1. Introduction

Video communication tools, such as video chat and video conferencing, have become common means of communication among users in remote locations, as display/capture devices, such as flat panel displays, projectors, and digital cameras, have evolved, and communication networks have become much faster. In addition, tools, such as PCs and smartphones, have become widely spread around the globe. However, although it is now very easy to use video communication, it has not completely replaced actual face-to-face (F2F) communication. It has been shown that F2F communication leads to longer conversations compared with online video communication \[1\]. Also, people tend to prefer F2F communication in important scenarios, such as business negotiations. In other words, it is widely considered that the "quality" of online video communication is lower than that of F2F communication.

In this study, we considered two major factors that affect the quality of video communications. The first factor is the decreased "sense of presence" of the other party at a remote location. In F2F, you can perceive that the person you are talking with is actually in front of you, and you can also feel their presence with your eyes. However, generally, video communication systems have several physiological issues that detract the presence of their users, e.g., low image resolution, improper life-size scaling of displayed users, and the stereoscopic effect absence in the display \[2\]. Therefore, in video communication, users do not properly feel the presence of the other party compared with F2F communication.

The second factor is that it is difficult to properly convey non-verbal information in video communication, including facial expressions, eye gaze information, gestures, and postures. For example, a coinciding eye gaze between two communicating parties makes eye contact possible, which contributes to building trust and having new impressions during conversations and when performing tasks between remote locations \[3\].

A typical video communication system configuration is shown in Fig. 1 (a). As shown in the figure, the camera is placed around the 2D display (typically at the top), and...
The user speaks while looking at the display to see the other party's video image. Therefore, a gap is generated between the directions of the user's eye gaze and the camera's optical axis, resulting in captured video images as downcast looks. In video communication, since video images are transmitted and received on both sides, communication is performed with both parties looking down, and eye contact is impossible. Thus, to realize eye contact, it is necessary to set up a camera on the extended line of the eye gaze direction of the user, that is, within the screen, as described by the dashed line in Fig. 1 (a), so as to capture the frontal image of the user. However, the display itself is an obstacle to taking videos. Also, cameras cannot be installed within or behind typical displays, such as liquid crystal display and organic light-emitting diode display.

There are two existing approaches for obtaining frontal images to achieve eye contact: a software (SW) approach [4-6], which synthesizes virtual frontal images through image processing, and a hardware (HW) approach [2, 7-9], which obtains optically correct frontal images using fast response projectors and liquid crystal screens or half mirrors. Compared with the SW approach, the HW approach has the advantage that the gap between the user's eye gaze direction and the camera's optical axis direction is small, even in case the display area is large, so natural frontal images can be obtained. However, this approach requires a large depth space in proportion to the size of the displayed video image, which is a big issue in system implementation. Also, a typical HW approach using a half mirror requires a minimum depth of $H_{\text{screen}} + D_{\text{camera}}$, where $H_{\text{screen}}$ and $D_{\text{camera}}$ denote the screen height and camera depth, respectively, as shown in Fig. 1 (b).

Nevertheless, this configuration causes displayed video images to be virtual images. Since the life-size virtual image display of a human has an effect of enhancing the sense of presence [10], this approach can be a solution to the first issue. Also, people prefer to have conversations in close proximity, especially when communicating with people close to them, such as family members and friends [11]. However, the system depth when using a tilted mirror hinders conversations with the opposite party, who in this case is projected as a life-size virtual image in a closer distance.

In view of the above issues, this study aims at establishing a video communication system that can realize virtual image projection and frontal image capturing and also reduce the system depth in front of users. We used a holographic optical element (HOE) [12], a type of diffraction grating with wavelength selectivity. HOE can be recorded by the interference of coherent lights, has high flexibility in optical design, and is a thin optical element. In earlier studies, the authors proposed dispersion compensation optical systems using diffractive optical elements (DOE) in single color [13] and in full color [14]. However, these studies were verified in either projection or capturing and had a system complexity problem in full-color systems. In this paper, we proposed a new thinner system configuration design that can simultaneously realize virtual image projection and frontal image capture in full color. The system uses three full-color HOEs as an off-axis mirror. One HOE, which replaces the tilted half mirror in conventional systems, is set up "upright" in front of the user to make the system thinner. The high flexibility in the optical design of HOEs makes the installation of equipment, such as displays and cameras, more flexible. This allows the equipment to be placed out of the user's field of view (below or at the lateral side of the user) and the entire system to be unobtrusively configured. The other HOEs are used for dispersion compensation to obtain blur-less full-color images. We also established a proof-of-concept system to simultaneously realize virtual image projection and frontal image capture. In addition, to make the system thinner, the upright setup makes the HOE attached to vertical surfaces, such as walls. Besides, HOEs with high transparency are visually unobtrusive, even when they are on walls. Due to these characteristics, the system using upright HOE can blend in with the environment when not in use. As a result, it can become a novel human interface device for video communications.

2. Virtual image projection and capture using HOE mirrors

During video communication, the video image of the
opposite party should be displayed upright in front of the user. However, in conventional setups, as shown in Fig. 1(b), the displayed image on the screen is formed as a virtual image at a symmetrical position with respect to the mirror plane by specular reflection. Therefore, the half mirror must be tilted to display the virtual image upright, so more depth is needed for the system installation. In our proposal, we used a HOE as an "off-axis" reflection optic system. Off-axis reflection means that the incidence and reflection angles are different. Also, an off-axis reflection can form a virtual image at the non-plane-symmetrical position of the screen with respect to the HOE plane, allowing the virtual image to be projected upright, while the HOE is also installed upright. In this paper, the used HOE is a type of reflective grating recorded in a volumetric material, and it can be referred to as the "HOE mirror." We used photopolymer exhibiting characteristics, such as high transmission and high diffraction efficiency, for the recording material of the HOE.

The HOE exposure is performed with signal and reference beams at an incident angle of 135 degrees. A near-eye display, such as a smart glass, focuses on the virtual image projection, where the HOE size is small [15]. Our video communication system not only aims at image projection but also at capturing the images of users. Therefore, a certain distance is needed between the HOE and the user, and a larger HOE size is needed compared with that for near-eye display. To expose a large HOE, the holographic printing method [16] was used, with which small holograms called cell holograms were sequentially exposed to enlarge the size.

The virtual image projection and frontal image capturing concept, in which a HOE mirror is used, is demonstrated in Fig. 2. The light entering the HOE mirror from below the user is diffracted toward the user's direction, as shown in Fig. 2 (a). Here when a display is set up below the user, the video image in the display is formed in an upright state behind the HOE. Similarly, the light entering the upright HOE from the user's side is diffracted toward below the user, as shown in Fig. 2 (b). As a camera is installed below the user, the user's frontal image diffracted by the HOE mirror can be taken. This is as if a virtual camera is set up in the front behind the HOE, which makes the depth space unnecessary for the camera. With this "reflective capturing" method, only one sheet of HOE needs to be set up in front of the user. Therefore, a very thin system configuration becomes possible. The thickness of the HOE is about 20 µm, and the total thickness including the used glass to laminate the HOE is less than 10 mm. The configuration needs enough space between the HOE and the user not to obstacle the optical path of the projection and capturing. However, replacing the specular reflection mirror with the off-axis reflection mirror makes it possible to make the system thinner.

Here, the HOE is a reflective diffraction grating that only reacts to the wavelengths around the wavelength at the time of exposure, and the used full width at half maximum (FWHM) centered on the peak wavelength of the holographic recording material is approximately 10 nm. At the peripheral wavelengths, even when light enters at the same incident angle, a slight difference is generated in the diffraction direction based on the grating equation. This is shown in Fig. 3 (a) as an example of virtual image projection. When the diffused from a video image plane such as a display or a screen on which a video image is projected by a projector enters the HOE, the light is reflected in a slightly different diffraction direction depending on the wavelength. This leads to a shift in the position where the image is formed as a virtual image, and the user perceives it as blurring of the virtual image eventually. As a solution, it is conceivable to use a light source having a sufficiently narrower bandwidth than that of the FWHM to which the HOE reacts on the video image plane, e.g. projection
using a laser projector. However, this requires a projector with a light source that uses a laser with almost the same wavelength of the exposure beam. Since such projectors are not general-purpose products, this approach would be costly.

To solve this issue, the authors proposed a blur compensation optical system that uses a general-purpose projector with a broad wavelength as a light source \[13\]. The compensation system projects wavelength-dispersed video images on diffusers (screens) via DOE. Further, it inversely disperses the video image in the HOE to finally cancel the perceived image blurring by the user. This optical system is shown in Fig. 3 (b). Note that a reflective holographic grating (1200 GPM VIS 50 mm by Edmund Optics) was used for the DOE.

Wavelength dispersion also occurs during video image capturing as well as in the case of projection. That is, as shown in Fig. 4 (a), when the HOE is captured as it is by the camera, a blurring occurs in the frontal image of the user. Although it is technically possible to use a narrow bandwidth laser as a light source for projection, as described above, it is difficult to assume such a narrow bandwidth light source for ambient light during video image capturing. Therefore, it is also necessary to provide a blur compensation optical system for image capturing, as shown in Fig. 4 (b). Due to the HOE diffraction, the virtual image of the user is formed on the upper part of the back of the HOE, but wavelength dispersion occurs. In this case, since the distance between the DOE for blur compensation and the virtual image is large, it is necessary to set up a very large DOE for blur compensation. Here, by setting up a convex lens between the DOE and the virtual image and by converting the divergent light into convergent light, once, a small-size real image is formed between the convex lens and the DOE. Of course, this real image is blurry; however, through the DOE diffraction, the frontal image of the user that is compensated for blurring is taken by the camera.

3. Full-color HOE mirror and dispersion compensation

In general, since video communications are usually required to be in full-color, full-color HOE should be considered. However, to fabricate a full-color HOE, it is necessary to use an RGB laser to emit an object beam and a reference beam at the same incident angle in each channel to perform exposure. For the exposure, by considering the enhanced diffraction efficiency, the method of laminating three exposed holographic recording materials for each single channel was used instead of the method of multiplexing RGB channels on a single holographic recording material \[17\]. In our experimental results, a photopolymer (HX200 by Covestro) was used as a holographic recording material, and the diffraction efficiency for R, G, and B was 36.6 %, 57.8 %, and 63.0 %, respectively.

Here, it is also necessary to perform blur compensation for the system using full-color HOE. The blur compensation in Fig. 3 (b) and Fig. 4 (b) can be adopted to one of the three RGB channels (monochrome). However, the diffraction angle of the DOE set up for blur compensation varies depending on the wavelength. Therefore, as shown in Fig. 5 (a), the light spatially diffracted by the DOE separates the RGB on the screen surface. Thus, when diffraction is performed again at the HOE, while blur compensation is performed in each channel, the virtual image is spatially separated between the color channels. To solve this issue, the authors proposed a full-color dispersion compensation

![Fig. 4. Dispersion compensation of the captured image ((a) is a captured dispersed image, and (b) is a non-dispersed image captured through compensation optics.)](image)

![Fig. 5. Dispersion compensation for full-color virtual image projection ((a) is the virtual image with color separation that occurs when using a DOE for compensation, and (b) is the virtual image without color separation and dispersion when using a HOE.)](image)
method not occurring the spatial separation between channels by setting up additional DOEs and projectors [14]. However, this method requires three sets of DOEs and projectors, which increases the system complexity. Therefore, we used another HOE for blur compensation. A full-color HOE that is laminated of three single-color HOEs enables the same diffraction characteristic between the color channels, preventing the diffracted RGB images from being spatially separated on the screen surface. The HOE for the compensation exhibits the same diffraction characteristics (the diffraction angle and peak wavelength of each channel) as those of the HOE mirror set up in front of the user. With the configuration in Fig. 5 (b), blur compensation can be realized in full color with a configuration equivalent to the blur compensation system in a single channel, as shown in Fig. 3 (b). Also, note that for the blur compensation in the projection, full-color capturing becomes possible by using the full-color HOE.

Fig. 6 shows a virtual image of the resolution chart captured by a camera set up at the user’s location. Fig. 6 (a) shows the case in which blur compensation was not performed, and it can be seen that the image is blurry in the horizontal direction. However, three DOEs and one HOE were used for blur compensation, as shown in Fig. 6 (b) and Fig. 6 (c), respectively. It can be seen that the blurring in the horizontal direction was improved in Fig. 6 (b) and Fig. 6 (c). Also, the obtained spatial resolutions by the analysis of the line profiles using the chart were 0.091 LP/mm at (a) and 0.33 LP/mm at (b) and (c). Note that Fig. 6 (c), which was subjected to blur compensation with the HOE, has a difference in color compared with Fig. 6 (a) and Fig. 6 (b). This is because the diffraction efficiency of the HOE for each channel is not the same and the colors are out of balance. In order to use HOEs in blur compensation optical systems, reproducing accurate color balances is a future challenge that must be overcome. Fig. 7 shows the virtual images projected in the setups of Fig. 6 (c). The distance between the screen and the HOE mirror was 60 cm, and we placed a real object 60 cm behind the HOE mirror. By changing the camera’s focus and position, the images that could be observed through the HOE attached above the holder showed the same changes in appearance as the real object. That is, the images were formed as virtual images at the same distance as the real object. For the captured images in full color, the HOE replaced the described DOE in Fig. 4 (b), as well as in the case of projection. Fig. 8 shows the captured images when a resolution chart was placed in front of the HOE mirror. Fig. 8 (a) shows the captured image in case the camera captured the diffracted image of the resolution chart directly without blur compensation and Fig. 8 (b) shows the captured image when blur compensation was performed. The results clearly show that the used HOE for the compensation could improve the horizontal blur in the captured images. The spatial resolutions were 0.12 LP/mm at (a) and 0.46 LP/mm at (b). Fig. 8 (b) shows a stronger blur in the vertical direction because of the spherical aberration of the used convex lens, but it can be suppressed using a compound lens.

4. Proof of the concept for full-color holographic video communication

So far, projection and capturing have been separately
described. However, in video communication systems, virtual image projection and frontal image capturing must be simultaneously realized, and it is necessary to separate the optical paths. Thus, we added a half mirror to separate the optical paths of the projection and capturing. Note that although the diffraction efficiency of HOEs varies depending on the exposure and laminating methods, the wavelength bandwidth in which HOEs react as optical elements is very narrow due to Bragg diffraction. Thus, there is a problem that the brightness of the projected/captured images is low, and it becomes noticeable in capturing, as it is difficult to control the bandwidth and power of the lighting source in actual situations. Therefore, we used a half mirror with a reflection transmission ratio of 3:1 to keep the capturing light intensity prioritized. Fig. 9 shows a proof-of-concept (PoC) system on an optical bench, where each value in the figure indicates a distance [cm] between two optical components. The solid and dashed lines describe the optical paths for projection and capturing, respectively.

In Figs. 2-5, the direction of the diffraction by the HOE mirror was vertical, and apparatuses like projector and camera were hindrances to reduce the depth. To deal with that, we rotated the HOE mirror by 90 degrees so that diffraction occurs in the horizontal direction. In the PoC system, only one HOE mirror existed in front of the user. Although space was still required to set up the apparatuses at the lateral side (left side in the figure) of the user, the PoC system realized a thin system configuration from the user's view. Note that we adopted a rear-projection screen considering the screen and half mirror positions.

Table 1 shows the list of used products for the system establishment. The projector and camera were connected to the PC, and it was possible to use the system for any video communication services available on PCs.

By using this PoC system, we observed the projected and captured images. Fig. 10 and Fig. 11 show the virtual images observed by the camera placed at the user's position when the lighting source for capturing was on. We laminated a sticker “surface of HOE” at the fixed position on the HOE surface as a reference object against the projected image. Fig. 10 shows the appearance of the virtual image observed through the HOE while changing the user's position. The positional relation between the reference sticker and projected image confirmed that the image was formed behind the

Table 1. Optical components using the PoC system

| Optical components | Product name |
|--------------------|--------------|
| Recording material | Bayfol HX200 (Covestro) |
| Lasers for exposure | R (660nm):Flamenco 100, G (532nm):Samba 100, B (473nm):Blues 50 (Cobolt) |
| Half mirror | B/S 75R/25T 254X356 MGF2 (Edmund Optics) |
| Projector | EB-1795F (EPSON) |
| Lens | SLB-100-150P (SIGMAKOKI) |
| Camera | Grasshopper3 (FLIR) |
| Camera Lens | DG Series Fixed Focal Length Lens, 50 mm (Edmund Optics) |
| Lighting source | JDR110V750W (Toshiba) |
| Screen (Diffuser) | UMU Smart Screen (NSG UMU PRODUCTS) |
We also verified the concept of a video communication system that simultaneously realizes virtual image projection and frontal image capturing with a thinner setup. These accomplishments are expected to enhance the sense of presence and the realization of eye contact in video communications. However, the system still has problems with the brightness of the images in both projection and capturing. As mentioned above, the captured image of the mannequin was not bright enough. Besides, the specular reflection image of the mannequin was more conspicuous than the projected virtual image when a strong lighting source for capturing was applied, which degrades the visibility of virtual images. Thus, it is essential to improve the contrast between specular reflection images and projected/captured images.

5. Conclusion

In this study, aiming at enhancing the quality of video communications, we proposed a system that enables virtual image projection and frontal image capturing by focusing on the sense of presence in video images and on eye contact. To realize the proposed system, an off-axis optical system using a HOE was adopted due to its ease of installation. The established verification system confirmed that it was possible to take frontal images of the user while viewing the image of the opposite party as a virtual image.

However, we cannot say that the PoC system provides sufficient image quality under general ambient light in either projection or capturing. Therefore, we still need to enhance the diffraction efficiency of HOEs. Separating the optical path for projection and capturing without using a half mirror will be another approach to deal with this problem. Besides, the apparatuses installation space at the lateral side of the user is still large. Since the HOEs used for PoC were exposed by the interference between parallel light, the projected virtual image almost had an equal magnification as that of the image on the screen. To downsize the entire system, we believe it would be effective to use a HOE exposed with diffused lights or convergent lights to increase the magnification. Based on these improvements, we would like to confirm through user evaluation whether this system contributes to enhancing the quality of communication.

Acknowledgment

The authors would like to thank Covestro Deutschland AG for providing the photopolymer...
holographic recording material.

References
1) R.S. Fish, R.E. Kraut and B.L. Chalfonte, "The VideoWindow System in Informal Communication", Proc. 1990 ACM Conf. Comput. Cooper. Work - CSCW ‘90, pp.1-11 (1990)
2) P. Harman, "Autostereoscopic Teleconferencing System", SPIE, 3957, pp.293-302 (2000)
3) L.S. Bohannon, A.M. Herbert, J.B. Pelz and E.M. Rantanen, "Eye Contact and Video-Mediated Communication: A Review", Displays, 34, 2, pp.177-185 (2013)
4) C. Kuster, J. Bazin and C. Gotsman, "Gaze Correction for Home Video Conferencing", ACM Trans. Graph., 31, 6, pp.2-7 (2012)
5) E. Wood, T. Baltrusaitis, L.P. Morency, P. Robinson and A. Bulling, "Gazedirector: Fully Articulated Eye Gaze Redirection in Video", Comput. Graph. Forum, 37, 2, pp.217-225 (2018)
6) B. Yip, "Face and Eye Rectification in Video Conference Using Affine Transform", Proc. IEEE Int. Conf. Image Process., 3, pp.3-6 (2005)
7) H. Ishii and M. Kobayashi, "ClearBoard", Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI ’92, pp.525-532 (1992)
8) M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A.V. Moere, O. Staadt, "Blue-C.", ACM Trans. Graph. Proc. Work. Virtual Environ EGVE’03, 22, 3, pp.819-827 (2003)
9) R. Vertegaal, I. Weersers, C. Sohn and C. Cheung, "Gaze-2," Proc. Conf. Hum. factors Comput. Syst. - CHI ‘03, 5, s.521 (2003)
10) M. Imoto, S. Uchida and K. Ashihaga, "A Concept of Immersive Telepresence 'Kirari!' (In Japanese)", Proc. Media Comput. Conf., pp.1287-1290 (2015)
11) H. Edward T, The Hidden Dimensions. 1966
12) N. Kim, Y.L. Piao and H.Y. Wu, "Holographic Optical Elements and Application", InterTechOpen, pp.99-131 (Mar. 2017)
13) T. Nakamura, S. Kimura, K. Takahashi, Y. Aburakawa, S. Takahashi, S. Igarashi, S. Torashima and M. Yamaguchi, "Off-Axis Virtual-Image Display and Camera by Holographic Mirror and Blur Compensation," Opt. Express, 26, 19, pp.24864-24880 (2018)
14) F. Watanabe, T. Nakamura, S. Torashima, S. Igarashi, S. Kimura, Y. Aburakawa and M. Yamaguchi, "Dispersion Compensation for Full-Color Virtual-Imaging Systems with a Holographic off-Axis Mirror", Proc. SPIE, 11306, Feb. (2020)
15) H. Mukawa, K. Akutsu, I. Matsumura, S. Nakano, Y. Yoshida, M. Kuwahara and K. Aiki, "A Full-Color Eyewear Display Using Planar Waveguides with Reflection Volume Holograms," J. Soc. Inf. Display, 17, 3, pp.185 (2009)
16) M. Yamaguchi, N. Ohyama and T. Honda, "Holographic Three-Dimensional Printer: New Method," Appl. Opt., 31, 2, pp.217-222 (1992)
17) J.A. Piao, G. Li, M.L. Piao and N. Kim, "Full Color Holographic Optical Element Fabrication for Waveguide-Type Head Mounted Display Using Photopolymer", J. Opt. Soc. Korea, 17, 3, pp.242-248 (June 2013)