Enhancing the Sensation of Depth in Mid-Air Image Interactions with Pictorial Depth Cues

Saki Kominato
kominato@media.lab.uec.ac.jp
The University of Electro-Communications
Department of Informatics
Chofu, Tokyo, Japan

Naoya Koizumi
koizumi.naoya@uec.ac.jp
The University of Electro-Communications
Department of Informatics
Chofu, Tokyo, Japan

Abstract

In virtual reality, visual information plays a critical role, and head-mounted displays are widely recognized as the primary means of presentation. However, non-wearable approaches such as projection mapping and mid-air images have also been explored. Mid-air images present content near real objects without screens, making them promising for mixed reality. Yet, their lack of physicality weakens depth perception and diminishes the sensation of pressing buttons. We tested whether pictorial cues (shading, shadow, size) enhance depth perception and button-press sensation in mid-air image UIs. Each experiment involved 14–16 participants. These cues increased perceived depth and improved pressing sensation. These findings suggest that pictorial cues can compensate for the absence of physical sensation and enhance the usability of mid-air image UIs.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**.

Keywords

mid-air image, perception, depth, thickness

ACM Reference Format:
Saki Kominato and Naoya Koizumi. 2025. Enhancing the Sensation of Depth in Mid-Air Image
Interactions with Pictorial Depth Cues. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3756884.3765994

1 Introduction

Two-dimensional (2D) images and videos can be perceived as three-dimensional (3D) by utilizing pictorial depth cues. Here, by pictorial depth cues we refer to occlusion, relative/familiar size, linear perspective, texture gradients, height in the visual field, aerial perspective, shading, and cast shadows [5]. These cues convey 3D structure and relations, shaping perceived distance and object thickness. The sense of depth refers both to perceived spatial distance between objects and to the perceived thickness of individual objects.

We study how pictorial depth cues can enhance 2D mid-air images—free-space real images formed by relaying a flat display. While pictorial cues have been widely explored for HMD-based AR/VR [1, 6], they are not directly applicable to mid-air images because shadow-based approaches generally need an extra surface or display [25]. Hence, we focus on in-image cues implementable without additional hardware.

Based on informal observation, cues such as shading or relative size can give users a sense of depth in mid-air images. If confirmed effective, this study would represent the first application of pictorial depth cues to mid-air image UI buttons, offering clear implications for future interface design.

In this study, we investigate the effect of adding pictorial depth cues to 2D mid-air images on the enhancement of perceived depth. Although several methods exist to measure perceived depth, we chose the pointing method to verify the effects of pictorial depth cues in an interactive experiment. In this method, users physically point to the location in space corresponding to the perceived depth, enabling the first quantitative measurement of perceived depth in mid-air images.

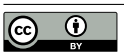
This paper makes the following contributions to interaction with 2D mid-air images:

- Motivation & problem framing: We articulate why 2D mid-air images inherently lack binocular disparity and motion parallax, motivating the use of pictorial depth cues.
- Methodological contribution: We operationalize a pointing-based measure of perceived depth for mid-air images and verify its accuracy for depth discrimination.
- Perceptual findings: We quantify the effects of three pictorial cues—shading, background shadow, and size difference—on perceived depth in 2D mid-air images.
- UI design implication: We show that combining size difference with a co-moving shadow enhances pressing sensation and increases penetration relative to the focal plane in mid-air button UIs.

2 Related Work

2.1 Mid-air Images

Mid-air images are images formed in real space by light emitted from a light source, which is reflected or refracted by optical elements. The optical systems for mid-air images can be categorized into those that form images symmetrically opposite to the light source display and those that form images on the same side as the light source. Examples of the former include Dihedral Corner Reflector Array (DCRA) [14], Micro Mirror Array Plate (MMAP) composed of two layers of slit mirror arrays, and aerial imaging by retro-reflection formed by combining retroreflective materials and



VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/2025/11
https://doi.org/10.1145/3756884.3765994

a half mirror [24]. Roof mirror array [11] is an example of the latter. These systems allow naked-eye viewing and multi-user interaction without HMDs(head-mounted displays); we use MMAP.

Interactive systems leveraging mid-air images have been extensively researched. In MARIO [10], users can stretch their hands toward mid-air images to manipulate blocks within a 3D space. MiTAi [15] enables continuous movement of mid-air images between mirror space and physical space, allowing users to touch and move mid-air images with their fingers. HaptoFloater [17] integrates haptic feedback into mid-air images, creating interactions where mid-air images feel as though they possess physical substance. AIR-range [8] displays mid-air images seamlessly rising from a table surface, offering a visual experience where mid-air images appear naturally present in the real-world space. Systems have also been proposed that reproduce occlusion relationships when users stretch their hands toward mid-air images, enhancing the sense that mid-air images exist in space [20]. Additionally, interactions have been achieved where the shapes of real objects are used to maintain geometric consistency with mid-air images, creating the illusion that the two are fused within the same space [2]. These studies contribute to natural interactions where users can directly manipulate mid-air images with their hands, enhancing the immersion of mid-air images as if they exist in real-world space.

However, many of these systems rely on 2D displays as light sources, limiting the display content to two-dimensional planes and lacking a sufficient sense of depth in the images. This issue stems from the limited adoption of 3D displays and implementation challenges, such as reversed depth relationships when 3D displays are used as light sources. Consequently, methods to provide more depth-rich images using 2D displays are required.

To address this challenge, this study attempts to add a sense of depth to mid-air images by utilizing pictorial depth cues. This approach aims to achieve natural depth representation in mid-air images, as if they exist in real-world space, without relying on 3D displays.

2.2 Depth Perception with Pictorial Cues

Pictorial depth cues, such as shading, shadows, and size differences, are known to influence depth perception. In this study, we focus on these three cues as the main factors to be examined. Previous studies have demonstrated the effects of pictorial depth cues in various environments. McIntosh et al. demonstrated that exaggerating pictorial depth cues, such as linear perspective and texture gradient, on backgrounds where targets are presented leads to an increased tendency in the perceived depth positions [16]. Adams et al. showed that presenting shadows beneath virtual objects in optical see-through augmented reality and video see-through augmented reality environments improves the accuracy of perceived positions [1]. Vienne et al. reported that environments with multiple cues (perspective, occlusion, shading) yield more precise depth perception [21]. Furthermore, studies have indicated that stereoscopic display environments benefit from improved depth perception owing to the effects of linear perspective [9].

On the other hand, the effects of pictorial depth cues on mid-air images remain insufficiently understood, particularly regarding depth perception in situations where shadows are not presented beneath the mid-air images. Kim et al. proposed a method to make depth positions easier to comprehend by displaying shadows beneath mid-air images [10]. Yano et al. reported that manipulating the length of shadows displayed beneath mid-air images alters the perceived thickness—the longer the shadow, the thicker the image is perceived; the shorter the shadow, the thinner it appears [25]. In this study, we investigate the effects of pictorial depth cues such as shading, background shadows, and size differences under conditions where mid-air images do not cast shadows beneath them, a setting that has received limited attention.

2.3 Methods for Estimating Depth Position

Several indirect approaches exist for measuring the perceived depth position of objects. The simplest method is verbal reporting, where participants estimate the perceived depth position using verbal descriptions. This method has been employed to measure the perceived positions of virtual objects in VR and video see-through augmented reality environments [22]. However, it relies on participants' subjective interpretations and may not accurately represent the actual depth position [3].

Many studies have used pointing methods, where participants indicate an object's position by pointing with a finger. For instance, McIntosh et al. evaluated depth perception by having participants directly point at targets displayed on backgrounds containing pictorial depth cues [16]. Mengdi et al. measured depth positions in VR environments by using a virtual hand cursor in the shape of a hand to point directly at the virtual objects [19]. Additional approaches include using a stick to point at depth positions of spheres displayed in stereoscopic environments [4, 13], and Fisher et al.'s method, where participants slide their finger beneath the stimuli to measure depth positions [7]. Unlike verbal reporting, pointing methods allow participants to directly indicate the perceived position of objects while they are still visible.

Pointing lets participants indicate perceived positions directly, avoids screen contact, matches real mid-air interactions, and is thus suitable for measurement. Since mid-air images are formed in space rather than on a display surface, there is no concern about the finger unintentionally contacting a screen when reaching for the perceived location. Moreover, in actual mid-air image interactions, it is assumed that users will reach toward the mid-air image [2, 10, 15, 17, 20], thus enabling evaluation under conditions that closely resemble practical usage. For these reasons, the pointing method is considered suitable for measuring the perceived position of mid-air images.

On the other hand, the pointing method also has some disadvantages. For example, variations in finger stability or individual motor ability can lead to inconsistencies in pointing accuracy. This issue is investigated in Experiment 1, the pointing accuracy experiment. Furthermore, when pointing toward a mid-air image, the hand may overlap with the image, potentially occluding the pictorial depth cues and affecting depth perception. To mitigate this issue, we adjusted the finger angles and pointing positions during the depth evaluation experiment.

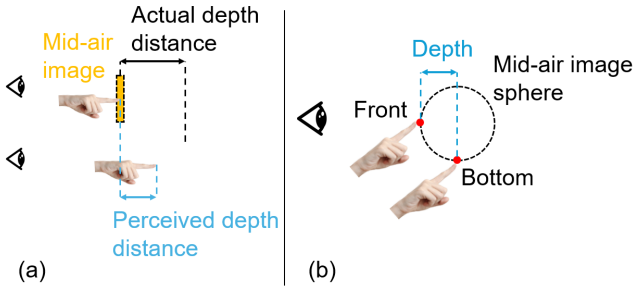


Figure 1: Pointing tasks: (a) Experiment 1 — Pointing accuracy, (b) Experiment 2 — Depth perception evaluation

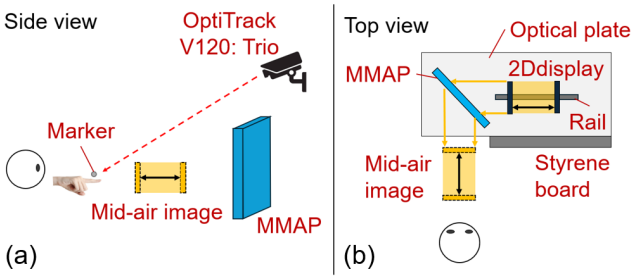


Figure 2: Optical design of the system: (a) Side view, (b) Top view

3 Overview of Experiments

The overall goal of this study was to validate the accuracy of the pointing method and to quantitatively evaluate depth perception in mid-air images using this method. To this end, we first conducted Experiment 1 to verify the method's accuracy, then Experiment 2 to measure perceived depth using the validated method, and finally Experiment 3 to explore the application of these findings to mid-air image UIs.

In Experiment 1 (pointing accuracy experiment), as shown in Figure 1 (a), participants were asked to reach out and point to the perceived depth position of the mid-air image. By comparing the pointed position with the actual imaging plane, we assessed the accuracy and reliability of the pointing method.

In Experiment 2 (depth perception evaluation experiment), as illustrated in Figure 1 (b), participants were asked to point to two different locations on the 2D mid-air image. The difference in the depth direction between the two pointed positions was measured to quantitatively evaluate the perceived magnitude of depth. Note that while Figure 1 (b) presents the mid-air image in a 3D representation for explanatory purposes, in the actual experiment, the mid-air image was formed on a 2D plane, and the pointing task was conducted accordingly.

In Experiment 3 (mid-air image UI experiment), we applied the effective pictorial depth cues identified in Experiment 2—shadow and size difference—to mid-air image buttons. We evaluated, using the pointing method, whether these cues enhance perceived depth and improve the sensation of pressing in user interactions.

4 Experiment 1: Pointing to Mid-Air Images

In this experiment, participants pointed to the perceived depth positions of mid-air images formed at different depth positions. By comparing the indicated positions with the actual imaging planes, we assessed the pointing accuracy.

4.1 Optical Design and Implementation

In this system, a 2D display and an MMAP were used to present 2D mid-air images. As shown in Figure 2 (b), light from the 2D display passes through the MMAP, forming a mid-air image at a position symmetrical to the display with respect to the MMAP plane. To adjust the depth position of the mid-air image, the display was mounted on a sliding rail installed beneath it. Additionally, an optical motion capture system was used to track the position of markers attached to the participants' fingers, enabling precise measurement of pointing positions. To prevent participants from inferring the depth position of the mid-air image based on the visible position of the display, the display was covered with a black styrene board.

The experimental system consisted of a 2D display, an MMAP, and an optical motion capture system. For the 2D display, we used an Intehill high-resolution mobile monitor F13NA (13.3-inch, 400 cd/m^2 brightness, 1920×1080 resolution). The MMAP was an ASKANET ASKA3D plate (360 × 360 mm)[1]. To measure the pointing positions, we used an optical motion capture system, the OptiTrack V120:Trio. To reduce stray light, a louver film was attached to the display. Furthermore, a sliding rail was installed on an optical breadboard, and the position of the display was precisely adjusted using the 2.5 cm spacing between holes on the breadboard as a reference.

4.2 Method

We presented seven depths (35–50 cm, 2.5-cm steps) at 80 cm viewing distance with a chinrest, the target image is shown in Figure 3. Consequently, from the observer's viewpoint, the mid-air images appeared to be located within a depth range of 30 cm to 45 cm. The range of depth positions was determined based on the near point (30 cm) and approximately two-thirds of the average arm length of young Japanese females (67.31 cm) [12, 18]. To measure the pointing positions, we mounted a marker near the index base, corrected to the fingertip offset, recorded 1 s per trial, and ran the study in a dark room.

4.3 Procedure

For each trial, participants pointed to the perceived depth position of the mid-air image, and the fingertip position was recorded. Seven imaging planes were positioned 35–50 cm from the MMAP center toward the observer in 2.5-cm steps. The presentation order was randomized using a Latin square design. Five trials were conducted for each imaging plane, resulting in a total of 35 trials per participant. Participants closed their eyes during changes of the imaging plane. The total time required for each participant was approximately 20 minutes.

14 individuals (12 males, 2 females) aged between 21 and 24 years participated in the experiment. Among them, 6 participants were

[1]https://aska3d.com/ja/technology.html?lang=en

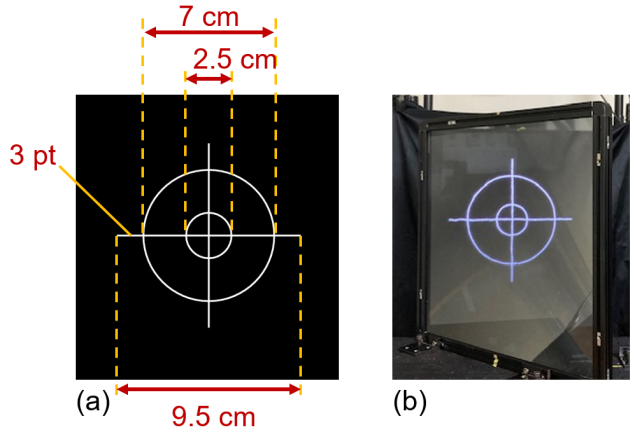


Figure 3: Stimuli for Experiment 1: (a) Image displayed on the 2D display, (b) Actual mid-air image

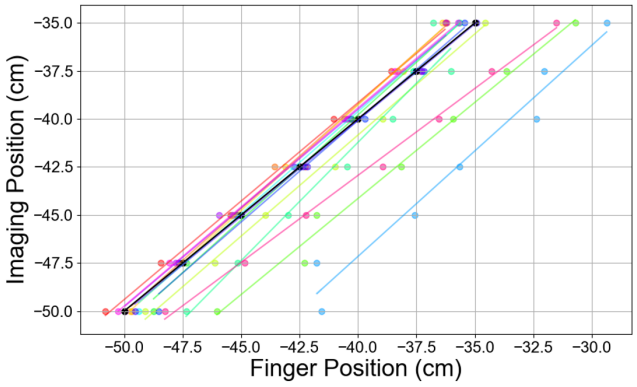


Figure 4: Perceived depth positions of the mid-air image at each imaging plane. Different colors represent individual participants. The black dots and black line indicate the ideal relationship where the pointing position perfectly matches the actual imaging plane.

engaged in research related to mid-air images. All participants had normal or corrected-to-normal vision, with no visual impairments affecting daily life. This study was approved by the Research Ethics Committee of the University of Electro-Communications (Approval No. H24053).

4.4 Results

Figure 4 shows a summary of the perceived depth positions of the mid-air image at each imaging plane for all participants. The horizontal axis represents the pointing positions measured by the participants (unit: cm), where larger values indicate positions farther away (toward the MMAP), and smaller values indicate positions closer to the participant. The vertical axis represents the actual imaging plane of the mid-air image (unit: cm), following the same convention. For each imaging plane, the average of the five trials per participant, using the median fingertip position over one second for each trial, was plotted to minimize small movements and ensure values closely reflected the perceived positions. Based on these data, a regression line was fitted for each participant.

Perceived positions tracked imaging-plane changes; mean slope error was 5.5%, indicating valid depth discrimination despite some individual offsets. Although some participants show offsets of 5.0–7.5 cm, these do not substantially affect depth perception, as there is no large difference in the slopes. Furthermore, the average slope of the regression lines for all participants deviated by only 5.5% from the ideal slope represented by the black line, confirming that the pointing method can validly indicate depth differences in Experiments 2 and 3. These results demonstrate that participants were able to accurately perceive the changes in the imaging plane and adjust their pointing positions accordingly.

Although some individual differences in pointing positions were observed, there was no overall tendency for large discrepancies between the pointing positions and the actual imaging plane. Figure 5 shows the deviations between the perceived depth positions indicated by pointing and the actual imaging plane. The horizontal axis represents the actual imaging plane (unit: cm), and the vertical axis represents the pointing positions (unit: cm). The blue dashed line represents the ideal pointing position for each condition. From this graph, it can be seen that the median pointing positions for all conditions were located near the actual imaging plane. Furthermore, the median deviations for all conditions were within 1 cm of the actual imaging plane. These results indicate that, despite some individual variation, the overall pointing accuracy was consistent and did not exhibit significant errors.

Across participants, pointing responses tracked imaging-plane changes with high linearity. Individual R^2 values were consistently high (median $R^2 = 0.9946$; range 0.9581–0.9983; mean 0.9894 ± 0.0128 SD), indicating that the method sensitively captured depth differences *within* participants. At the group level, *Group* R^2 *(mean)* was very high ($R^2 = 0.9986$), whereas *Group* R^2 *(pooled)* was lower ($R^2 = 0.8222$) when fitting a single line to all $N = 14 \times 7 = 98$ points. Both group regressions yielded essentially the same slope and intercept for the underlying trend ($b = 0.942$, $a = -1.475$ in the units of y), consistent with the per-participant slope distribution (mean 0.942 ± 0.071 SD).

The findings indicate that the pointing method could be a promising means of estimating perceived depth in mid-air images. Additionally, it was shown that mid-air images, which lack physical substance, enable direct depth indication through pointing, which is difficult to achieve with conventional displays. Based on these results, the next experiment measures the perceived magnitude of depth of mid-air images using the pointing method.

5 Experiment 2: Evaluation of Perceived Depth of Mid-air Images Using Pointing Method

In this experiment, participants pointed to two different locations on the 2D mid-air image. The depth difference between the two pointing positions was measured to quantify the perceived magnitude of depth.

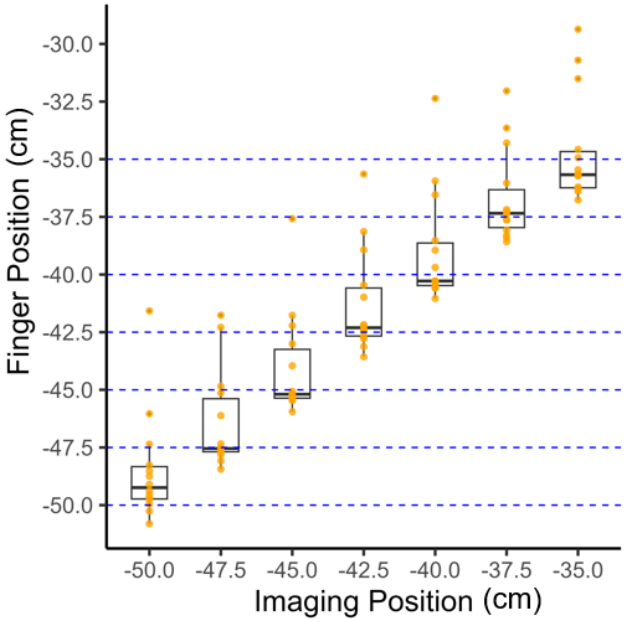


Figure 5: Deviations between the perceived depth positions of the mid-air image and the actual imaging plane. Larger values indicate positions farther away (toward the MMAP), while smaller values indicate positions closer to the participant.

5.1 Optical Design and Implementation

A 2D display and an MMAP were used to present the mid-air images. The imaging plane was positioned 47 cm in front of the center of the MMAP, and the observation distance was set to 77 cm in front of the MMAP center. The pointing positions were recorded using an optical motion capture system (OptiTrack V120: Trio).

As shown in Figure 6, the experimental setup consisted of a 2D display, an MMAP, and an optical motion capture system. The same high-resolution mobile monitor (Intehill) used in Experiment 1 was employed as the 2D display. The MMAP was an ASKA3D plate (488 mm × 488 mm) manufactured by ASKANET, and the motion capture system was provided by OptiTrack. The mid-air image content was created using Unity software (version 2022.3.19.f1).

5.2 Method

In this experiment, the perceived depth of mid-air images was evaluated using the pointing method under various conditions involving three types of pictorial depth cues (shading, shadow, and size difference), as shown in Figure 7. These conditions were selected based on a public demonstration conducted at a university, where questionnaire results indicated that these factors were frequently associated with enhanced depth perception.

In the shading condition, a 10 cm sphere was shown either uniformly bright or with shading on the lower part, illuminated by a single directional light positioned 60° above to emphasize surface shading. In the shadow condition, two overlapping colored planes were used to evaluate the difference in perceived depth depending on whether the front object cast a shadow on the rear object. The

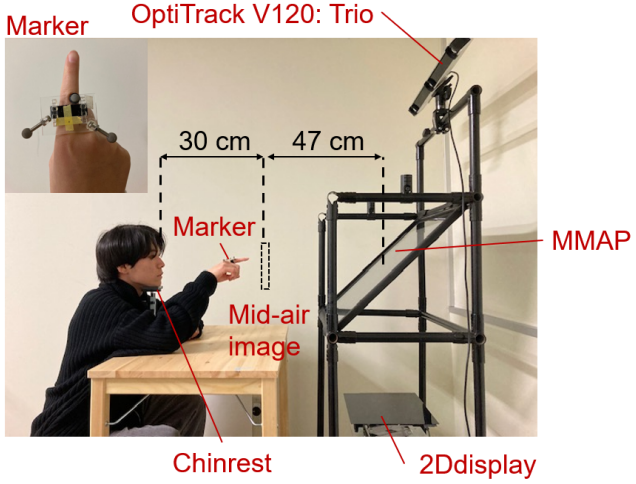


Figure 6: Experimental setup (hardware)

planes were separated by 10 cm along the depth axis to make the separation visible while keeping it within finger reach. In the no shadow condition, no shadow was projected onto the rear plane. In the with shadow condition, a shadow from the front object was projected onto the rear plane. The shadow was generated in Unity by enabling Cast Shadows for the red plane, with a white directional light placed 12° above and 17° horizontally, producing a natural soft shadow on the blue plane. In the size difference condition, multiple circles were arranged based on the texture gradient principle to manipulate depth perception through size variation. In the no size difference condition, all circles were displayed at the same size. In the with size difference condition, the circles gradually increased in size from the top to the bottom of the display, creating the visual effect that larger circles appear closer and smaller circles appear farther away, with frontal illumination provided by a single directional light.

These images were displayed as 2D mid-air images in front of the participants, who were instructed to point sequentially at two designated locations in the depth direction for each condition:

- Sphere: The front-most protruding part (front surface) and the lowest protruding part (bottom surface)
- Two planes: The center of the front red plane and the center of the rear blue plane
- Multiple circles: The red circle at the front-right position and the red circle at the back-left position

The mid-air images were displayed continuously, and participants' heads were stabilized using a chin rest. The experiment was conducted in a dark room.

5.3 Procedure

For each condition, participants were instructed to point at the two designated locations, and their finger positions were recorded for one second. A total of six mid-air image conditions were presented in a randomized order based on a Latin square design, with five trials conducted for each condition, resulting in 30 trials in total. The experiment took approximately 30 minutes per participant.

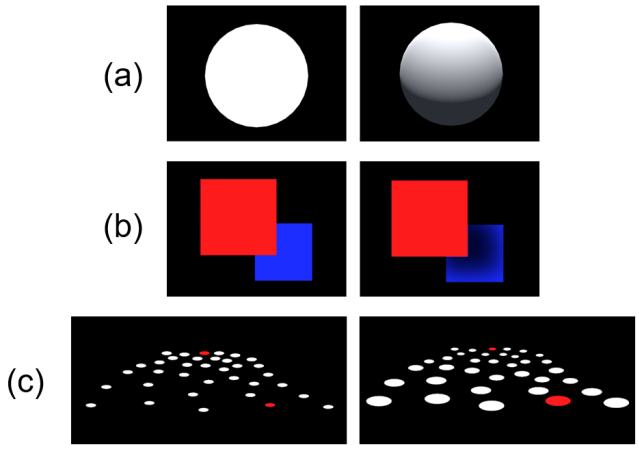


Figure 7: Mid-air images used in the experiment: (a) shading condition, (b) shadow condition, (c) size difference condition

The participants consisted of 14 individuals (12 males and 2 females) aged between 21 and 24, who also took part in the pointing accuracy experiment.

5.4 Results

Figure 8 shows the perceived depth results for mid-air images under conditions with and without each pictorial depth cue. The vertical axis represents the magnitude of perceived depth (unit: cm), calculated as the difference in the average pointing positions along the depth direction for the two designated locations. The Wilcoxon signed-rank test was used to assess the statistical significance of differences between conditions.

As shown in Figure 8 (a), a significant difference was observed depending on the presence or absence of shading ($p < 0.01$; means: without = 1.24 cm, with = 3.02 cm, Cohen's $d_z = 0.91$). Many participants commented that "the sphere without shading appeared flat, whereas the shaded sphere exhibited a clear sense of depth."

Shading significantly enhanced perceived depth, reflected in pointing. A 5 cm radius sphere yielded a median perceived depth of 2.4 cm, about half of the expected, suggesting participants recognized it as 3D and that sensing areas should be volumetric.

Figure 8 (b) shows that the presence or absence of shadow also had a significant effect on perceived depth ($p < 0.01$; mean: without = 0.48 cm, with = 1.09 cm, Cohen's $d_z = 1.13$), although the overall depth magnitude was smaller compared to the shading condition. Some participants stated that "without shadow, the two planes appeared on the same plane, whereas with shadow, some depth was perceived." However, there were also comments such as "even with shadow, I could hardly perceive any depth." One possible reason for this is the small size difference between the red and blue planes, which may not have provided sufficient depth cues. Increasing the size difference, for example by making the blue plane smaller, could enhance perceived depth.

Figure 8 (c) demonstrates that size difference also had a significant effect on perceived depth ($p < 0.01$; mean: same = 1.30 cm, different = 3.46 cm, Cohen's $d_z = 0.80$). In the size difference condition, although the two red circles were actually placed on the same

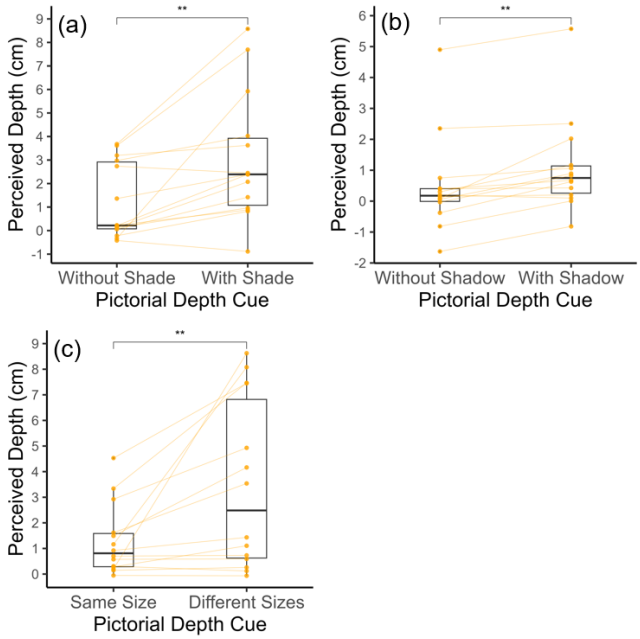


Figure 8: Perceived depth results: (a) presence or absence of shading, (b) presence or absence of shadow, (c) presence or absence of size difference

plane, a median depth of 2.5 cm was perceived. Many participants stated that "the size difference clearly enhanced the sense of depth," while a few commented that "the depth effect was weak." This reduction in perceived depth can be attributed to the flat appearance of the circles. It is expected that adding thickness, such as rendering the circles as cylinders, would further enhance depth perception. Additionally, because the circles were small, they were easily obscured by the participants' fingers when pointing, which reduced the discomfort caused by the inability to focus on the imaging plane compared to the shadow condition. This suggests that appropriately reducing the size of mid-air images to a degree where they can be occluded by fingers may help alleviate the discomfort caused by focusing mismatches. Future work should further investigate this issue, including the use of focal adjustment mechanisms.

In summary, these results indicate that pictorial depth cues such as shading, shadow, and size difference significantly influence the perceived depth of 2D mid-air images. Furthermore, this experiment demonstrated that combining 2D mid-air images with the pointing method enables the quantitative evaluation of pictorial depth cues. In the next experiment (Experiment 3), we investigate how applying these pictorial depth cues to user interfaces (UIs) in mid-air images affects depth perception.

6 Experiment 3: Mid-Air Image Buttons

In this experiment, we focused on mid-air image buttons as an example of an interface that involves user interactions along the depth direction.

Since conventional 2D mid-air image buttons are displayed only on a plane without physical thickness or depth-related movement,

it is difficult to provide a natural sense of depth during button pressing. In contrast, physical buttons produce a clear sense of pressing because the button physically sinks in response to the user's action, resulting in a tangible change along the depth direction. Although methods such as physically moving the display to change the focal position of mid-air images have been proposed to replicate this sensation in mid-air interfaces [10], such approaches increase device complexity and introduce response delays.

Therefore, in this experiment, we explored a method to visually present a depth sensation without physically moving the focal plane. We displayed a shadow behind the button and reduced the size of both during pressing to evoke an inward push.

6.1 Optical Design and Implementation

The optical design and system configuration were the same as in Experiment 2. The focal plane of the mid-air image was set 47 cm in front of the center of the MMAP, and the viewing distance was set to 77 cm from the MMAP center. The experiment was conducted in a dark room, with participants' heads stabilized using a chin rest to ensure observation of the mid-air image from the front. The visual content for the mid-air image buttons was created using Unity (version 2022.3.38.f1).

6.2 Method

As shown in Figure 9, four types of mid-air image buttons were presented in random order. In each condition, identical buttons were displayed on the left and right sides, and the conditions never differed between the two sides. The four button conditions were as follows:

- Without shadow & Same size (Figure 9 (a))
- With shadow & Same size (Figure 9 (b))
- Without shadow & Different sizes (Figure 9 (c))
- With shadow & Different sizes (Figure 9 (d))

The buttons were illuminated by a single directional light, and the shadows, placed at the lower right, were opaque gray with blurred edges, adopted based on a pilot study showing they are easier to perceive as shadows. For each condition, participants performed a button-pressing task, and the amount of penetration was measured—defined as the distance the fingertip extended beyond the focal plane of the mid-air image. This penetration depth served as an objective index, indicating how pictorial depth cues affected the actual extent of finger movement relative to the focal plane. Participants were also asked to indicate the perceived depth of the button press using a pointing gesture, and to rank the subjective sense of pressing depth for each condition. This perceived depth of the button press (hereafter termed perceived pressing depth) served as a subjective index, reflecting how deeply the button was felt to sink based on memory of the pressing sensation without visual stimuli. Thus, the two measures captured complementary aspects of depth: the objective motor response versus the subjective perceptual impression.

6.3 Procedure

The experiment consisted of three repetitions of a set containing the four button conditions, including both the button-pressing tasks and the evaluation of perceived pressing depth. First, a practice

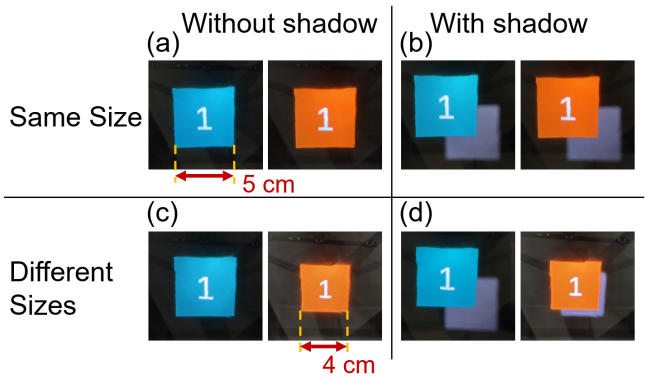


Figure 9: Mid-air image buttons displayed in the experiment: (a) Without shadow & Same size, (b) With shadow & Same size, (c) Without shadow & Different sizes, (d) With shadow & Different sizes. In each image, the left (light blue) shows the button before pressing, and the right (orange) shows the button after pressing.

phase was conducted to familiarize participants with operating the mid-air image buttons. During the main experiment, participants were instructed to press the button with a number displayed on it using the index finger of their dominant hand. The numbered button was randomly presented on either the left or right side. The depth position of the fingertip was recorded at the moment of pressing, and the penetration amount relative to the focal plane was calculated. To provide clear feedback that the button was successfully pressed, the button's color changed when the fingertip reached the button. Each condition included 10 press trials, repeated three times, resulting in a total of 120 trials per participant. The order of the conditions was randomized based on a Latin square design to minimize order effects. After completing the pressing tasks for all four conditions, participants indicated the perceived depth of pressing for each button using a pointing gesture. In this reporting, no stimuli were presented, and participants pointed based on their memory of the pressing sensation. Finally, participants ranked the conditions in order of the perceived magnitude of the pressing sensation. The total time required per participant was approximately 20 minutes.

A total of 16 participants (12 males and 4 females) aged between 21 and 24 years took part in the experiment. The participants were newly recruited for this study. Of these, 6 participants were involved in research related to mid-air images, while the remaining 10 had no such experience. All participants had normal or corrected-to-normal vision, sufficient for daily activities.

6.4 Results

First, Figure 10 shows the penetration depth of participants' fingers for each condition and the results of a two-way repeated-measures ANOVA with "presence/absence of shadow" and "presence/absence of size difference" as within-subject factors. The vertical axis represents the penetration depth of the finger from the focal plane

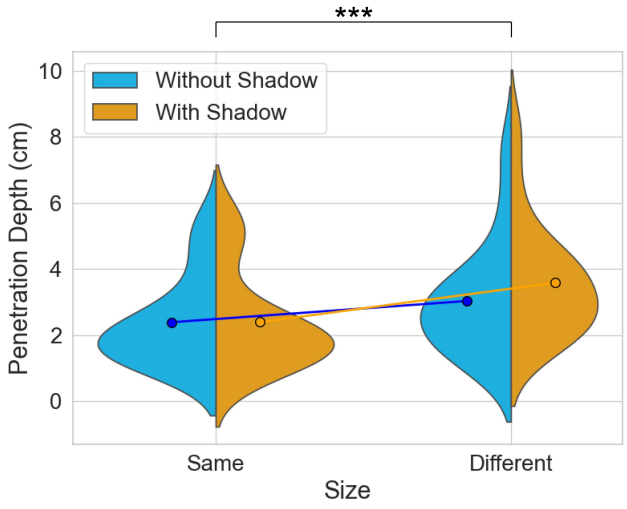


Figure 10: Penetration depth of participants' fingers and interaction effects for each condition

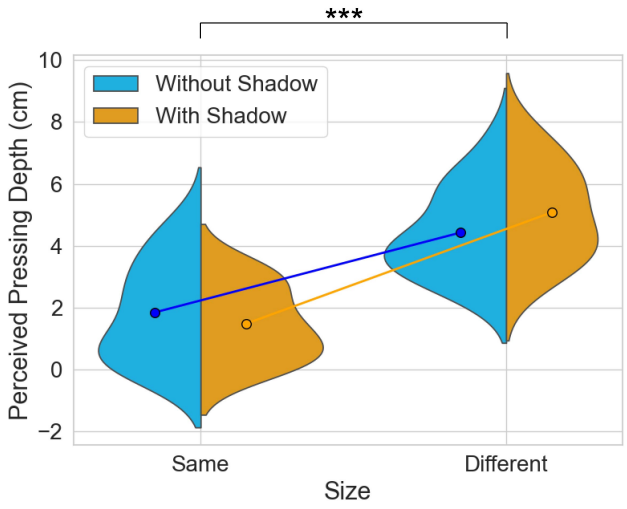


Figure 11: Perceived pressing depth and interaction effects for each condition

(unit: cm). Since the Shapiro-Wilk test showed that the penetration depth data were not normally distributed, an aligned rank transform (ART) procedure [23] was applied prior to the ANOVA.

Next, Figure 11 shows the perceived pressing depth for each condition and the results of a two-way repeated-measures ANOVA based on the within-subject factors of "presence/absence of shadow" and "presence/absence of size difference." The vertical axis indicates the perceived depth of button pressing, in terms of "how deeply the button felt to be pressed." When significant interactions were found in penetration depth and perceived pressing depth, simple main effects were tested for each.

The results revealed that conditions involving size differences—particularly when combined with shadows—significantly increased both the penetration depth and the perceived pressing depth. As shown in Figure 10, the mean penetration depth increased by approximately 1 cm under the different sizes condition compared to the same size condition. The two-way ANOVA showed a significant main effect of size difference ($F(1, 15) = 48.43$, $p < 0.001$) and a significant interaction between shadow and size difference ($F(1, 15) = 5.18$, $p < 0.05$). On the other hand, no significant main effect of shadow was observed ($F(1, 15) = 3.66$, $p > 0.05$). Similarly, Figure 11 shows that the perceived pressing depth increased by approximately 3–4 cm under the different sizes conditions compared to the same size conditions. The ANOVA revealed a significant main effect of size difference ($F(1, 15) = 150.03$, $p < 0.001$) and a significant interaction between shadow and size difference ($F(1, 15) = 4.08$, $p < 0.05$), while the main effect of shadow was not significant ($F(1, 15) = 0.34$, $p > 0.05$). To clarify these interactions, simple main effects were tested with paired t-tests and Wilcoxon tests. Results showed that size difference had a significant effect on both penetration depth and perceived pressing depth, and the effect was amplified by the presence of shadows. These findings suggest that size difference effects depend on shadow, whose changes in size and position enhanced depth perception and pressing sensation. In this experiment, in the condition where both shadow and different sizes were present, not only the size but also the position of the shadow changed to appear closer to the button. This was intended to mimic the real-world phenomenon where both the size and position of shadows change when an object moves along the depth axis under lighting conditions. Thus, the combination of size and positional changes of the shadow further emphasized the depth perception and pressing sensation induced by the size difference.

Subjective evaluations revealed that perceived pressing depth increased notably under the different sizes condition, especially when combined with shadows. As shown in Figure 12, approximately 90% of participants reported the strongest pressing sensation under the condition with both shadow and size difference, followed by the condition with size difference alone. In contrast, under same size conditions, the perceived depth remained low regardless of shadow presence. These findings suggest that size-induced visual change is a key factor in enhancing pressing sensation. Several participants also noted that shadows increased the button's three-dimensionality, making it feel "more dynamic and satisfying." Notably, when both cues were present, the simultaneous motion of the button and its shadow amplified visual change, further strengthening the pressing experience.

In summary, these results demonstrate that even for intangible mid-air image buttons, it is possible to effectively generate the sensation of pressing by providing visual cues. In particular, the combination of visual size difference and shadow significantly increased both the penetration depth and perceived pressing sensation, enhancing the user's sense of interaction.

7 Discussion

Through three experiments on interactions using mid-air images, this study demonstrated methods for evaluating depth perception

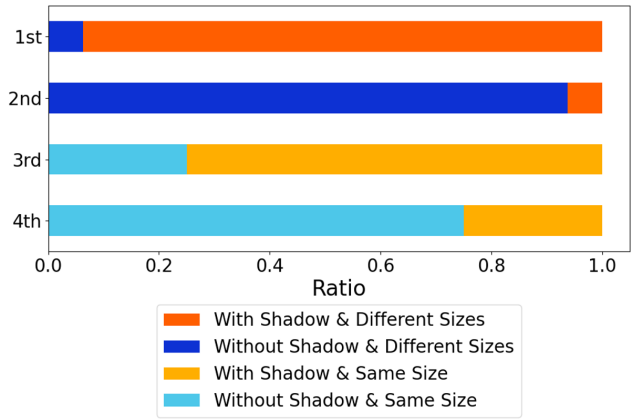


Figure 12: Subjective Ranking Results of Perceived Pressing Sensation

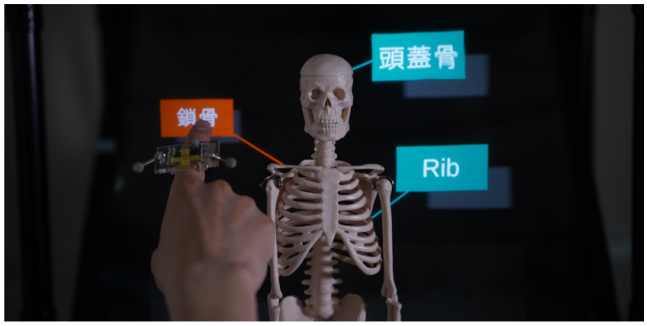


Figure 13: An example of interaction where real objects and mid-air images coexist.

and explored possibilities for the design of mid-air image presentations. As an application, a museum display could use pictorial cues to emphasize spatial relations and interactivity as shown in Figure 13. In this paper, our contributions are as follows:

- Experiment 1 demonstrated that the pointing method used toward mid-air images enables responses to depth variations with an accuracy having only a 5.5 % error.
- Experiment 2 enabled quantitative evaluation of pictorial cues using 2D mid-air images and pointing (e.g., a shaded 5-cm sphere yielded 2.4-cm perceived depth).
- Experiment 3 demonstrated that, even in the case of mid-air image buttons without physical substance, the sensation of pressing can be induced by manipulating the size of the button and its shadow. By employing both penetration depth (objective) and perceived pressing depth (subjective), we showed that pictorial depth cues affected not only motor control but also users' pressing experience.

While the foregoing contributions and avenues for application are promising, several limitations qualify our findings. Below we catalog the limitations identified in each experiment.

In Experiment 1, pointing positions changed with depth variations, but discrepancies with the actual mid-air image position remained. However, this discrepancy can be corrected by placing a physical object at the same depth position as the mid-air image, which may not pose a significant problem in practical applications.

Next, in Experiment 2, compared to physical spheres, mid-air images could not reproduce the full range of depth perception and tended to be perceived as having less depth than physical objects. The influence of highlights, shading, and shadows on depth perception was not fully examined, leaving room for further study. Additionally, in the size difference condition, both size and perspective cues were used, based on preliminary tests indicating their effectiveness. Therefore, this study did not fully isolate pictorial depth cues, and future work should examine them separately. Depth cues were tested only in binary form; future work should vary parameters such as shading intensity, shadow blur, and size gradients for finer analysis.

The evaluation in Experiment 3 was limited to specific UI designs, and it remains unclear whether similar effects generalize to other designs. Specifically, when only shadows were presented, the pressing sensation was not confirmed, which suggests that the position and color of shadows may not have been effective in conveying the pressing sensation. Thus, using designs where these elements are altered may result in different pressing sensations.

In the future, it will be necessary to clarify which UI designs can more effectively provide a sense of depth during button pressing, by examining factors such as optimal size variations, shadow position and color, and other pictorial cues like button shapes and perspective. It will also be important to explore the effects of visual parameters (e.g., shading intensity, shadow blur, size gradients) in a more systematic manner, rather than using only binary conditions. In addition, incorporating haptic feedback for mid-air images or auditory feedback when buttons are pressed could provide a more realistic interaction experience. Furthermore, participants in this study were mostly young men; future research should involve broader age and gender groups, including presbyopic adults, to better understand differences in mid-air image perception.

8 Conclusion

We examined whether pictorial cues enhance depth in mid-air images and support pressing sensation in UI design. Across three experiments, we validated the pointing method, quantified depth cue effects, and applied them to button UIs. The experimental results showed that pictorial depth cues, such as differences in size, shadows, and shading, enhance the sense of depth in 2D mid-air images. Combining size differences with shadows effectively enhanced pressing sensation, showing potential for mid-air button UIs. These results indicate that pictorial depth cues can effectively enhance interaction in 2D mid-air UIs, even in the absence of physical substance or focal depth changes.

Acknowledgments

This study was supported by JST FOREST Program, Grant Number JPMJFR216L.

References

[1] Haley Adams, Jeanine Stefanucci, Sarah Creem-Regehr, and Bobby Bodenheimer. 2022. Depth Perception in Augmented Reality: The Effects of Display, Shadow,

and Position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 792–801. doi:10.1109/VR51125.2022.00101

[2] Shohei Ando and Naoya Koizumi. 2024. An Optical Design for Interaction With Mid-Air Images Using the Shape of Real Objects. *IEEE Access* 12 (2024), 39129–39138. doi:10.1109/ACCESS.2024.3374782

[3] Jeffrey Andre and Sheena Rogers. 2006. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & psychophysics* 68, 3 (April 2006), 353—361. doi:10.3758/bf03193682

[4] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300437

[5] James E. Cutting and Peter M. Vishton. 1995. Chapter 3 - Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth*. In *Perception of Space and Motion*, William Epstein and Sheena Rogers (Eds.). Academic Press, San Diego, 69–117. doi:10.1016/B978-012240530-3/50005-5

[6] Catherine Diaz, Michael Walker, Danielle Albers Szafir, and Daniel Szafir. 2017. Designing for Depth Perceptions in Augmented Reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 111–122. doi:10.1109/ISMAR.2017.28

[7] S Kay Fisher and Kenneth J Ciuffreda. 1988. Accommodation and Apparent Distance. *Perception* 17, 5 (1988), 609–621. doi:10.1068/p170609

[8] Tomoyo Kikuchi, Yuchi Yahagi, Shogo Fukushima, Saki Sakaguchi, and Takeshi Naemura. 2023. AIR-range: Designing optical systems to present a tall mid-AIR image with continuous luminance on and above a tabletop. *ITE Transactions on Media Technology and Applications* 11, 2 (2023), 75–87. doi:10.3169/mta.11.75

[9] Beomjin Kim and Thomas Bolinger. 2024. Analyzing Depth Perception in Virtual Environments: A Comparative Study of Varied Scaling Factors. In *Proceedings of the 2023 5th International Conference on Video, Signal and Image Processing* (Harbin, China) *(VSIP '23)*. Association for Computing Machinery, New York, NY, USA, 200–205. doi:10.1145/3638682.3638712

[10] Hanyuool Kim, Issei Takahashi, Hiroki Yamamoto, Satoshi Maekawa, and Takeshi Naemura. 2014. MARIO: Mid-air Augmented Reality Interaction with Objects. *Entertainment Computing* 5, 4 (2014), 233–241. doi:10.1016/j.entcom.2014.10.008

[11] Takafumi Koike and Yasushi Onishi. 2018. Aerial 3D Imaging by Retroreflective Mirror Array. In *Proceedings of the 2018 ACM Companion International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) *(ISS '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 25–29. doi:10.1145/3280295.3281365

[12] Makiko Kouchi and Masaaki Mochimaru. 2006. Japanese 3-D body shape and dimensions data 2003, National Institute of Advanced Industrial Science and Technology, H18PRO-503. https://www.airc.aist.go.jp/dhrt/fbodydb/index.html. Accessed: 2025-01-18.

[13] Chiuhsiang J. Lin and Bereket H. Woldegiorgis. 2017. Egocentric distance perception and performance of direct pointing in stereoscopic displays. *Applied Ergonomics* 64 (2017), 66–74. doi:10.1016/j.apergo.2017.05.007

[14] Satoshi Maekawa, Kouichi Nitta, and Osamu Matoba. 2006. Transmissive optical imaging device with micromirror array. In *Three-Dimensional TV, Video, and Display V*, Bahram Javidi, Fumio Okano, and Jung-Young Son (Eds.), Vol. 6392. International Society for Optics and Photonics, SPIE, 63920E. doi:10.1117/12.690574

[15] Motohiro Makiguchi, Ayaka Sano, Takahiro Matsumoto, Hiroshi Chigira, and Takayoshi Mochiduki. 2023. Implementation of Interactive Mirror-Transcending Aerial Imaging System. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) *(SUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 12, 11 pages. doi:10.1145/3607822.3614536

[16] Robert D. McIntosh, Matthew H. Iveson, Sebastian Sandoval Similä, and Antimo Buonocore. 2023. Pictorial depth cues always influence reaching distance. *Neuropsychologia* 190 (2023), 108701. doi:10.1016/j.neuropsychologia.2023.108701

[17] Rina Nagano, Takahiro Kinoshita, Shingo Hattori, Yuichi Hiroi, Yuta Itoh, and Takefumi Hiraki. 2024. HaptoFloater: Visuo-Haptic Augmented Reality by Embedding Imperceptible Color Vibration Signals for Tactile Display Control in a Mid-Air Image. *IEEE Transactions on Visualization and Computer Graphics* 30, 11 (2024), 7463–7472. doi:10.1109/TVCG.2024.3456175

[18] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. doi:10.1145/237091.237102

[19] Mengdi Shi, Tao Hu, and Jiawen Yu. 2022. Pointing Cursor Interaction in Virtual Reality from the Perspective of Distance Perception. *Traitement du Signal* (2022), 475–483. doi:10.18280/ts.390209

[20] Mayumi Takazaki and Shinji Mizuno. 2020. A Method for Appropriate Occlusion between a Mid-air 3DCG Object and a Hand by Projecting an Image on the Hand. In *ACM SIGGRAPH 2020 Emerging Technologies* (Virtual Event, USA) *(SIGGRAPH '20)*. Association for Computing Machinery, New York, NY, USA, Article 14, 2 pages. doi:10.1145/3388534.3407285

[21] Cyril Vienne, Stéphane Masfrand, Christophe Bourdin, and Jean-Louis Vercher. 2020. Depth Perception in Virtual Reality Systems: Effect of Screen Distance, Environment Richness and Display Factors. *IEEE Access* 8 (2020), 29099–29110. doi:10.1109/ACCESS.2020.2972122

[22] Franziska Westermeier, Larissa Brübach, Carolin Wienrich, and Marc Erich Latoschik. 2024. Assessing Depth Perception in VR and Video See-Through AR: A Comparison on Distance Judgment, Performance, and Preference. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2140–2150. doi:10.1109/TVCG.2024.3372061

[23] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[24] Hirotsugu Yamamoto, Yuka Tomiyama, and Shiro Suyama. 2014. Floating aerial LED signage based on aerial imaging by retro-reflection (AIRR). *Optics Express* 22, 22 (November 2014), 26919–26924. doi:10.1364/OE.22.026919

[25] Yutaro Yano and Naoya Koizumi. 2023. Mid-air Image's Background Changes the Impression of a Mid-air Image. In *ICAT-EGVE 2023 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Jean-Marie Normand, Maki Sugimoto, and Veronica Sundstedt (Eds.). The Eurographics Association. doi:10.2312/egve.20231309