
 \end{aligned} \tag{5}$$

where $X_n, Y_n,$ and Z_n are the $X, Y,$ and Z values on the reference white surface, respectively, and $X_n = 95.04, Y_n = 100.0,$ and $Z_n = 108.89$ were used here. When we let the difference of $L^*, a^*,$ and b^* under the conditions with and without a hot mirror be $\Delta L^*, \Delta a^*,$ and $\Delta b^*,$ respectively, ΔE is obtained using Equation (6).

$$\Delta E = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{6}$$

The results demonstrate that the color difference tends to increase as the elevation angle decreases and the azimuth angle increases. Similar to the luminance results, $(\theta_E, \theta_A) = (15, 20)$ is not measured properly due to stray light (*2).

C. DISCUSSION

Figure 9 presents that the impact of our method on image sharpness is minimal. The two nearly coincident curves indicate that the sharpness is not significantly different with or without the hot mirror.

TABLE 2. Chromaticity measurement results. The values show the color difference with and without a hot mirror.

Color difference (ΔE)		Azimuth angle (θ_A)				
		0°	5°	10°	15°	20°
Elevation angle (θ_E)	-15°	21.9	18.4	21.8	20.1	18.9
	-10°	8.8	8.3	14.6	10.6	11.2
	-5°	6.6	6.9	7.2	6.6	7.8
	0°	5.3	5.4	5.1	4.5	4.5
	5°	4.5	4.4	4.6	3.9	3.5
	10°	3.9	4.5	4.4	3.5	3.4
	15°	4.3	3.5	3.9	5.5	4.0(*2)

The hot mirror slightly reduced the luminance and chromaticity when the elevation angle θ_E was small. This is assumed to be caused by light passing through the MMAP at a large incident angle to the hot mirror. The hot mirror used in this implementation was manufactured to reflect IR light incident at a 45-degree angle. Visible light also shows different reflectance, transmittance, and absorbance depending on the incidence angle. It is considered that the smaller the elevation angle θ_E (i.e., the higher the incidence angle of visible light to the hot mirror), the lower the transmittance of visible light.

These differences should not significantly affect perception when users view mid-air images using this system. Tsuchiya et al. [32] measured a 55% luminance attenuation rate due to MMAP. However, our measurements indicate that hot mirrors have only a 5% to 10% luminance attenuation rate, which is smaller than that of MMAP, which is the dominant factor in the luminance attenuation of mid-air images. A light source with a high luminance is typically recommended to compensate for the luminance attenuation caused by MMAP; thus, the effect of the additional luminance attenuation in this system seems small. Additionally, the small change in chromaticity can be seen from the comparison with the color temperature reproduced by a typical display. There are two color temperatures that can be changed by the settings of the display used in this study, 6500 K and 9300 K, and the measured color difference (ΔE) for these was 26.4. The worst color difference measured in the experiment was smaller than this value, indicating that color representation is possible, although it is not suitable for cases where strict color reproduction is required. Incidentally, it is preferable to observe from a position with a large elevation angle to maintain high luminance and chromaticity. This condition tends to be satisfied during interaction with mid-air images because mid-air images are inevitably placed below the face for ease of operation. Considering this, we consider that the addition of a hot mirror is unlikely to cause perceptual problems with the quality of the mid-air images.

V. APPLICATIONS

We developed several examples of interaction with mid-air images to confirm that our system meets the requirement (3) described in Section I. We demonstrated the synergy between the characteristics of mid-air images, which can show real-time CG, and real objects, which can be directly manipulated because they have substance. This enables the provision of



FIGURE 10. In our system, a miniature umbrella is used to protect the character displayed as a mid-air image from getting wet in the rain. The picture-in-picture image shows that a depth camera can capture the shape of real objects.

experiences that cannot be achieved in CG space or real space alone. We exhibited the following applications at a domestic conference and campus events and observed more than 60 people enjoying the interaction between mid-air images and real objects with geometric consistency. Each of the applications is listed below.

A. HOLDING OUT AN UMBRELLA FOR THE CHARACTER

Our system realizes direct interaction with CG images using real objects rather than indirect interaction, such as using a controller in video games. Figure 10 shows an interaction that helps the mid-air image character by blocking the rain falling on her with a real object. The shapes of the miniature umbrella and hand at the mid-air image location are acquired by the ToF sensor and mapped into Unity space. The mapped real objects and the rain are judged for collision in Unity space, and the rain is drawn such that it is repelled at the collision location. Geometric consistency in the user's viewpoint is ensured because the rain image in the air is not rendered at the position where the umbrella or hand exists.

B. CREATING THE ROUTE FOR MID-AIR IMAGE OBJECTS

Users play with actual objects and CG images that work in a physical simulation to realize cyber-physical Rube Goldberg machines, for example. Figure 11 shows an interaction in which LEGO bricks form a course along which apples of the mid-air image roll. Similar to the aforementioned application, the shape of the LEGO bricks is acquired, and collision determination and physics calculations with the CG image are applied. When a real object (red LEGO block) is below another real object (yellow LEGO block), as shown in Figure 11, the sensor placed above the mid-air image (Figure 3, Sensor A) is unable to acquire the shape, whereas sensing from behind the mid-air image, as in our method, can measure the shape.

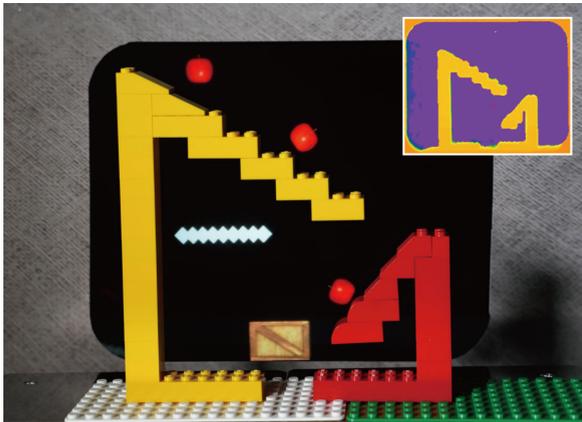


FIGURE 11. Mid-air image objects roll along a user-created course with LEGO bricks.



FIGURE 12. In a dollhouse, the cat finds a shape to land on and jumps there.

C. TROTTING CAT IN THE DOLLHOUSE

With our method, the shape of the real objects inside can be measured even if the area where the user can view the mid-air image is covered by walls on the top, bottom, left, and right sides. Figure 12 is an interaction with mid-air images in a dollhouse. The shape of the real objects inside is detected, and the mid-air image cat jumps onto a shape where it can land. Sensing from an oblique direction to the mid-air image, as shown in Sensor B in Figure 3, results in an area that is shielded because the ceiling, floor, and walls cover the top, bottom, left, and right sides of the mid-air image. This type of interaction within a space surrounded by real objects is expected to be applied in amusement facilities. For example, by placing cases in the front and hiding the optical elements, it is possible to create an experience as if the characters of an animation or video game are present in the stage or environment in the real space. The cases in this method can also hide the limitation of the viewing range of mid-air images and the problem of stray lights, creating an experience in which CG images and real objects are naturally integrated.

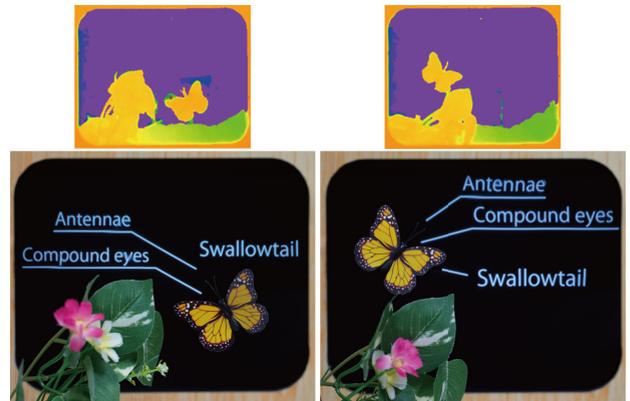


FIGURE 13. Annotations of mid-air images are displayed at positions that do not overlap with the shape of real objects obtained from the depth sensor.

D. ATTACHING THE ANNOTATION

Our system facilitates searching for a position where the entire mid-air image can be rendered (i.e., there is no part that is unrenderable) and presents the user with digital information that is appropriate for the real objects. Figure 13 shows the application for education, presenting annotations to users observing real objects. Grasping the shape of the real objects and displaying the image in a position in which they are not overlapped are necessary to annotate real objects using mid-air images. In addition, because the displayable area of mid-air images is limited, the relative position of the accompanying mid-air image must change as the real objects move. Our system contributes to deciding the proper positioning between them. Like this application, it is possible to implement content that supports the observer's understanding by adding digital information to realistic educational materials.

VI. DISCUSSION AND FUTURE STUDIES

Our method cannot acquire images using visible light because the sensor is limited to an IR sensor. Therefore, it is impossible to obtain color information on the surroundings of the user using our system. To use an RGB camera while hiding it from the user, we can consider (1) using a half-mirror rather than the hot mirror, (2) using polarized light [33], and (3) transmitting the camera images [32]. Method (1) cannot obtain clear images because the sensor also captures the light from the display transmitted through the half-mirror. Method (2) has difficulty obtaining polarization elements, and its combination with MMAP has not yet been developed. The images obtained using method (3) have low sharpness, and the viewpoint is positioned at the mid-air image position. However, combining method (3) with our system may allow color images to be used as supplemental images.

There are also limitations to the environments in which our system can be used. Some objects may be difficult to detect using IR sensors because the absorptivity and reflectivity of IR light vary from object to object. For example, black

objects, such as hair, tend to absorb IR light, and ToF sensors may not receive enough reflected light. Meanwhile, shiny surfaces, such as eyeglass frames and metal, reflect excessive IR light, causing IR sensors to detect false signals. Other questions remain regarding the acceptability of using real objects with shapes that are too complex. However, these are expected to be solved by using an IR camera and image processing. In addition, using outdoors can be difficult because sunlight interferes with IR and mid-air images, as with other motion capture systems for VR or AR, which recommends indoor use. As the effects of the aforementioned factors remain uncertain, future research is anticipated to focus on delineating the environmental conditions conducive to the successful application of our method.

Furthermore, the viewpoint to observe the mid-air image is limited to a certain range. Because of the principle of MMAP, the area of the mid-air image shielded by the real objects changes when the user's viewpoint moves. This also changes the position where the real objects and the mid-air image appear to be in contact with each other; thus, it is better for the user to view the mid-air image from the front of MMAP. This limitation may be solved by dynamically moving the sensor position depending on the user's viewpoint.

The advantage of our method is that it can be used to solve the occlusion problem of mid-air images. Based on the principle of optics, mid-air images cannot present a correct occlusion relationship with objects further back than the mid-air image. Hunter et al. [21] addressed this problem by using hand tracking to drill holes in the image where the hand and the mid-air image overlap. Our method can be used to take similar measures when interacting with mid-air images using real objects as well as hands.

Our method has scalability particularly for mid-air images in AR. It is not adaptable for existing AR technologies, such as projection mapping and HMD, but it can be applied to other mid-air image optics, such as AIRR and DCRA, as described in Section II. Additionally, it is conceivable to combine our system with some IR-based hand tracking, gesture recognition, and face detection, which are developed for AR systems. This would realize a rich extension of the interaction with mid-air images, which is becoming more widespread, to suit the application.

For future work, interaction using not only the shape of the real objects but also information on its three-dimensional position is expected. Moving the light source allows the mid-air image to move in the depth direction since the mid-air image's position depends on the position of the light source. Therefore, three-dimensional interaction can be implemented by moving the mid-air image position depending on the depth value acquired by the ToF sensor. Because our method is based on sensing from behind the mid-air image, it can also measure the depth movement of the real objects.

VII. CONCLUSION

In this study, we demonstrated a system for interaction with mid-air images that can measure the shape of real objects

to realize geometric consistency. Our method uses IR light reflection from a hot mirror to overcome the limitations of conventional optical systems, where sensors cannot be placed behind mid-air images. This enables the measurement of the shapes of real objects virtually from behind the mid-air images, resulting in an interaction between any real objects and the mid-air images with geometric consistency. We confirmed that the hot mirror has little effect on the quality of the mid-air image in our system. Furthermore, we designed several applications and verified that interactive mid-air images with geometric consistency can be realized with our system. This system is expected to be applied in the field of human computer interaction, such as the interaction with CG images using the affordances of real objects.

REFERENCES

- [1] R. Langner, M. Satkowski, W. Büschel, and R. Dachselt, "MARVIS: Combining mobile devices and augmented reality for visual data analysis," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–17.
- [2] F. Heinrich, L. Schwenderling, F. Joeres, K. Lawonn, and C. Hansen, "Comparison of augmented reality display techniques to support medical needle insertion," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 12, pp. 3568–3575, Dec. 2020.
- [3] G. A. Alshafeey, M. M. Lakulu, M. A. Chyad, A. Abdullah, and G. Salem, "Augmented reality for the disabled: Review articles," *J. ICT Educ.*, vol. 6, pp. 46–57, Jun. 2019. [Online]. Available: <https://ojs.upsi.edu.my/index.php/JICTIE/article/view/2637>
- [4] D. Eckhoff, A. Cassinelli, T. Liu, and C. Sandor, "Psychophysical effects of experiencing burning hands in augmented reality," in *Proc. Int. Conf. Virtual Reality Augmented Reality*. Cham, Switzerland: Springer, 2020, pp. 83–95.
- [5] S. Zollmann, T. Langlotz, M. Loos, W. H. Lo, and L. Baker, "ARSpectator: Exploring augmented reality for sport events," in *Proc. SIGGRAPH Asia Tech. Briefs*, Nov. 2019, pp. 75–78.
- [6] I. Sato, Y. Sato, and K. Ikeuchi, "Acquiring a radiance distribution to superimpose virtual objects onto a real scene," *IEEE Trans. Vis. Comput. Graphics*, vol. 5, no. 1, pp. 1–12, Jan./Mar. 1999.
- [7] M. A. Norasikin, D. Martinez-Plasencia, G. Memoli, and S. Subramanian, "SonicSpray: A technique to reconfigure permeable mid-air displays," in *Proc. ACM Int. Conf. Interact. Surf. Spaces*, Nov. 2019, pp. 113–122.
- [8] Y. Tokuda, M. A. Norasikin, S. Subramanian, and D. Martinez Plasencia, "MistForm: Adaptive shape changing fog screens," in *Proc. CHI Conf. Human Factors Comput. Syst.* New York, NY, USA: Association for Computing Machinery, May 2017, pp. 4383–4395.
- [9] I. Rakkolainen, J. Landkammer, M. Laitinen, M. Piirto, and K. Palovuori, "The interactive FogScreen," in *Proc. ACM SIGGRAPH Emerg. Technol.*, 2005, p. 8.
- [10] R. Morales, A. Marzo, S. Subramanian, and D. Martínez, "LeviProps: Animating levitated optimized fabric structures using holographic acoustic tweezers," in *Proc. 32nd Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2019, pp. 651–661.
- [11] Y. Ochiai, K. Kumagai, T. Hoshi, J. Rekimoto, S. Hasegawa, and Y. Hayasaki, "Fairy lights in femtoseconds: Aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields," *ACM Trans. Graph. (TOG)*, vol. 35, no. 2, pp. 1–14, 2016.
- [12] S. Maekawa, K. Nitta, and O. Matoba, "Transmissive optical imaging device with micromirror array," *Proc. SPIE*, vol. 6392, pp. 130–137, Oct. 2006.
- [13] Y. Maeda, D. Miyazaki, and S. Maekawa, "Aerial imaging display based on a heterogeneous imaging system consisting of roof mirror arrays," in *Proc. IEEE 3rd Global Conf. Consum. Electron. (GCCE)*, Oct. 2014, pp. 211–215.
- [14] T. Koike and Y. Onishi, "Aerial 3D imaging by retroreflective mirror array," in *Proc. ACM Companion Int. Conf. Interact. Surf. Spaces*. New York, NY, USA: Association for Computing Machinery, Nov. 2018, pp. 25–29, doi: [10.1145/3280295.3281365](https://doi.org/10.1145/3280295.3281365).

- [15] H. Yamamoto, Y. Tomiyama, and S. Suyama, "Floating aerial LED signage based on aerial imaging by retro-reflection (AIRR)," *Opt. Exp.*, vol. 22, no. 22, pp. 26919–26924, Nov. 2014. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-22-22-26919>
- [16] J. Seifert, S. Boring, C. Winkler, F. Schaub, F. Schwab, S. Herrdum, F. Maier, D. Mayer, and E. Rukzio, "Hover Pad: Interacting with autonomous and self-actuated displays in space," in *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol.*, 2014, pp. 139–147.
- [17] C.-H. Wang, C.-Y. Hsieh, N.-H. Yu, A. Bianchi, and L. Chan, "HapticSphere: Physical support to enable precision touch interaction in mobile mixed-reality," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2019, pp. 331–339.
- [18] E. Whitmire, M. Jain, D. Jain, G. Nelson, R. Karkar, S. Patel, and M. Goel, "DigiTouch: Reconfigurable thumb-to-finger input and text entry on head-mounted displays," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–21, Sep. 2017.
- [19] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson, "HoloDesk: Direct 3D interactions with a situated see-through display," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, May 2012, pp. 2421–2430.
- [20] A. Butler, O. Hilliges, S. Izadi, S. Hodges, D. Molyneaux, D. Kim, and D. Kong, "Vermeer: Direct interaction with a 360° viewable 3D display," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2011, pp. 569–576.
- [21] S. Hunter, R. Azuma, J. Moisant-Thompson, D. MacLeod, and D. Disanjh, "Mid-air interaction with a 3D aerial display," in *Proc. ACM SIGGRAPH Emerg. Technol.*, 2017, pp. 1–2.
- [22] M. Takazaki and S. Mizuno, "A method for appropriate occlusion between a mid-air 3DCG object and a hand by projecting an image on the hand," in *Proc. ACM SIGGRAPH Emerg. Technol.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–2.
- [23] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung, "Touching the void: Direct-touch interaction for intangible displays," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 2625–2634.
- [24] Y. Monnai, K. Hasegawa, M. Fujiwara, K. Yoshino, S. Inoue, and H. Shinoda, "HaptoMime: Mid-air haptic interaction with a floating virtual screen," in *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2014, pp. 663–667.
- [25] Y. Makino, Y. Furuyama, S. Inoue, and H. Shinoda, "Haptoclone (haptic-optical clone) for mutual tele-environment by real-time 3D image transfer with midair force feedback," in *Proc. CHI*, San Jose, CA, USA, 2016, pp. 1980–1990.
- [26] H. Kim, I. Takahashi, H. Yamamoto, S. Maekawa, and T. Naemura, "MARIO: Mid-air augmented reality interaction with objects," *Entertainment Comput.*, vol. 5, no. 4, pp. 233–241, Dec. 2014.
- [27] H. Kajita, N. Koizumi, and T. Naemura, "SkyAnchor: Optical design for anchoring mid-air images onto physical objects," in *Proc. 29th Annu. Symp. User Interface Softw. Technol.*, Oct. 2016, pp. 415–423.
- [28] Y. Matsuura and N. Koizumi, "Fairlift: Interaction with mid-air images on water surface," in *Proc. ACM SIGGRAPH Emerg. Technol.*, Aug. 2018, pp. 1–2.
- [29] S. Ando and N. Koizumi, "Floagent: Interaction with mid-air image via hidden sensors," in *Proc. SIGGRAPH Asia Emerg. Technol.*, Dec. 2022, pp. 1–2.
- [30] J. Rekimoto, K. Uragaki, and K. Yamada, "Behind-the-mask: A face-through head-mounted display," in *Proc. Int. Conf. Adv. Vis. Interfaces*, 2018, pp. 1–5.
- [31] M. Stengel, S. Grogorick, M. Eisemann, E. Eisemann, and M. A. Magnor, "An affordable solution for binocular eye tracking and calibration in head-mounted displays," in *Proc. 23rd ACM Int. Conf. Multimedia*, Oct. 2015, pp. 15–24.
- [32] K. Tsuchiya and N. Koizumi, "An optical design for avatar-user co-axial viewpoint telepresence," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2020, pp. 108–116.
- [33] K. Fujii, N. Endo, N. Hagen, M. Yasugi, S. Suyama, and H. Yamamoto, "Aerial video-calling system with eye-matching feature based on polarization-modulated aerial imaging by retro-reflection (p-AIRR)," *Opt. Rev.*, vol. 29, no. 5, pp. 429–439, Oct. 2022.
- [34] N. Kawagishi, R. Kakinuma, and H. Yamamoto, "Aerial image resolution measurement based on the slanted knife edge method," *Opt. Exp.*, vol. 28, no. 24, pp. 35518–35527, Nov. 2020. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-28-24-35518>
- [35] N. Koizumi and A. Sano, "Optical system to display mid-air images on a glossy plane and remove ground images," *Opt. Exp.*, vol. 28, no. 18, pp. 26750–26763, Aug. 2020. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-28-18-26750>



SHOHEI ANDO received the B.E. degree from The University of Electro-Communications, Japan, in 2022. His research interests include mid-air imaging and human–computer interactions.



NAOYA KOIZUMI (Member, IEEE) received the B.E. and M.E. degrees from The University of Electro-Communications, Japan, in 2004 and 2006, respectively, and the Ph.D. degree in media design from Keio University, in 2012. He was a JSPS Postdoctoral Researcher with Keio University, in 2013, and a Project Researcher with The University of Tokyo, from 2013 to 2016. He is currently an Associate Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications. His research interests include entertainment computing and augmented reality.

• • •