IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 August 2022, accepted 4 September 2022, date of publication 6 September 2022, date of current version 14 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204874

RESEARCH ARTICLE

HiddenGazeStereo: Hiding Gaze-Contingent Disparity Remapping for 2D-Compatible Natural 3D Viewing

TAIKI FUKIAGE[1] **AND SHIN'YA NISHIDA**[1,2]
[1]NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Kanagawa 243-0198, Japan
[2]Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

Corresponding author: Taiki Fukiage (t.fukiage@gmail.com)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the NTT Communication Science Laboratories Ethics Committee under Approval No. R02-002, and performed in line with the Helsinki Declaration.

ABSTRACT Stereoscopic 3D displays (S3D), the most popular consumer display devices for 3D presentation, have a few problems that degrade the natural visual experience, such as unnatural relationships between eye vergence and accommodation, and severe image blurring (ghost) for viewers without stereo glasses. To simultaneously solve these problems, we combine gaze-contingent disparity remapping with Hidden Stereo in a manner that mutually compensates for their respective shortcomings. Gaze-contingent disparity remapping can reduce the vergence-accommodation conflict by shifting the disparity distribution around the gaze position to be centered on the display plane. Hidden Stereo can synthesize 2D-compatible 3D stereo images that do not produce any ghosting artifacts when the images for the two eyes are linearly fused. Thus, by using our new gaze-contingent display, while one viewer with glasses enjoys natural 3D content, many other glassless viewers enjoy clear 2D content. To enable real-time synthesis, we accelerate Hidden Stereo conversion by limiting the processing to each horizontal scanline. Through a user study using a variety of 3D scenes, we demonstrate that Hidden Stereo can effectively hide disparity information to glassless viewers despite the dynamic disparity manipulations. Moreover, we show that our method can alleviate the limitation of Hidden Stereo—the narrow reproducible disparity range—by manipulating the disparity so that the depth information around the gaze position is maximally preserved.

INDEX TERMS Stereoscopic 3D, backward compatible stereo, gaze-contingent display.

I. INTRODUCTION

Stereoscopic 3D (S3D) displays can convey 3D depth information by presenting two images of a stereo pair separately to the left and right eyes. Typical S3D displays present stereo images either in a spatial or temporal multiplexing way. Spatial multiplexing presents the left and right images in odd and even rows of the screen, respectively, while temporal multiplexing temporally alternates the left and right stereo images. In both cases, specialized 3D glasses are required in order to deliver the left and right images of the stereo image coming from the screen to the corresponding eyes. While there are other types of S3D displays, such as autostereoscopic displays and head-mounted displays, S3D displays that use 3D glasses have advantages such as higher spatial resolution or support for multiple viewers [1]. However, the S3D displays using 3D glasses have the backward compatibility problem that when the viewer does not wear 3D glasses, the left and right images appear to be overlapped on the screen, producing "ghost" or image blur. Therefore, viewers without 3D glasses cannot enjoy image content presented on S3D due to degraded image quality [2].

In addition to the lack of backward compatibility, S3D displays have a limitation in their ability to reproduce

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

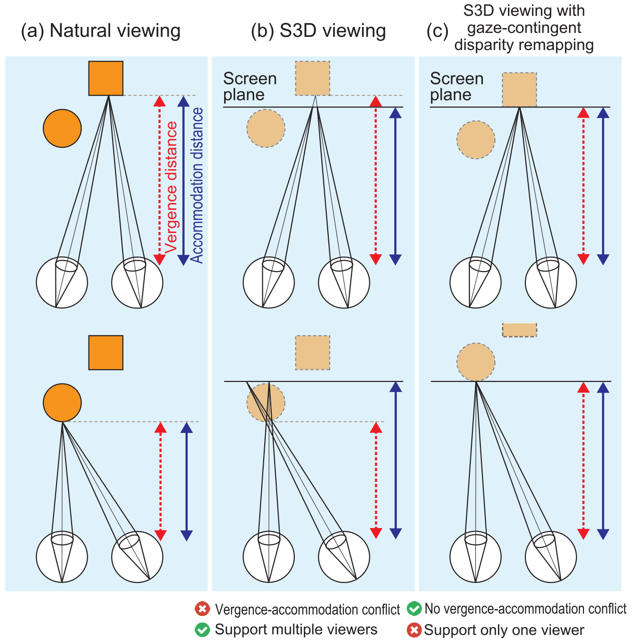
(a) Natural viewing
(b) S3D viewing
(c) S3D viewing with gaze-contingent disparity remapping
Vergence distance
Accommodation distance
Screen plane
Screen plane
✖ Vergence-accommodation conflict
✓ Support multiple viewers
✓ No vergence-accommodation conflict
✖ Support only one viewer

FIGURE 1. **Vergence and accommodation under three different viewing conditions. (a) Under natural viewing conditions, vergence and accommodation present consistent cues about the distance to the viewing object. (b) Under viewing with S3D displays, vergence and accommodation cues are decoupled because the accommodation distance is fixed to the screen plane. (c) Gaze-contingent disparity remapping techniques resolve this issue by dynamically changing the disparity so that the vergence distance to the currently viewed object is the same as the distance to the screen plane. As a trade-off, however, the number of viewers is limited to only one.**

natural binocular viewing. When looking at an object naturally, our eyes rotate to allow the object to be seen in the fovea of each eye (vergence) while at the same time, focal distances of the lens in our eyes are adjusted to get sharp retinal images (accommodation). Under natural viewing conditions, vergence and accommodation change cooperatively depending on the depth of the object being focused on (Fig. 1 (a)). In S3D, however, this relationship is broken, and while the accommodation is fixed on the display plane, the vergence changes according to the disparity (Fig. 1 (b)). As the vergence and accommodation act as depth cues in the visual system, this conflict of information can produce problems such as visual discomfort and visual fatigue for 3D viewers [3], [4]. To tackle this problem, gaze-contingent disparity retargeting techniques have been proposed [5], [6], [7], [8]. These techniques reproduce the natural viewing condition by shifting disparity values around the gaze point to be centered on the display plane (Fig. 1 (c)). However, gaze-contingent disparity manipulation makes the backward compatible problem even more pronounced because for secondary viewers without glasses, the image ghost appears to change dynamically and unexpectedly depending on the gaze behavior of the primary viewer. Furthermore, since the gaze-contingent technique can only deal with one viewer, the number of viewers who can view the content (either in 3D or 2D) is limited to just one. This completely eliminates the advantage of S3D displays over goggle-type displays of being not fully personalized, allowing additional users to share visual experiences in the same place.

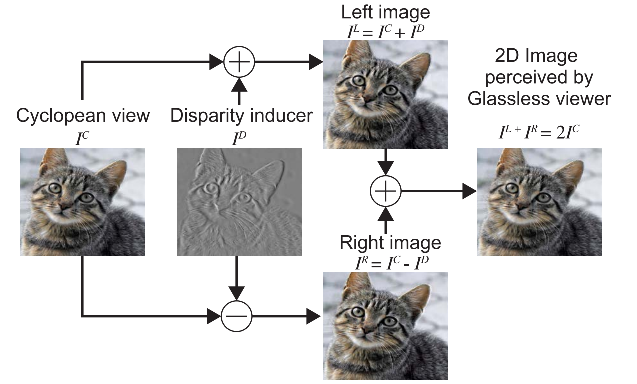

FIGURE 2. **Stereo image synthesis by Hidden Stereo. Hidden Stereo generates the left/right stereo images by adding/subtracting the disparity inducer pattern to/from the input cyclopean view. When viewed without glasses, the left and right images appear to be linearly combined on a S3D display. The linear fusion of the Hidden Stereo images cancels out the disparity inducer components and brings the image back to the original image.**

Therefore, in this work, we propose to incorporate a recently proposed technique, Hidden Stereo (HS), into gaze-contingent disparity manipulation. HS is a technique to achieve perfect backward compatibility on existing S3D devices [9] (Fig. 2). Presenting stereo images in Hidden Stereo format allows an unlimited number of viewers to enjoy the 2D version of the same content without any ghosting artifacts and dynamic image distortions caused by the gaze-contingent image manipulations (Fig. 3). In addition to the above advantage, our method can mitigate the limitation of HS in the reproducible disparity range by manipulating the disparity so that the depth information around the gaze point is maximally represented within this effective range (see the next section for details).

In order to realize the above idea, we have to significantly accelerate the HS algorithm, and we do so by restricting the computation to a single dimension (i.e., horizontal scanline) and parallelizing it on GPU. This enables real-time synthesis that is necessary for gaze-contingent retargeting. To ensure the smooth transition of the retargeting state across fixations, we use the seamless gaze disparity manipulation technique proposed by Kellnhofer *et al.* [8]. Through a user study using a variety of 3D scenes, we confirmed that the perfect backward compatibility of HS is preserved even when the disparity is dynamically changed by gaze-contingent retargeting. We also ensured that the proposed method could enhance depth impressions of HS while maintaining binocular image quality at an acceptable level under the effective disparity range.

A. BACKGROUND: HIDDEN STEREO

In HS, a stereo image pair is generated in such a way that when the left and right images are linearly combined, it results in a ghost-free 2D image representing the view from the

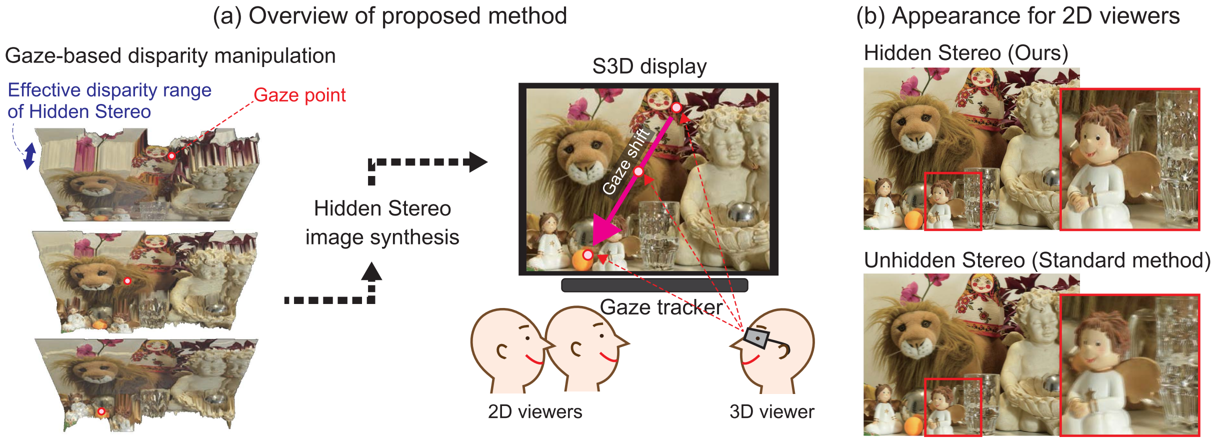
(a) Overview of proposed method
Gaze-based disparity manipulation
Effective disparity range of Hidden Stereo
Gaze point
Hidden Stereo image synthesis
S3D display
Gaze shift
Gaze tracker
2D viewers
3D viewer
(b) Appearance for 2D viewers
Hidden Stereo (Ours)
Unhidden Stereo (Standard method)

FIGURE 3. Overview of our method. (a) We retarget the disparity map based on the current gaze position such that the disparity information around the attended region is maximally preserved within the effective range of Hidden Stereo. Then, the stereo images synthesized by Hidden Stereo are presented on a S3D display. As a result, we achieve ghost-free viewing of stereo images for glassless 2D viewers while maintaining depth impressions superior to the original Hidden Stereo without gaze-contingent disparity manipulation. (b) Comparison of appearance for 2D viewers with our method (top) and a standard stereo synthesis method that explicitly shifts the input view (bottom).

intermediate point between the left and right eyes (i.e., cyclopean view). This is achieved by generating left/right images by adding/subtracting a disparity inducer pattern to/from the cyclopean view image (Fig. 2). The disparity inducer pattern is generated by shifting the spatial phase of the cyclopean image by $\pi/2$ with appropriate weights applied, such that after the addition / subtracting of the pattern, the phase of the cyclopean image is shifted to produce apparent disparities. In practice, the phase manipulation is operated in the multiscale bandpass representation [10]. Because the same disparity inducer pattern is either added to or subtracted from the cyclopean image to generate the left or right stereo image, linear fusion of the two images cancels out the disparity inducer components and brings the image back to the original cyclopean image. The detailed algorithm will be presented later in Section III-A.

In exchange for perfect backward compatibility, HS has a limitation in the reproducible disparity size because it relies on the principle of additive phase shift. Specifically, the higher the spatial frequency of the image, the more difficult it becomes to add a large disparity. Fortunately, the sensitivities of disparity detectors in the human visual system (HVS) are tuned to the mid-frequency range and limited in high frequency bands [11]. Thus, HS is still capable of adding a modest amount of depth to natural images, which usually contain broad-band frequency information.

The effective range of HS is close to Panum's fusion area of 10 min, that is, the maximum disparity size that humans can binocularly fuse without vergence eye movements [12], [13]. Under a natural viewing environment or using the standard stereo presentation method, for large disparities beyond Panum's fusion area, a vergence eye movement will help the viewer perceive the binocular image clearly. In our proposed method, gaze-contingent disparity manipulation replaces the functional role of the vergence eye movements, thereby allowing representation of a wider range of scene disparity beyond the effective range of the original HS display.

II. RELATED WORK

A. GAZE-CONTINGENT BINOCULAR IMAGE MANIPULATION

Since the vergence eye movement is made in the direction of reducing the disparity of the gazing target, the histogram of the disparity around the gaze position becomes a distribution centered on zero under natural viewing conditions [14]. To reproduce the natural disparity distribution on S3D displays, techniques have been proposed that shift the zero disparity plane to the depth of the current gaze position [5], [6], [7], [15]. In [8], additional nonlinear disparity compression was performed to ensure that the disparity falls within the comfort zone. It was also confirmed that the gaze-driven disparity manipulation improves visual comfort in terms of both objective [6] and subjective measures [7].

A challenge that the gaze-driven techniques face is that the dynamic change in disparity mapping may be noticed by the viewer during eye movements. To overcome this issue, Kellnhofer *et al.* [8] developed a model that can predict the limits of the HVS to detect transient disparity changes, allowing for seamless disparity manipulation. In this work, we also rely on their visible disparity change predictor to achieve seamless disparity remapping for our system.

As an approach complementary to manipulating disparity, Maiello *et al.* [16] proposed to simulate dioptic blur on the peripheral retina based on the depth differences between the gaze position and the other image region. They reported that the peripheral blurring could facilitate binocular fusion when disparity was large. However, we do not incorporate gaze-contingent image blurring in our technique because it inevitably eliminates the backward compatibility to

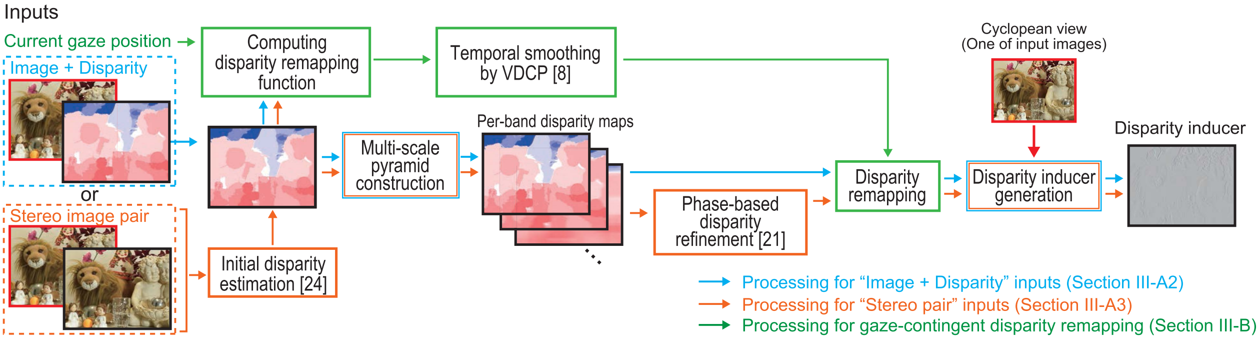
Inputs
Current gaze position
Image + Disparity
or
Stereo image pair
Computing disparity remapping function
Temporal smoothing by VDCP [8]
Multi-scale pyramid construction
Per-band disparity maps
Initial disparity estimation [24]
Phase-based disparity refinement [21]
Disparity remapping
Cyclopean view (One of input images)
Disparity inducer generation
Disparity inducer
Processing for "Image + Disparity" inputs (Section III-A2)
Processing for "Stereo pair" inputs (Section III-A3)
Processing for gaze-contingent disparity remapping (Section III-B)

FIGURE 4. **Process overview of the proposed method. The diagrams indicated by light blue, orange, and green represent the processing for the disparity inducer synthesis from an image-disparity pair (Section III-A2), the disparity inducer synthesis from a standard stereo pair (Section III-A3), and gaze-contingent disparity remapping (Section III-B), respectively.**

2D viewers. In addition, since our technique only provides the disparity range where sensory fusion is possible, the benefits of blurring to facilitate binocular fusion are considered to be marginal.

B. PHASE-BASED VIEW SYNTHESIS TECHNIQUES

The stereo image synthesis algorithm used in the proposed method is based on the recent development of phase-based view synthesis techniques. When generating a view based on a modified disparity map, the most common approach is grid-based warping of an original view image [17], [18], [19]. However, warping an image based on per-pixel disparity representation cannot correctly handle complex scenes containing specularity, defocus/motion blur, or transparency, where multiple depth can present at the same image region. As an alternative approach, Didyk *et al.* [20] proposed a phase-based disparity manipulation technique. In this technique, the left-right input stereo images are decomposed into multi-scale bandpass pyramids, and disparity is represented as phase differences between the left-right pair in the corresponding bands. Then, novel views are generated by inverting the pyramid representation after interpolating/extrapolating the phase differences. This technique can better represent scenes with complicated depth structures because of its ability to simultaneously represent different disparities for each frequency band as well as the sub-pixel accuracy.

However, the phase-based approach has a limitation in the supported disparity range. To overcome this limitation, Kellnhofer *et al.* [21] proposed to combine both approaches: they first compute a per-pixel disparity map as a rough estimate and then refine this map based on the phase-based approach to obtain per-band per-pixel disparity representation. A novel view is generated by translating bandpass images based on the manipulated disparity map in the corresponding band. Their technique retains the advantages of the phase-based approach while being capable of handling large disparities as in the image warping method.

Hidden Stereo (HS), which we use to generate ghost-free stereo images, is similar in essence to the phase-based technique, as both manipulate disparity by shifting the phase information. The key difference between them is that HS uses a linear operation (i.e., the addition of a disparity inducer) to shift phase in order to achieve perfect backward compatibility to 2D viewing. When converting a standard stereo pair into HS format, we also utilize Kellnhofer *et al.*'s technique [21] to accurately represent large disparities in the input stereo pair without potential mismatches in high-frequency bands (Section III-A3).

III. METHOD

In this section, we describe the realization of the proposed method in detail. The overview of the process is presented in Fig. 4. We assume either a pair of an image and corresponding disparity map or a standard stereo image pair for input. Additionally, the current gaze position, represented as x-y coordinates in screen space, is assumed to be given. The input images are used to compute a per-band disparity representation as well as to generate a disparity inducer pattern, which is then used to synthesize a Hidden Stereo (HS) image pair. The gaze position is used to compute a disparity remapping function that shifts and compresses the original disparity maps.

In the following, we first describe the basic algorithm of HS, and then explain how to generate disparity inducers from general image inputs. After that, we describe how gaze-contingent disparity retargeting is applied during the above process.

A. VIEW SYNTHESIS BY HIDDEN STEREO

1) BASIC ALGORITHM

The basic algorithm we use is the same as the original method in [9]. However, we restrict the computation to one-dimension (i.e., horizontal scanline of the image) to make it feasible for real-time disparity manipulation.

Let us assume a simple example where the original image is a sinusoidal pattern with a spatial frequency of ω (Fig. 5). Although the actual image does not take negative intensity, we can think of this example as a single frequency component after Fourier decomposition. To produce disparity (i.e., horizontal displacement) in this pattern, we add/subtract a quadrature($\pi/2$)-phase-shifted version of the original pattern, scaled by a weight A. According to the basic formula

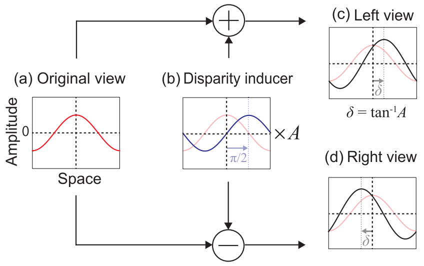

FIGURE 5. **The basic mechanism to produce disparity in Hidden Stereo. Here we assume an image whose intensity profile is a sinusoidal wave in a horizontal scanline (a). The disparity inducer (b) is generated by shifting the original phase by $\pi/2$. The left and right stereo images, (c) and (d), are generated by addition of (b) to (a) and subtraction of (b) from (a), respectively. The resulting phase shift size δ in (c) and (d) can be controlled by multiplying a weight A to the disparity inducer. Linear fusion of the pair (c) and (d) cancels out the disparity inducer (b) and makes the intensity profile the same as the original wave (a).**

for composite trigonometric functions, the above operation yields a pattern of the same spatial frequency with its phase shifted by $\phi = \arctan A$:

$$\begin{aligned}\sin \omega x + A \sin(\omega x + \frac{\pi}{2}) &= \sqrt{1 + A^2} \sin(\omega x + \phi),\\ \sin \omega x - A \sin(\omega x + \frac{\pi}{2}) &= \sqrt{1 + A^2} \sin(\omega x - \phi), \quad (1)\end{aligned}$$

Therefore, we can control the amount of disparity by adjusting the weight A. When we want to produce a disparity of size d, the required phase shift size is $\phi = \omega d/2$. The required weight value is thus

$$A = \tan \frac{\omega d}{2}. \quad (2)$$

It should be noted that there is a limit in the reproducible disparity size. In theory, a sinusoidal pattern cannot be displaced beyond half of its wavelength. Thus, the maximum disparity that can be achieved is π/ω.

2) SYNTHESIS FROM AN IMAGE-DISPARITY PAIR

In practice, we apply the above operation after decomposing an image into multiple bandpass components as shown in Fig. 6. Inspired by [21], we use 1D versions of filters employed in the complex steerable pyramid [10]. However, we do not use the lowpass residual component in the pyramid as it is not necessary to construct disparity inducer patterns.

Here, we assume that a single 2D image I and a corresponding disparity map D are given. Positive and negative disparity values in D represent the disparity closer to (i.e., crossed disparity) and farther away from the viewer than the screen plane (i.e., uncrossed disparity), respectively. We first decompose the input image by applying a series of filters ψ to discrete Fourier transform (DFT) of the I. When applying DFT, we use the Periodic Plus Smooth Decomposition technique [22] to efficiently remove artifacts caused by the periodic boundary condition imposed by DFT. Let $\mathcal{F}$ and $\mathcal{F}^{-1}$ be the DFT and its inverse function, respectively, the complex bandpass responses of I in the f-th spatial frequency band are

$$B_f = \mathcal{F}^{-1}\left(\psi_f \mathcal{F}(I)\right). \quad (3)$$

Then, we get the quadrature-phase shifted responses $\tilde{B}$ by taking the imaginary part:

$$\tilde{B}_f = \mathrm{Im}[B_f]. \quad (4)$$

Next, we compute the weight values A that are applied to the quadrature-phase components. The weights are determined based on the input disparity map D. To prevent aliasing, we first construct a multiscale pyramid representation of the disparity map, G^D, so that G^D in the f-th level (G^D_f) only contains spatial frequency bands less than and equal to $\tilde{B}_f$. To compute G^D_f, we average disparities in its local neighborhood along the horizontal scanline over a range equal to the wavelength of each band f, following [21]. Then, according to Eq. 2, the weight function A can be written as:

$$A_f = \tan \frac{\omega_f G^D_f}{2}, \quad (5)$$

where ω_f denotes the peak frequency of the f-th frequency band. Note that the above operation preserves the signs of disparity values, and if the weight A is positive, the resulting disparity inducer produces crossed disparity, while if the weight A is negative, the resulting disparity inducer shifts the image in the opposite direction, producing uncrossed disparity.

The weight values computed by the above equation can become infinitely large as the input disparity value is close to a half of the wavelength of each frequency band. Therefore, following [9], we limit the maximum absolute weight to one as:

$$A'_f = \min(\max(A_f, -1), 1), \quad (6)$$

This means that the maximum phase shift of each bandpass component is limited to $\pm\pi/4$. Although limiting the weight values in this way can cause inconsistency in disparity sizes across spatial frequencies, it does not produce significant problems as long as the disparity sizes do not exceed a certain limit (which was subjectively measured as the effective disparity range of HS in [9]). This is because the HVS has independent disparity detection mechanisms, each tuned to various ranges of spatial frequencies [12], [23]. A small inconsistency across disparity detected by those mechanisms will be resolved later in the integration process (please refer to Appendix in [9] for details).

Finally, we can obtain the disparity inducer I^D by reconstructing the weighted quadrature-phase components. The reconstruction can be easily performed by summing up all the band-pass components after applying the same filters as used in the decomposition:

$$I^D = \mathcal{F}^{-1}\left(\sum_f \psi_f \mathcal{F}(A'_f \tilde{B}_f)\right). \quad (7)$$

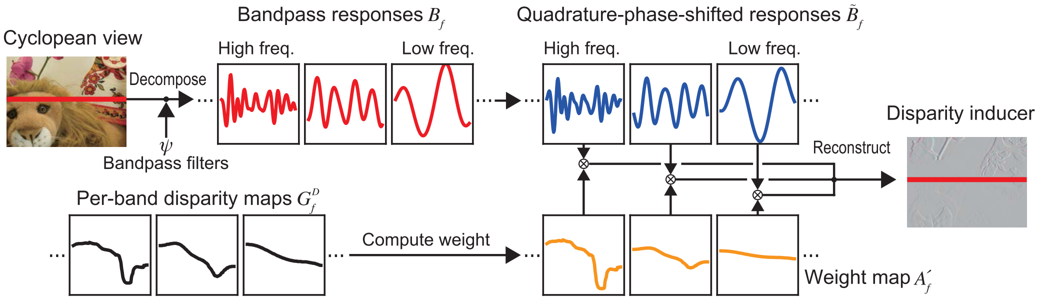

FIGURE 6. Process of generating a disparity-inducer at a horizontal scanline (i.e., the red horizontal line in the input/output images).

3) SYNTHESIS FROM A STANDARD STEREO PAIR

In another practical scenario, one may want to convert standard stereo images into ghost-free HS images. A simple solution to achieve this is to compute a disparity map and generate a disparity inducer for one of the input stereo images as described in Section III-A2. As an alternative way, Fukiage *et al.* [9] directly computed phase differences between bandpass components of the input stereo pair and used them to obtain weight values A for each bandpass component. The latter approach has the advantage that it can represent per-band disparities for each pixel, which better captures disparities in complex scenes containing specularity, defocus/motion blur, and transparent objects [20]. However, the phase-based technique has a limitation in that the disparity estimation fails when the disparity range in the stereo pair is relatively large. To overcome this limitation, we first compute a per-pixel disparity map as a rough estimate and refine it for each spatial frequency band based on the phase-based approach as done in [21].

Following [21], we first compute a rough disparity estimate D using a method of Hosni *et al.* [24] from an input stereo pair I^L and I^R. Here, we assume that the stereo pair is rectified so that we can perform the following processing independently within each horizontal scanline. From the per-pixel disparity map D, we initialize the per-band disparity maps $G_f'^D$ by averaging disparities in its local neighborhood over a range equal to the wavelength of each band f.

The initial disparity maps are refined using residual phase differences between the input stereo pair. For this, we first decompose I^L and I^R into complex bandpass responses B_f^L and B_f^R, respectively. Then, we find the correspondences between B_f^L and B_f^R using $G_f'^D$; for each position x in the left bandpass response $B_f^L(x)$, the corresponding right bandpass response is found at the closest pixel to $x - G_f'^D(x)$. The per-band disparity maps are refined using the phase differences between these corresponding bandpass responses $\Delta\phi$ as

$$G_f^D = G_f'^D + \frac{\Delta\phi}{\omega_f}. \tag{8}$$

After the refined disparity maps are obtained, the process to generate the disparity inducer is the same as in Section III-A2 (Eqs. 5-7). Finally, HS image pairs I'^L and I'^R are obtained by adding/subtracting the disparity inducer I^D to/from the input left image I^L.

4) ADDITIONAL PROCESSING

Below, we describe a few additional implementation details when generating a disparity inducer.

a: CLIPPING DISPARITY INDUCER

The addition of a disparity inducer can cause the intensities of the resulting stereo images to exceed the displayable dynamic range. We handle this problem by clipping the disparity inducer wherever the intensities in the resulting stereo images exceed predefined bounds (e.g., [0,255]). Please refer to [9] for detail.

b: GAMMA CORRECTION

Care must be taken so that the disparity inducer is perfectly canceled out when images are actually presented on a 3D monitor. Typical display devices have a nonlinear response function. In order to compensate for this, images are usually encoded as gamma-corrected values (i.e., sRGB color space). Thus, we first linearize the input image(s) and generate a Hidden Stereo pair in the linear color space. Then, we convert the Hidden Stereo pair back into the gamma-corrected space before sending them to the monitor.

c: COLOR PROCESSING

When processing color images, we simply process each of the RGB channels independently to obtain an HS image pair. However, the computational cost can be reduced by applying HS conversion only to the luminance channel, exploiting the fact that the luminance information is dominant in human stereopsis [12]. In this case, we first convert the input images into the YUV color space and apply HS conversion to the Y channel only. Then, the HS image pair in the YUV color space is converted back to the original RGB color space. For this color conversion, we followed the formula provided in ITU-R Rec. 601 [25].

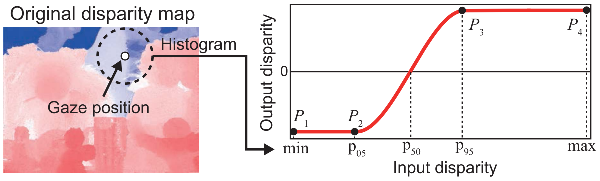

FIGURE 7. **Process to construct a disparity mapping function. A histogram of disparity values around the gaze position is constructed. The 5 and 95 percentiles of the histogram as well as the minimum and maximum disparity values are chosen for control points (P_1 to P_4) for the disparity mapping function. The control points are also vertically shifted so that the 50 percentile point of the histogram comes close to zero disparity. The control points are then smoothly interpolated to construct a disparity mapping function.**

B. GAZE-CONTINGENT DISPARITY RETARGETING

To make the most of the effective disparity range of HS, we retarget the disparity range depending on the current gaze position. The retargeting algorithm we use is based on Kellnhofer *et al.*'s technique [8]. In this technique, the disparity values around the gaze position are retargeted within a certain range while the transition of the disparity remapping function is smoothed so that the temporal artifact due to disparity manipulation becomes imperceptible.

1) CONSTRUCTING THE DISPARITY REMAPPING FUNCTION

Here, we describe how the disparity remapping function is defined given an input disparity map D and gaze position x. In the case of stereo image conversion (Section III-A3), we use the initial disparity estimate for D.

Figure 7 shows the process used to construct a disparity remapping function. As the first step, a histogram of disparity values around x is constructed. The contribution of each disparity value to the histogram is weighted according to a Gaussian function centered at x. The standard deviation of the Gaussian function is set to 2.5 deg, following [8]. This range was determined based on the fact that stereo acuity is significantly impaired beyond it [26].

Then, the disparity values from the 5th percentile (p_{05}) to the 95th percentile (p_{95}) of the histogram are retargeted to a pre-specified target range $[d_{min}, d_{max}]$. We assume that the target range is set to fall within the effective range of HS, which is around 6-8 min according to the previous work [9]. We will also investigate the optimal range using the current implementation in a user study (Appendix D). The remapping function is defined by four control points P_1-P_4 defined as:

$$\begin{aligned} P_1 &= [\min(D), d_{min}], \\ P_2 &= [\mathrm{p}_{05}, d_{min}], \\ P_3 &= [\mathrm{p}_{95}, d_{max}], \\ P_4 &= [\max(D), d_{max}]. \end{aligned} \tag{9}$$

To prevent the remapping function from magnifying disparity values beyond the original values, the slope between P_2 and P_3 is constrained so as not to exceed one, as done in [8].

We also vertically shift all the control points so that the line connecting P_2 and P_3 crosses the point $[p_{50}, 0]$. This makes the disparities of the gaze position closer to zero disparity (i.e., screen plane). Then, we additionally clip the disparity range to $[d_{min}, d_{max}]$ as the control points can deviate from the range by the vertical shift. Although this clipping moves the zero-crossing point slightly away from $[p_{50}, 0]$, we do not perform further refinements since the perceptual gain achieved does not justify the additional computational cost.

Finally, the intermediate points between these control points are smoothly interpolated by the piecewise cubic Hermite interpolating polynomial. The disparity mapping function is applied to the per-band disparity maps G_f^D for each spatial frequency band f.

2) TEMPORAL SMOOTHING OF DISPARITY REMAPPING

Directly applying the remapping function computed for each frame produces sudden disparity changes that are visible to users with 3D glasses. To prevent this, the remapping function is temporally smoothed so that the amount of transition in terms of disparity scaling and shifting does not exceed a certain threshold. For the threshold, we use twice the detection threshold predicted by the visible disparity change predictor (VDCP) developed by Kellnhofer *et al.* [8] because this was found to be the best compromise between the depth reproduction and stability for natural scenes in their subjective experiment.

IV. RESULT

A. IMPLEMENTATION AND PERFORMANCE EVALUATION

We evaluated the performance of the proposed algorithm for both input cases: Synthesis from an image-disparity pair (Section III-A2) and synthesis from a standard stereo pair (Section III-A3). The number of pyramid levels was determined depending on the width W of the input image as $\lceil \log_2 W - 3 \rceil$. We implemented all the processing, including gaze-contingent disparity manipulation and HS image conversion, on GPU using CUDA. Here, we assume that even in the case of having a standard stereo pair as input, the disparity map D is either given or precomputed and stored as an image together with the input stereo pair.

For each input type, we tested the three different variants of implementation to process color images.

- *RGB* Each of the RGB channels is independently processed to obtain an HS image pair.
- *Lum.* The image is processed only in the luminance channel after being converted to the YUV color space as described in Section III-A4.c.
- *Lum. half* To further reduce the computational cost, the disparity inducer is computed with half the resolution in the horizontal direction. The generated disparity inducer is then upscaled to the original size by linear interpolation and added to, or subtracted from, the cyclopean view image. This approach works because the disparity detection mechanisms in the HVS are not very sensitive to high frequency patterns [12].

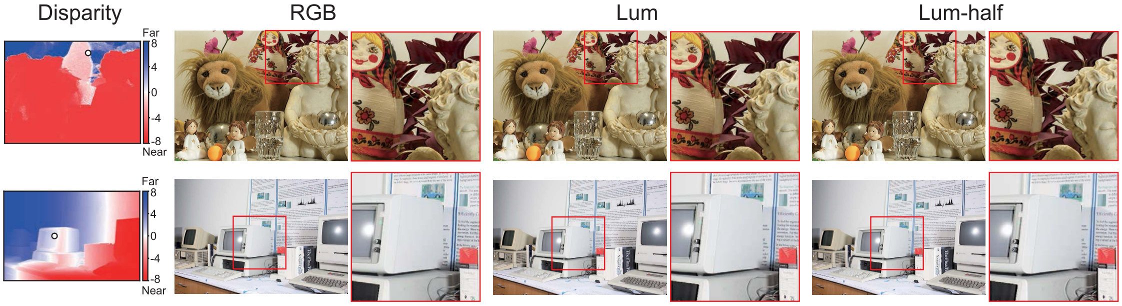

FIGURE 8. Comparison of Hidden Stereo images obtained by generating the disparity inducer in different ways. The leftmost column shows disparity maps for each scene. The white dot in the disparity map indicates the gaze position used to compress disparity. The second to last columns show the left image of the stereo pairs generated by each method. The area in the red rectangle is enlarged and presented on the right side. (*RGB*) Disparity inducers are generated by independently processing each of the RGB channels. (*Lum*) Disparity inducers are generated by only processing the luminance components of the input images. (*Lum-half*) Disparity inducers are generated by only processing the luminance components of the input images with half the resolution in the horizontal direction. The results show that the quality comparable to *RGB* can be obtained by *Lum* or *Lum-half*.

TABLE 1. Performance of the proposed method measured in frames per second.

Input type	Input size	Performance (FPS)		
		RGB	Lum.	Lum. half
Image + disp.	1280 × 720	42.8	139.7	211.6
	1920 × 1080	26.5	76.5	128.0
Stereo pair	1280 × 720	30.1	59.7	106.8
	1920 × 1080	17.1	31.5	55.0

Figure 8 shows some examples of comparison between the approximated versions of implementation. The results demonstrate that the differences are negligibly small.

For each of the above three implementations (denoted respectively as *RGB*, *Lum.*, and *Lum. half*), we measured the overall performance of our algorithm in frames per second (FPS) for two different input sizes (i.e., 1280 × 720 and 1920 × 1080) for each input type (i.e., image-disparity pair and stereo pair). The performance was measured on a desktop computer with a NVIDIA Geforce RTX 3090 (24GB GPU memory).

The results indicate that the proposed method can run at more than 30 FPS for Full-HD resolution inputs when processed in the single luminance channel or with reduced internal resolution. The large performance difference between the two input types is due to the presence/absence of the phase-based disparity refinement process, which accounts for about 40% (*RGB*) or 60% (*Lum*) of the entire processing time. The computational time required for the disparity remapping was negligibly small (up to a few percent) compared to the view synthesis process.

B. EFFECT OF GAZE-CONTINGENT DISPARITY REMAPPING

Examples of the results obtained by the proposed method are presented in Fig. 9. Here, disparities in the original input stereo pair (top row) are compressed to $[-8, 8]$ min, which is around the effective disparity range of HS [9]. In global compression (the second row), we globally compressed the original disparity map by the disparity remapping function obtained by using the 5th and 95th percentiles of the disparity histogram constructed from the entire image. Due to the excessive compression, the depth profile appears to be unclear in some parts of the image, especially in the lower and upper areas. Using gaze-contingent compression (the third and fourth rows), the depth structures are preserved around the gaze position (indicated by the white dot in the disparity map), and the depth impressions comparable to the original stereo image can be perceived in that area. For more results including videos with dynamic disparity remapping, please refer to the supplementary material. In Appendix C, we quantitatively show the degree to which gaze-contingent disparity remapping can improve the local disparity range around gaze positions.

C. COMPARISON WITH AN ALTERNATIVE VIEW SYNTHESIS METHOD

We also compare our results with those obtained by an alternative view synthesis technique that does not hide disparity information for 2D viewers. We refer to this alternative method as UnHidden Stereo, or UHS for brevity.

For this comparison, we used a view synthesis technique developed by Kellnhofer *et al.* [21] as it also relies on the same per-band per-pixel disparity representation. In this technique, a stereo image pair is generated by displacing each bandpass image of one of the input stereo images (i.e., the left image) in two opposite directions according to the target per-band disparity map. For this, the wavelet coefficients around occluded regions are first attenuated to avoid mixing foreground and background signals. Then, the non-uniform Fourier transform is used to displace the wavelet position according to the disparity map. Please refer to the original paper for details.

Figure 10 presents the comparison results. Here, disparities in the original input stereo pair are compressed to $[-8, 8]$ min according to the gaze position indicated by the white dot. The comparison of stereo images shown in the first to third columns demonstrate that the perceived depth impression in

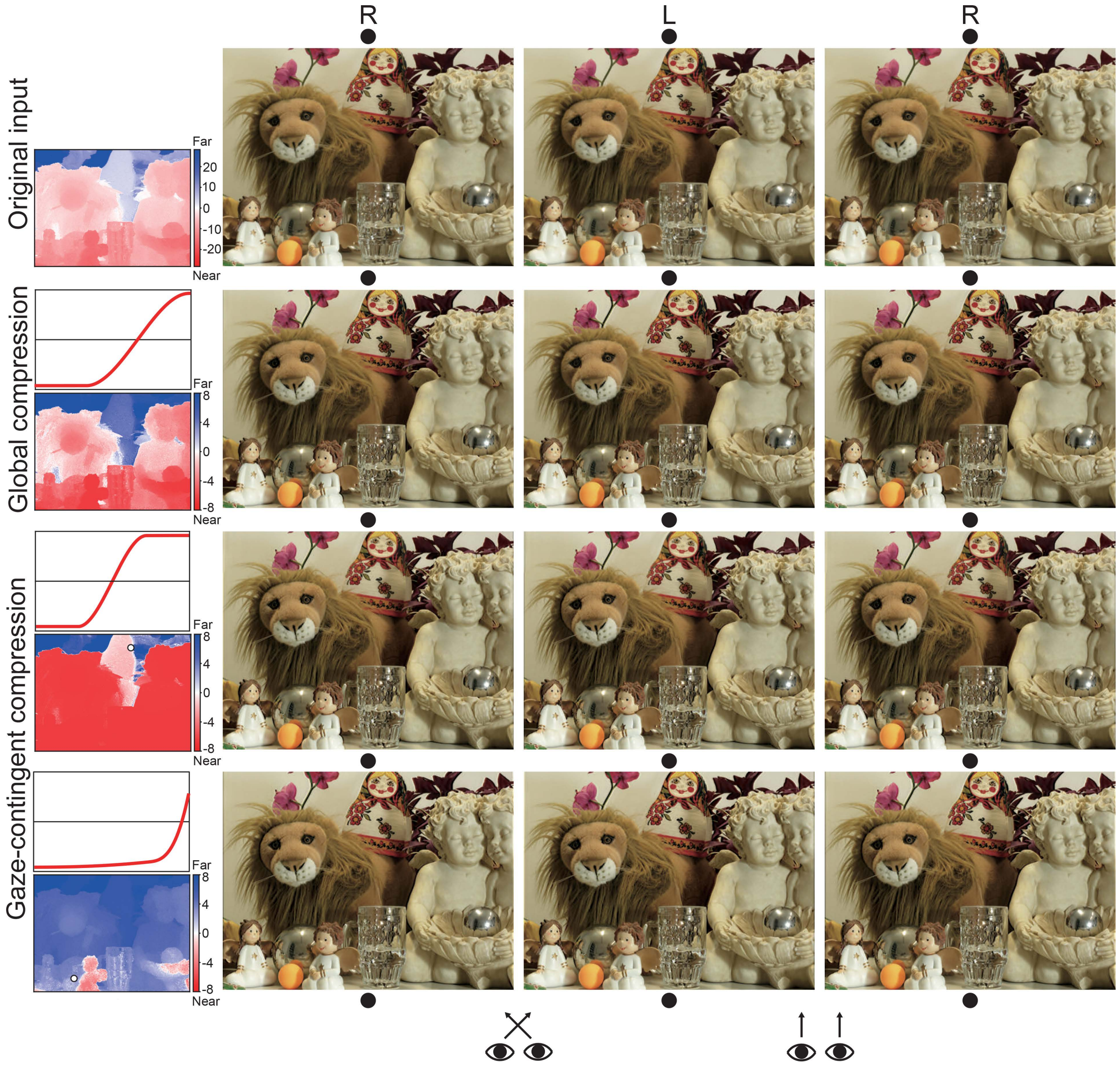

FIGURE 9. Results of gaze-contingent remapping applied to Hidden Stereo. The top row presents the input stereo pair. The second row shows the result of global disparity compression. The third and fourth rows show results of gaze-contingent remapping obtained with two different gaze positions. The leftmost column shows the disparity remapping functions and disparity maps used to generate the Hidden Stereo images. The white dot in the disparity map indicates the gaze position used to compress disparity. The image pairs in the second and third columns are for cross fusion, and those in the third and fourth columns are for parallel fusion. (The images in the second and fourth columns are identical.) The images are best viewed from a distance that is 1.5 times the image width.

HS is comparable to that in UHS. On the other hand, the simulated appearances of stereo images for glassless viewers (the rightmost column) show that UHS exhibits visible ghosts outside the gaze position while HS does not.

V. SUBJECTIVE EVALUATION

To demonstrate the effectiveness of our method, we conducted a user study. The primary objective of the study was to ascertain if our method can retain backward compatibility for 2D viewers under dynamic disparity manipulation. However, we were also interested in the depth impressions and image quality obtained by our method perceived by 3D viewers compared to those obtained by different variants of stereo synthesis methods. Therefore, the participants evaluated the stereo images both with and without 3D glasses. The study employed a pairwise comparison task, where the participants compared the stereo images generated by two different methods and chose the one they preferred.

A. METHOD

1) COMPARISON METHOD

We compare the proposed Gaze-Contingent Hidden Stereo technique (GC-HS) to the following three different

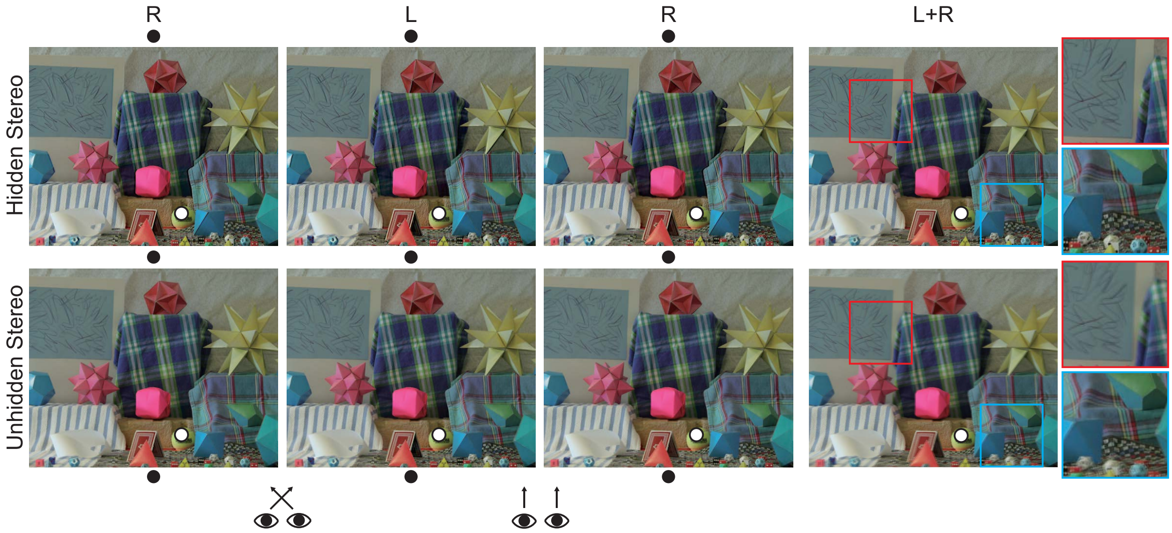

FIGURE 10. Comparison of stereo images generated by Hidden Stereo (top) and Unhidden Stereo (bottom). The image pairs in the first and second columns are for cross fusion, and those in the second and third columns are for parallel fusion. (The images in the first and third columns are identical.) The rightmost column presents the appearance of stereo images for glassless viewers, which were synthetically generated by averaging the left and right images. Unhidden Stereo exhibits visible ghosts outside the gaze position (the white dot) while Hidden Stereo does not. Please see the enlarged area enclosed in red and blue rectangles. The images are best viewed from a distance of 1.5 times the image width. The disparity range is compressed within [−8, 8] min according to the gaze position indicated by the white dot.

variants of the stereo synthesis methods: Hidden Stereo without gaze-contingent disparity remapping (HS), Unhidden Stereo (UHS), and Unhidden Stereo in conjunction with Gaze-Contingent disparity remapping (GC-UHS). We implemented all of the comparison techniques on GPU using CUDA to achieve real-time performance.

Below, we describe the comparison techniques in detail. For all the methods including the proposed method (GC-HS), stereo images are generated by using a standard stereo image pair as input.

a: WITH VS WITHOUT GAZE-CONTINGENT DISPARITY REMAPPING

Regarding the gaze-contingent methods (GC-HS and GC-UHS), the disparity map is compressed to a given range $[-d, d]$ by the disparity remapping function conditioned on the gaze location as in Section III-B. In the other methods (HS and UHS), the disparity map is globally compressed by the disparity remapping function obtained by using the 5th and 95th percentiles of the disparity histogram constructed from the entire image for d_{min} and d_{max}, respectively.

b: HIDDEN STEREO VS UNHIDDEN STEREO

For the methods using Hidden Stereo (GC-HS and HS), the process after the disparity map is obtained is exactly the same as described in Section III-A3. In the methods using Unhidden Stereo (UHS and GC-UHS), a stereo image pair is generated by displacing the bandpass image of one of the input stereo images (i.e., the left image) in two opposite directions as described in Section IV-C.

FIGURE 11. 3D scenes used in user studies.

2) APPARATUS

A 31-inch SONY LMD-X310MT monitor equipped with a passive (polarizing) 3D system was used as the presentation device. The resolution of the screen was 4096 × 2160, and the left and right stereo images were displayed in odd and even horizontal scan lines, respectively. The participants wore polarized 3D glasses to separately view the stereo pair in different eyes. A Tobii 4C EyeTracker (90Hz) was used to monitor gaze locations. The experiments were conducted in a dimly lit room. The observation distance was set at 80 cm so that the accuracy of the eye tracker could be kept sufficiently high.

3) STIMULI

The eight different scenes shown in Fig. 11 were used to generate stimuli. Five were camera-captured scenes taken from MPI Light Field Archive (MPI-LFA) [27] and Middlebury Stereo Datasets (MSD) [28], [29]. The remaining three were computer-generated scenes taken from the 4D Light Field Benchmark (4D-LFB) [30].

To ensure a fair comparison between the HS and UHS methods, the images were processed independently for each color channel because processing only in the luminance channel results in visible color artifacts especially for the UHS method under large disparity conditions. To maintain high frame rates in both methods, the images were resized to smaller scales when synthesizing stereo images and then upscaled to fit the screen size using bilinear interpolation. The image sizes used for image generation and presentation for each scene are summarized in Table 2.

The initial disparity estimates D of the two scenes from MPI-LFA were computed by Hosni *et al.*'s method [24]. For the remaining scenes, the ground truth disparity maps provided in each dataset were used for D because we wanted to eliminate the effect of disparity estimation accuracy as much as possible. Note, however, that even when ground truth disparity is present, we still used the phase-based disparity refinement process (Eq. 8) because it has an advantage especially when the scene contains transparent or reflective material as in *Kitchen*. The disparity ranges of the stereo pairs used for input to generate stimuli are summarized in Table 2. In Appendix A, we also evaluate the objectively measured image quality of the synthesized images generated by both the HS and UHS methods when using the tested 3D scenes for input.

4) PARTICIPANTS

Fourteen participants (aged between 20 and 45) who were unaware of the purpose of the study took part in the study. All had a normal or corrected-to-normal vision, and had experience viewing stereo video content at home, in movie theaters, or amusement parks. They also passed a stereo-blindness test using a random-dot stereogram.

Written informed consent was obtained from all the participants prior to the start of the experiment. The experimental protocol was approved by the Research Ethics Committee of our organization and was conducted in accordance with the ethical standards of the Declaration of Helsinki.

5) TASK

The experiment was divided into 2D sessions and 3D sessions, depending on whether the participants observed the stimuli with or without glasses. In both sessions, the participants compared stereo images generated by the proposed method (GC-HS) and one of the other methods and selected the preferred one by pressing a button. In the 2D session, the participatns were asked about their preference about image quality as "Which image showed better image quality?" In the 3D session, in addition to the above question, they were asked about their preference about depth impressions as "Which image showed a stronger sense of depth?"

For each session, the two stereo images were sequentially presented in random order for 10 seconds each, with a 2-second blank interval inserted in between. The participants were instructed to move their eyes to see various image areas. The presentation of the stimuli could be repeated as needed. Between the trials, a white dot was presented at the gaze position measured by the eye tracker, which allowed the participants to check if the eye tracker was monitoring their gaze position correctly. In the 2D session, we used gaze positions pre-recorded during the 3D session of the same scene to remap disparity in the gaze-contingent methods. The rationale for this was to simulate the situation in which a 2D viewer observes the scene manipulated based on the gaze position of a different 3D viewer.

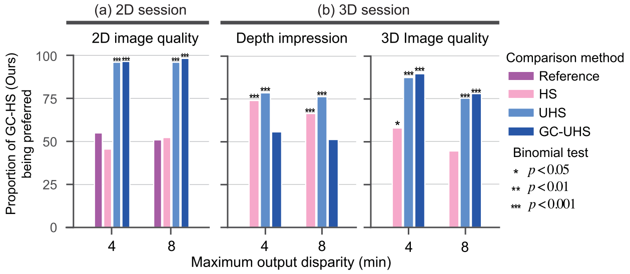

FIGURE 12. Results of pairwise comparisons in User study 1. Percentages for our method (Gaze-Contingent Hidden Stereo, GC-HS) being preferred in terms of 2D image quality (left), depth impressions (center), and 3D image quality (right) are shown. The 2D image quality results were obtained in the 2D session (without wearing 3D glasses) while those for depth impressions and 3D image quality were obtained in the 3D session (while wearing 3D glasses). The asterisks show the statistical significance obtained with a binomial test.

6) CONDITIONS

The disparity ranges of the input stereo pairs were compressed to either $[-4, 4]$ or $[-8, 8]$ min. The disparity ranges were determined based on the effective disparity range (i.e., 6-8 min) suggested in the previous work [9]. In the 2D session, the stereo image generated by the proposed method (GC-HS) was compared with the clean 2D reference image (where an identical image was presented to the two eyes) in addition to the stereo images of the three comparison methods. The reference image was a binocular presentation of the same monocular image (i.e., the input left image). The comparison with respect to the reference was included to check if the proposed method could actually "hide" the disparity information. In total, there were 64 conditions (4 comparison pairs × 8 scenes × 2 disparity ranges) in the 2D session and 48 conditions (3 comparison pairs × 8 scenes × 2 disparity ranges) in the 3D session.

7) PROCEDURE

Each participant ran the 2D and 3D sessions successively and performed each pairwise comparison twice in a randomized order. The order of sessions was counterbalanced across participants. At the beginning of each of the 2D and 3D sessions, the participants completed a practice session consisting of stimuli generated by both the HS and UHS methods using two scenes not used for the main experiment. During the practice session, the participants were instructed on what kind of image quality degradation could be typically observed. For example, we showed that unnatural lustrous

TABLE 2. Details about the input 3D scenes used in the user study. Columns 5 through 8 show the minimum, maximum, 5th percentile, and 95th percentile of disparity values in the scene, respectively.

Name	Dataset	Original size (px)	Presented size (deg)	Min (min)	Max (min)	5 %tile (min)	95 %tile (min)
FairyCollection	MPI-LFA	1024 × 720	34.4 × 24.2	-28.2	16.1	-13.6	16.1
Bikes	MPI-LFA	1024 × 720	34.4 × 24.2	-16.1	8.1	-12.1	8.1
Kitchen	4D-LFB	512 × 1024	24.6 × 24.6	-17.1	19.3	-15.3	9.0
Medieval	4D-LFB	512 × 1024	24.6 × 24.6	-18.0	21.0	-14.5	17.5
Tomb	4D-LFB	512 × 1024	24.6 × 24.6	-15.6	20.6	-15.0	14.1
Vintage	MSD	512 × 676	24.6 × 16.2	22.3	322.0	35.8	213.4
Shopvac	MSD	512 × 867	24.6 × 20.8	22.0	671.9	26.5	643.7
Moebius	MSD	512 × 817	24.6 × 19.6	0.0	57.8	18.6	49.8

impressions (typical artifacts in the HS method with large target disparity sizes) could be observed in the 3D session, and blur and double images could be observed in the 2D session. A 10-minute break was taken after every 20 trials. The entire experiment took approximately 4.5 hours for each subject.

B. RESULTS

The percentages for our method (GC-HS) being preferred over the other three methods in the 2D session are presented in Fig. 12(a). The results for individual scenes are presented in Fig. 14 in Appendix B. Binomial tests revealed a significantly higher preference rate for our method than for the two Unhidden Stereo methods (UHS and GC-UHS) in both disparity ranges. On the other hand, the preference rate was similar between our method and HS, and between our method and the monocular reference image. These results indicate that the disparity information was effectively hidden to viewers without glasses despite the dynamic disparity manipulations, while the ghost caused by the stereo disparity significantly degraded the image quality in the UHS methods.

Figure 12(b) presents the 3D session results. The results for individual scenes are presented in Fig. 15 in Appendix B. Binomial tests revealed that our method produces significantly stronger depth impressions than the two compared methods with global disparity compression (HS and UHS) in both of the tested disparity ranges. There was no significant difference between the depth impressions for our method and GC-UHS and the proportion was nearly 50/50. This indicates that our method can produce depth impressions equivalent to those of the unhidden stereo synthesis method under the tested disparity ranges (4 min and 8 min).

As for the image quality for 3D viewers, the results showed that the image quality of our method is as good as that of HS. This indicates that gaze-contingent disparity manipulation effectively improves the depth impressions of HS without degrading the 3D image quality. We also found that the image quality of our method (GC-HS) is not inferior to that of the two methods using unhidden stereo synthesis (UHS and GC-UHS). Rather, the image quality obtained with our method was unexpectedly superior. This might be because HS is more resistant to disparity estimation errors that are often observed in high frequency components due to mismatches in the phase-based refinement process. Alternatively, participants might simply prefer HS images because they tended to have slightly higher contrast due to its additive nature. See Appendix E for further discussion.

In summary, the subjective evaluation demonstrated that gaze-contingent disparity remapping and Hidden Stereo actually work together complementarily to compensate for each other's weaknesses: Hidden Stereo completely hides the dynamic disparity manipulations to 2D viewers while gaze-contingent disparity remapping enhances the depth impressions produced by Hidden Stereo without degrading 3D image quality for a 3D viewer. In an additional user study presented in Appendix D, we further show that the same conclusion is obtained by using the absolute quality rating with respect to reference images.

VI. LIMITATION AND FUTURE WORK

We limit the disparity range to the effective range of Hidden Stereo estimated in the previous work [9]. However, the optimal range depends on the contrast, spatial frequency of the content as well as the eccentricity. Building a perceptual model to predict this effect and optimizing disparity ranges for each content and for each eccentricity is therefore an important challenge for the future.

As a general limitation of gaze-contingent disparity remapping techniques, improvements achieved by gaze-contingent remapping are reduced in cases where a relatively wide range of disparity values are contained within a local region. This is typically observed when there is a complicated object with fine structures or transparent material.

In addition, temporal smoothing of the disparity mapping function [8] employed in this work degrades the effect of gaze-contingent manipulations when a viewer keeps moving his/her eyes. A promising solution is to exploit saccadic suppression to allow more aggressive disparity changes during saccadic eye movements (e.g., [31]).

VII. CONCLUSION

In this paper, we combined gaze-contingent disparity remapping with Hidden Stereo. This made it possible to present clear 2D images to an unlimited number of viewers without glasses, while improving the impression of depth obtained with Hidden Stereo for a viewer with glasses. Although our technique presents disparity sizes only within the upper

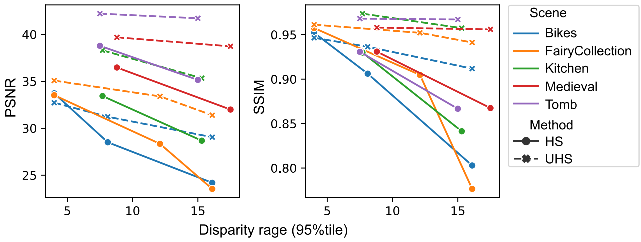

FIGURE 13. **Objective image quality scores of the synthesized images measured by PSNR (left) and SSIM (right).**

limit of binocular sensory fusion, gaze-contingent remapping replaces the functional role of the actual vergence eye movements by constantly shifting the disparity around the gaze position toward the screen plane. To enable real-time image synthesis, we simplified the original Hidden Stereo algorithm such that the computation is restricted to the 1D scanline and parallelized it on GPU. The subjective evaluation demonstrated the effectiveness of the proposed method.

APPENDIX A
COMPARISON OF OBJECTIVE IMAGE QUALITY OF SYNTHESIZED STEREO IMAGES PRODUCED BY HS AND UHS METHODS

To see the objectively measured quality of stereo images generated by the HS and UHS methods, we computed PSNR and SSIM between the synthesized views and the corresponding ground truth views. The images were generated from the 3D scenes tested in the user studies (Table 3). The input disparity levels 1, 2, and 3 were used for *FairyCollection*, *Bikes* and the input disparity levels 2 and 3 were used for *Kitchen*, *Medieval*, and *Tomb*. The other scenes were not included because the ground truth views that correspond to the synthesized views were not available. Note that we did not apply any disparity compression to synthesize these views. We computed the quality scores for the left and right images of each synthesized stereo pair and averaged them to obtain a single score for each input.

The results are presented in Fig. 13. We found that both PSNR and SSIM were higher overall in UHS than HS. In addition, the scores for HS significantly degraded as the disparity level increased while those for UHS remained moderately high even with large disparity levels. This is expected as HS has a limitation in the reproducible disparity range. However, in the user study, we found that the subjectively rated image quality under the 3D viewing conditions was significantly higher for HS when the disparity size was within the effective range of 8 min. Thus, these conventional image quality metrics are not a good predictor for binocularly presented stereo images. Whether the objective quality metrics specifically developed for stereo images can better handle HS-synthesized images is a topic to be investigated in future work.

APPENDIX B
INDIVIDUAL SCENES RESULTS OF THE USER STUDY

Figures 14 and 15 respectively present the individual scene results of the 2D and 3D sessions in the user study of the main paper.

APPENDIX C
QUANTITATIVE ANALYSIS ON THE EFFECT OF GAZE-CONTINGENT DISPARITY REMAPPING

In this section, we objectively analyze the degree to which gaze-contingent disparity remapping can improve the local disparity range around the gaze positions. To this end, we constructed local disparity histograms around gaze positions while participants observed five scenes (i.e., FairyCollection, Bikes, Kitchen, Medieval, and Tomb) in the user study described in the next section (Appendix D). The contribution of each disparity value to the histogram was weighted according to a Gaussian window with standard deviation equal to 2.5 deg. For each scene, three different input disparity levels (summarized in Table 3) were remapped to different disparity ranges. Specifically, the stereo pairs with disparity level 3 were remapped to the disparity ranges of $[-4, 4]$, $[-8, 8]$, and $[-16, 16]$ min, those with disparity level 2 were remapped to $[-4, 4]$ and $[-8, 8]$ min, and those with disparity level 1 were remapped to $[-4, 4]$ min. The disparity was remapped either gaze-contingently or globally as in Section V-A1.a. Then, for each input scene, we computed the average disparity range (95th percentile minus 5th percentile) over all fixations of all participants for each of the three conditions: (1) Original disparity (disparity in the input stereo image), (2) Gaze-contingently compressed disparity, (3) Globally compressed disparity.

The local disparity ranges obtained in this analysis are shown in Fig. 16. Here, the green data points indicate the disparity ranges of the original (input) scenes. The light red points (with dotted lines) indicate those after global compression. The red points (with solid lines) indicate those after gaze-contingent remapping. As expected, the global compression significantly reduced the local disparity range around gaze positions when the target disparity range (abscissa) was below the original one. On the other hand, the gaze-contingent remapping successfully increased the disparity range under these conditions.

APPENDIX D
SECOND USER STUDY: ABSOLUTE QUALITY RATING

The user study presented in our main paper demonstrated the effectiveness of our method relative to the other comparison methods using a pairwise comparison task. However, it is also important to investigate the perceived fidelity of each synthesized image to the original, the physically correct one, on an absolute scale. Therefore, in the second study, the participants compared a stereo image obtained with each method with the input stereo image and rated the image quality and depth strength in terms of degradation from the original input.

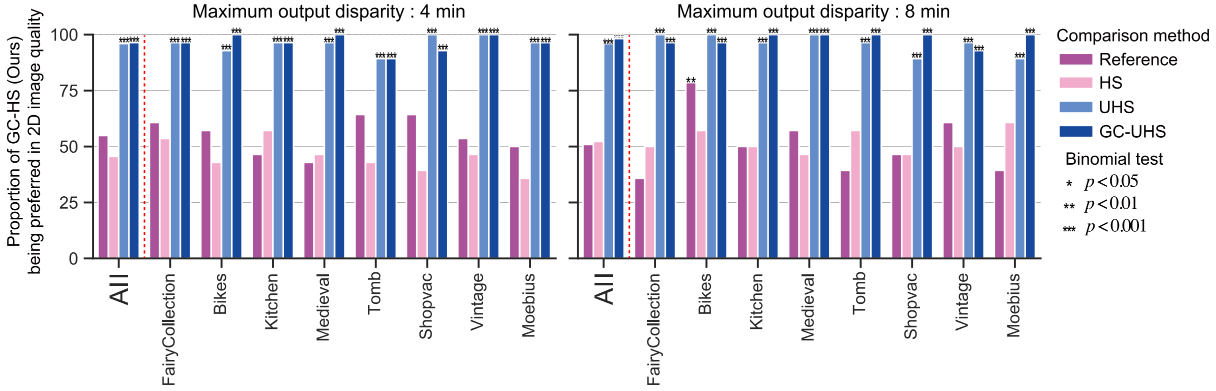

FIGURE 14. Results of pairwise comparisons in the user study (Section V in the main paper) when viewed without 3D glasses (2D viewing condition). Percentages of our method (GC-HS) being preferred in terms of image quality are shown. The results indicate that the 2D image quality of our method is as good as those of 2D clean reference images and HS, while much higher than those of unhidden stereos (UHS and GC-UHS).

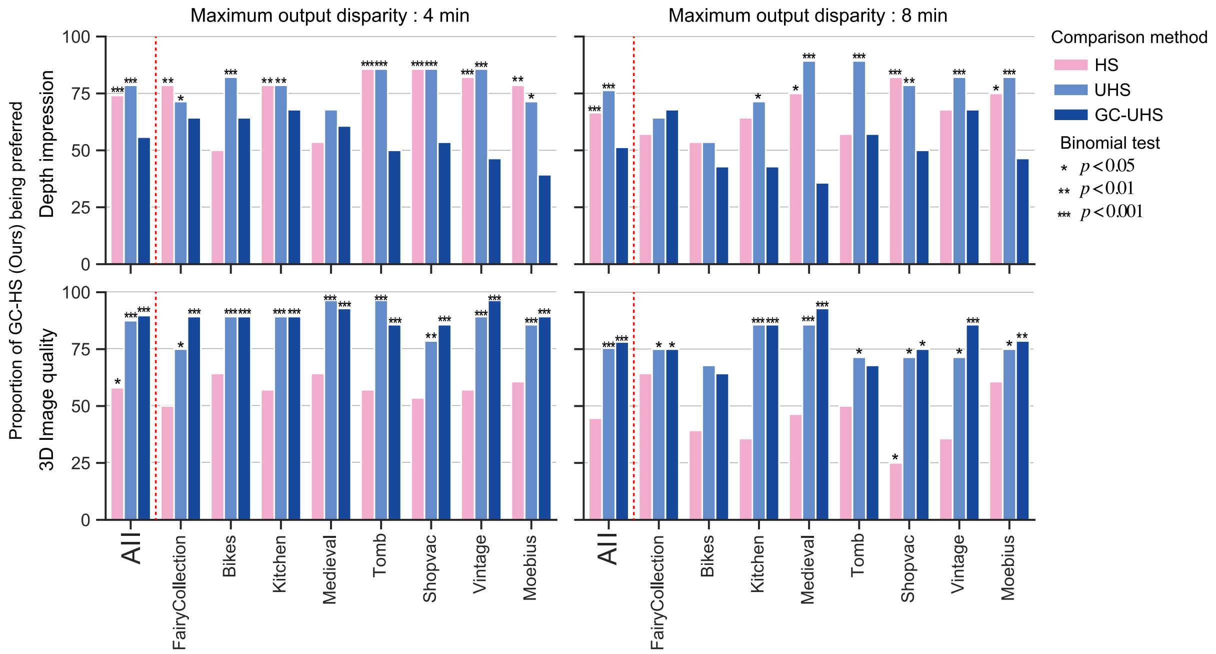

FIGURE 15. Results of pairwise comparisons in the user study (Section 5 in the main paper) when viewed with 3D glasses (3D viewing condition). Percentages of our method (GC-HS) being preferred in terms of depth impressions (top row) and image quality (bottom) are shown. The results indicate that our method improves depth impressions for the 3D viewer in comparison with those without gaze-contingent disparity remapping (HS and UHS), while retaining 3D image quality comparable to that of HS or better than those of unhidden stereos (UHS and GC-UHS).

A. METHOD

1) COMPARISON METHODS AND APPARATUS

The same comparison methods (i.e., GC-HS, HS, GC-UHS, and UHS) and apparatus that were used in the first user study were also used in this study.

2) PARTICIPANTS

Fifteen participants (aged between 20 and 45) who were unaware of the purpose of the study took part in the experiment. All had a normal or corrected-to-normal vision, and had experience viewing stereo video content at home, in movie theaters, or amusement parks. They also passed a stereo-blindness test using a random-dot stereogram.

3) TASK

As in the first experiment, the experimental sessions were divided into 2D sessions and 3D sessions. In both sessions, the participant compared a reference image with a synthesized stereo image (test) the disparity of which was compressed by either GC-HS, HS, GC-UHS, or UHS methods. In the 2D session, binocular presentation of the same monocular image (i.e., the input left image) was used as the reference, and the participants evaluated the impairment in perceived image quality by pressing a button. In the 3D session, the input stereo image was used as the reference, and participants evaluated the impairment in both perceived image quality and depth strength by pressing a button.

TABLE 3. Details of the input 3D scenes used in the quality rating experiment.

Name	Dataset	Original size (px)	Presented size (deg)	Disparity level	Min (min)	Max (min)	5 %tile (min)	95 %tile (min)
FairyCollection	MPI-LFA	1024 × 720	34.4 × 24.2	level 1	-8.1	4.0	-4.0	4.0
				level 2	-12.1	8.1	-12.1	8.1
				level 3	-28.2	16.1	-13.6	16.1
Bikes	MPI-LFA	1024 × 720	34.4 × 24.2	level 1	-4.0	4.0	-4.0	0.0
				level 2	-8.1	4.0	-8.1	4.0
				level 3	-16.1	8.1	-12.1	8.1
Kitchen	4D-LFB	512 × 1024	24.6 × 24.6	level 1	-4.3	4.8	-3.8	2.2
				level 2	-8.5	9.7	-7.7	4.5
				level 3	-17.1	19.3	-15.3	9.0
Medieval	4D-LFB	512 × 1024	24.6 × 24.6	level 1	-4.5	5.3	-3.6	4.4
				level 2	-9.0	10.5	-7.2	8.8
				level 3	-18.0	21.0	-14.5	17.5
Tomb	4D-LFB	512 × 1024	24.6 × 24.6	level 1	-3.9	5.1	-3.8	3.5
				level 2	-7.8	10.3	-7.5	7.0
				level 3	-15.6	20.6	-15.0	14.1

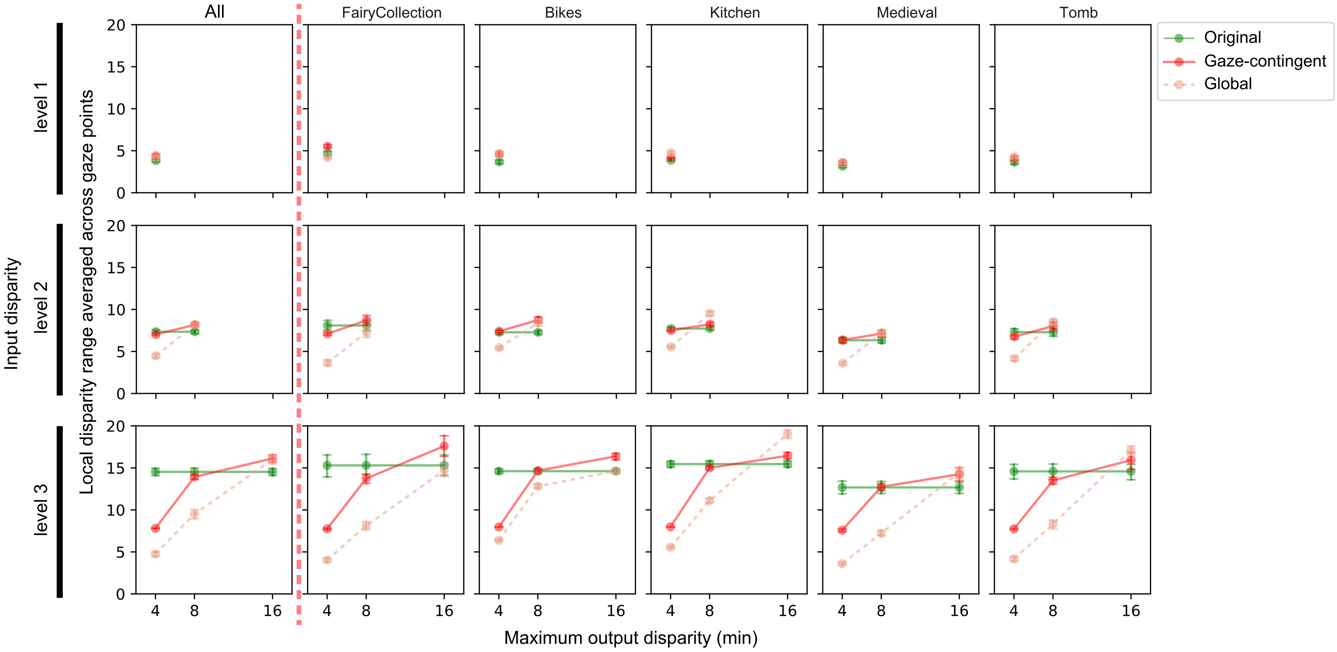

FIGURE 16. Local disparity ranges (95th percentile - 5th percentile) averaged across gaze points measured during the second user study. The first column shows the aggregated results over all the test scenes. The second to last columns show the results of individual scenes. The error bars represent 95% confidence intervals. For details, please refer to Appendix C.

In each trial, the reference and test images were presented sequentially for 10 seconds each, with a 2-second blank interval inserted in between. After that, the response screen was presented and participants were asked about the image quality (in both the 2D and 3D sessions) and the depth strength (only in the 3D session), namely, “Compared to the reference stimulus, to what extent did the image quality of the test stimulus appear to be degraded from the reference stimulus?” and “Compared to the reference stimulus, to what extent did the depth strength of the test stimulus appear to be degraded from the reference stimulus?” According to the double-stimulus impairment scale (DSIS) method [32], the participants rated the impairment on the following 5-point scale: 5 (imperceptible), 4 (perceptible, but not annoying), 3 (slightly annoying), 2 (annoying), and 1 (very annoying). The participants were instructed to move their eyes to see various image areas. The presentation of the stimuli could be repeated as needed. In the 2D session, we used gaze positions pre-recorded during the 3D session of the same scene to remap disparity in the gaze-contingent methods.

4) STIMULI AND CONDITION

Stereo pairs from the same scenes used in the local disparity analysis (Appendix C) were used as input to generate stimuli (summarized in Table 3). As in Appendix C, the stereo pairs with disparity level 3 were remapped to the disparity ranges of [−4, 4], [−8, 8], and [−16, 16] min, those with disparity level 2 were remapped to [−4, 4] and [−8, 8] min, and those with disparity level 1 were remapped to [−4, 4] min. Otherwise, the stimuli were generated in the same way as in the first user study. In the 2D session, a control condition in which the input left image was presented both as a reference and test

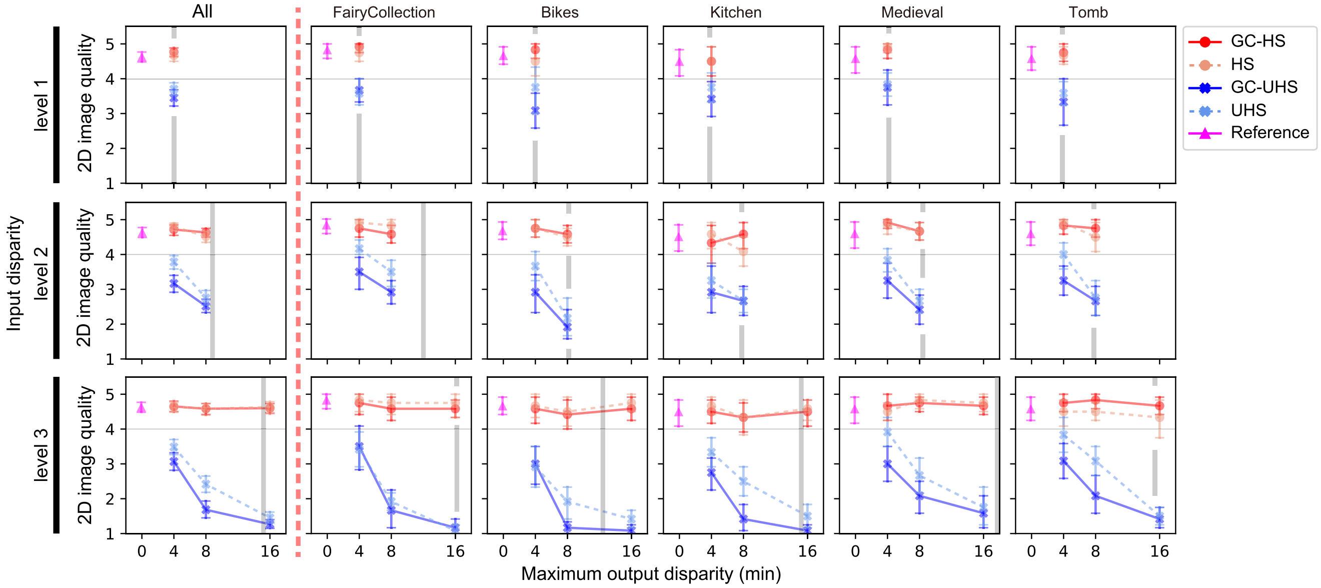

FIGURE 17. Image quality scores averaged across participants in the 2D session of the second user study. The first column shows the aggregated results over all the test scenes. The second to last columns show the results of individual scenes. The error bars represent 95% confidence intervals. Each row shows the results at different input disparity levels as shown in Table 3. The gray horizontal line in each plot indicates the tolerable quality threshold (i.e., *perceptible, but not annoying*. The thick vertical line in each plot indicates the 95 %tile of absolute disparities in the reference (input) 3D scenes. The average of the 95 %tiles across individual scenes are presented for the aggregated result (All).

was included for each scene to investigate the "upper limit" of the quality score (Note that the participants do not always select the fifth grade, that is, "imperceptible" even in this condition due to response errors or perceptual fluctuations). In total, there were 120 conditions (4 comparison methods × 5 scenes × 6 input-output disparity combinations) in the 3D session and 125 conditions (120 + 5 control conditions) in the 2D session.

5) PROCEDURE

Each participant ran the 2D and 3D sessions successively and evaluated all the conditions in a randomized order. The order of sessions was counterbalanced across participants. At the beginning of each of the 2D and 3D sessions, the participants completed a practice session consisting of stimuli generated by both HS and UHS methods using two scenes not used for the main experiment. During the practice session, the participants were instructed on what kind of image quality degradation could be typically observed in the same manner as in the first user study. A 10-minute break was taken after every 20 trials. The entire experiment took approximately 4.5 hours for each participant.

B. RESULTS

1) RESULTS OF SUBJECTIVE EVALUATION

a: 2D SESSION

Figure 17 presents the 2D session results. The scores were averaged across participants. The first column shows the aggregated results over all the test scenes while the remaining columns show the results of individual scenes. The results show that regardless of the target disparity range, the HS and GC-HS methods showed quality scores comparable to the control condition where the same reference image was presented as the test image. On the other hand, the quality scores for the UHS and GC-UHS methods decreased rapidly as the disparity range increased. Multiple comparison tests (Tukey's HSD) revealed significant differences in every combination between the two HS-based methods and the two UHS-based methods for all six disparity conditions ($p < 0.001$). Note that we also found significant differences between UHS and GC-UHS in the 4-min disparity condition at input disparity level 2 and the 4- and 8-min disparity conditions at input disparity level 3. This is likely because the temporal changes produced by the gaze-contingent disparity remapping appeared more annoying for viewers without glasses. In contrast, there was no difference between HS and GC-HS, which indicates that the temporal changes caused by disparity remapping were successfully made invisible in the proposed method.

b: 3D SESSION

Figure 18 presents the 3D session results. Again, the first column shows the aggregated results over all the test scenes while the remaining columns show the results of individual scenes. The depth strength scores shown in the odd rows of the figure were not substantially different across comparison methods. As a result of multiple comparison tests (Tukey's HSD) performed for each disparity condition on the scene-aggregated result (the first column), significant differences were found only between HS and UHS ($p < 0.01$) and between HS and GC-UHS ($p < 0.01$) in the 16-min output disparity condition at input disparity level 3. However, when we replotted the data for each condition without aggregating test scenes against the proportion of local disparity reproduction (the ratios of local disparity range relative to the original

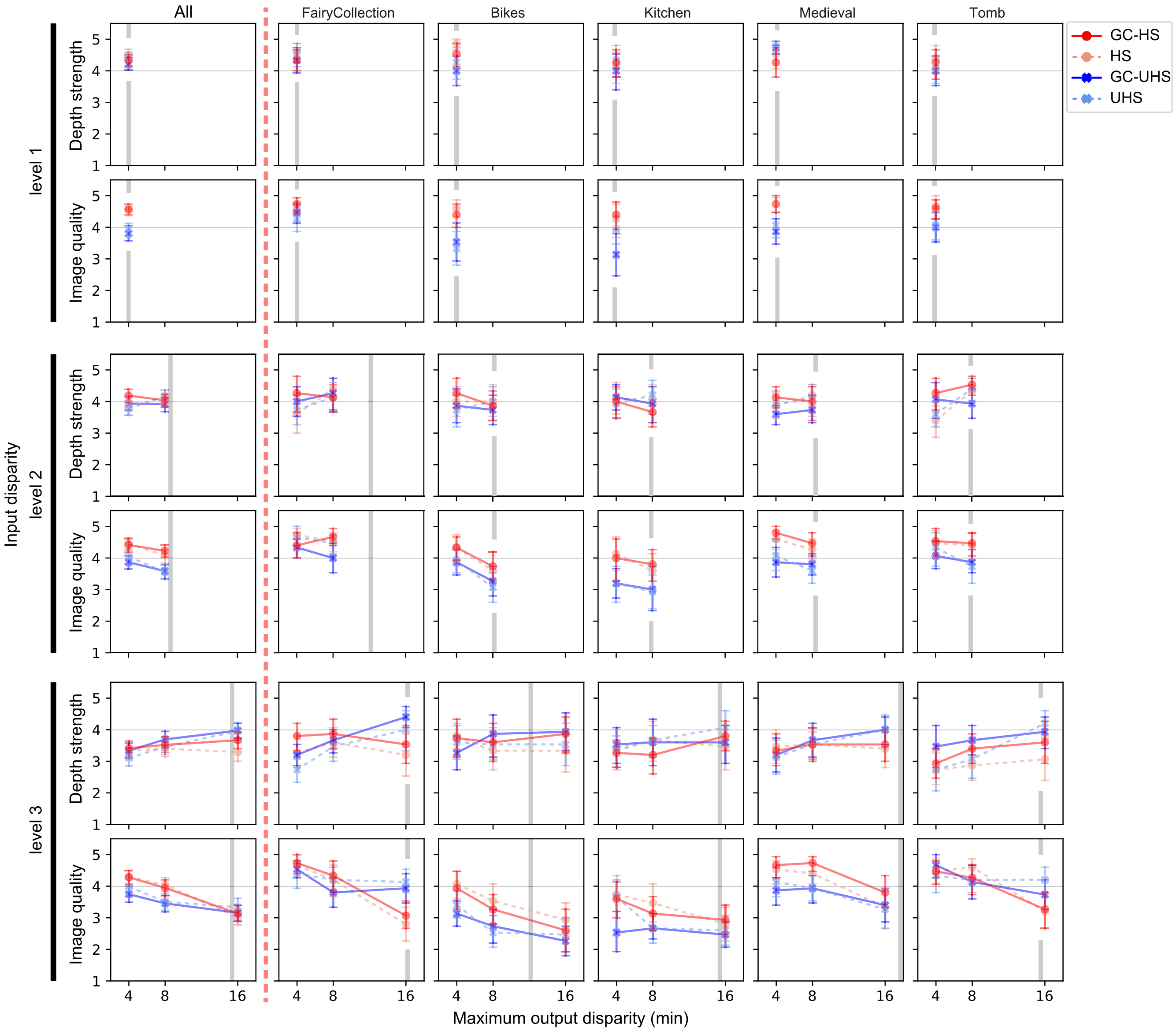

FIGURE 18. Depth (odd rows) and image (even rows) quality scores averaged across participants in the 3D session of the second user study. The first column shows the aggregated results over all the test scenes. The second to last columns show the results of individual scenes. The error bars represent 95% confidence intervals. Each two successive rows show the results at different input disparity levels as shown in Table 3. The gray horizontal line in each plot indicates the tolerable quality threshold (i.e., *perceptible, but not annoying*. The thick vertical line in each plot indicates the 95 %tile of absolute disparities in the reference (input) 3D scenes. The average of the 95 %tiles across individual scenes are presented for the aggregated result (All).

disparity range in Fig. 16), we found a clear trend suggesting that the increase in the local disparity range contributed to the improvement in perceived depth quality (Fig. 19). We computed the correlation between the depth quality and the proportion of disparity reproduction separately for each stereo synthesis method (i.e., HS and GC-HS, UHS and GC-UHS are grouped together to compute the correlations). As a result, we found positive correlations for both the HS-based method ($\rho = 0.513, p < 10^{-4}$) and the UHS-based method ($\rho = 0.764, p < 10^{-11}$). Therefore, it was confirmed that the depth impressions of Hidden Stereo, measured in terms of fidelity to the physically correct stereo image, can be improved by gaze-contingent disparity manipulation.

The results of 3D image quality scores (the even rows in Fig. 15) show that HS-based methods are clearly superior to UHS-based methods, consistent with the results of the first user study. According to the scene-aggregated results (the first column), both the HS and GC-HS methods maintained an acceptable image quality of 4 (perceptible, but not annoying) up to the 8-min disparity condition, while the UHS and GC-UHS methods fell below the acceptable quality at the 4-min disparity condition. Multiple comparison tests (Tukey's HSD) revealed that both the HS and GC-HS methods scored significantly higher than either the UHS method or the GC-UHS method in the 4-min and 8-min disparity conditions at all the input disparity levels

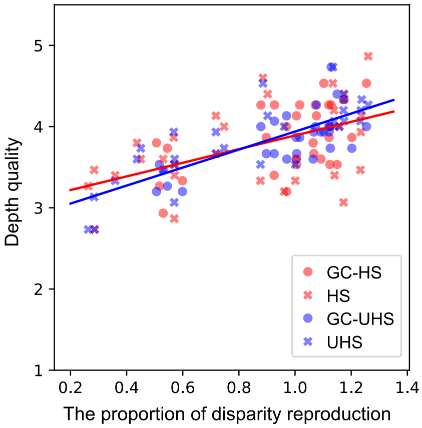

FIGURE 19. Depth quality scores vs. the ratio of remapped disparity ranges to the original disparity range. Each data point represents one of the conditions (the input scenes and the target disparity ranges) in the second user study. The red and blue lines are the linear fit to the data of HS-based methods (GC-HS and HS) and those of UHS-based methods (GC-UHS and UHS), respectively.

($p < 0.001$ for all comparisons). However, we also found that the image quality varies depending on the tested scene. Building a model that explains this content-dependent effect is thus an important future work.

C. DISCUSSION

Consistent with the results of the first user study presented in the main paper, the 2D session results showed that the disparity information was effectively hidden to 2D viewers by HS, even with dynamic disparity remapping. We also demonstrated that the UHS-based stereo images suffered from significant image quality degradation for 2D viewers even when the disparity range was compressed within as small as 4 min. The results also showed that the temporally varying ghost caused by dynamic disparity remapping further degraded the 2D image quality in the UHS method. Thus, the advantage of HS becomes markedly greater in the gaze-contingent S3D displays.

The 3D session results revealed the improvements in depth strength not only in objective measures (i.e., the local disparity range) but also in subjective evaluation. However, the results of the subjective evaluation indicated that the scores mostly fluctuated between 3 and 4, meaning that the participants tended to be satisfied even when disparities were significantly compressed. The reason for this may be because the sufficient amount of monocular depth cues that are needed to derive the depth structures already exist in natural scenes [33], [34]. However, in cases where monocular cues are unreliable and subtle depth differences have an impact on the task objective (e.g., estimating the height of buildings from a satellite image), we believe that the improved depth impressions (as demonstrated in the user study in the main paper) should directly affect user experiences.

The 3D session results suggested that the optimal disparity of the proposed method is about 8 min in visual angle, regardless of whether it is combined with gaze-contingent disparity remapping. This range closely matches the effective range of HS estimated in the original work [9]. Gaze-contingent remapping effectively assigns the local disparities around the attended position to this range and improves the depth impressions. The effective range of HS is close to Panum's fusion area of 10 min, that is, the maximum disparity size that humans can binocularly fuse without vergence eye movements [12], [13]. Beyond Panum's fusion area, a small vergence eye movement is required to perceive the binocular image clearly. Therefore, assuming that both the vergence and accommodation is fixed on the display plane, our proposed approach may be considered to be near optimal with respect to visual comfort on S3D displays.

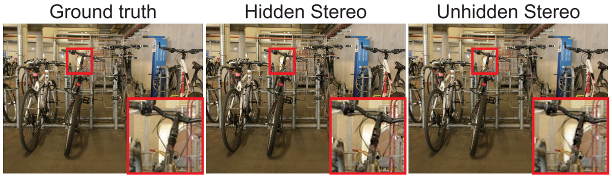

FIGURE 20. Comparison of synthesized images. Ground truth view is shown on the left. Both HS and ES images are generated based on the same per-pixel per-band disparity map. Stereo images generated by the Unhidden Stereo method sometimes exhibit artifacts due to mismatches in high-frequency-band disparity, as seen in the enlarged area. HS can suppress these artifacts by limiting the amount of disparity shift to $\pi/4$ for each band.

APPENDIX E
WHY DOES HIDDEN STEREO EXHIBIT BETTER 3D IMAGE QUALITY THAN UNHIDDEN STEREO?

Throughout the two user studies, we consistently found that HS could provide better 3D image quality than the UHS method. As discussed in the main paper, we think that two factors can account for this result. First, stereo images generated by HS tended to be slightly higher in contrast than the original input due to its additive nature. It is possible that the participants preferred higher contrast images and saw them as having "better" image quality. Second, artifacts produced by the stereo image synthesis process were more visible in the UHS-based methods (Fig. 20). Although the per-band disparity representation used in both HS and UHS can better represent disparities in complex scenes, its quality is not perfect and less precise in high frequency bands due to mismatches in the disparity refinement process. HS effectively suppresses these artifacts produced by imprecise disparity estimates by limiting the amount of disparity shift to $\pi/4$ for each band. Therefore, our results suggest that unless large viewpoint shifts are required, it is better to focus on providing the disparity cues that are the least necessary for the HVS to perceive depth.

REFERENCES

[1] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540–555, Apr. 2011.

[2] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H. Seidel, "A perceptual model for disparity," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–10, Jul. 2011.

[3] M. Lambooij, M. Fortuin, I. Heynderickx, and W. Ijsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, pp. 30201-1–30201-14, 2009.

[4] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *J. Vis.*, vol. 11, no. 8, p. 11, Aug. 2011.

[5] M. Fisker, K. Gram, K. K. Thomsen, D. Vasilarou, and A. Kraus, "Automatic convergence adjustment for stereoscopy using eye tracking," in *EG 2013—Posters*, 2013, pp. 23–24.

[6] M. Bernhard, C. Dell'mour, M. Hecher, E. Stavrakis, and M. Wimmer, "The effects of fast disparity adjustment in gaze-controlled stereoscopic applications," in *Proc. Symp. Eye Tracking Res. Appl.*, New York, NY, USA, Mar. 2014, pp. 111–118.

[7] P. Hanhart and T. Ebrahimi, "Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience," *Proc. SPIE*, vol. 9011, Mar. 2014, Art. no. 90110D.

[8] P. Kellnhofer, P. Didyk, K. Myszkowski, M. M. Hefeeda, H.-P. Seidel, and W. Matusik, "GazeStereo3D: Seamless disparity manipulations," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–13, Jul. 2016.

[9] T. Fukiage, T. Kawabe, and S. Nishida, "Hiding of phase-based stereo disparity for ghost-free viewing without glasses," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–17, Jul. 2017.

[10] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 3. Washington, DC, USA, Oct. 1995, pp. 444–447.

[11] B. Lee and B. Rogers, "Disparity modulation sensitivity for narrow-band-filtered stereograms," *Vis. Res.*, vol. 37, no. 13, pp. 1769–1777, Jul. 1997.

[12] I. P. Howard, *Perceiving in Depth Volume 1 Basic Mechanisms*, vol. 1. London, U.K.: Oxford Univ. Press, 2012.

[13] C. Schor, I. Wood, and J. Ogawa, "Binocular sensory fusion is limited by spatial resolution," *Vis. Res.*, vol. 24, no. 7, pp. 661–665, 1984.

[14] Y. Liu, A. C. Bovik, and L. K. Cormack, "Disparity statistics in natural scenes," *J. Vis.*, vol. 8, no. 11, pp. 1–14, Aug. 2008.

[15] E. Peli, T. R. Hedges, J. Tang, and D. Landmann, "A binocular stereoscopic display system with coupled convergence and accommodation demands," in *Proc. SID Int. Symp.*, 2001, vol. 32, no. 1, pp. 1296–1299.

[16] G. Maiello, M. Chessa, F. Solari, and P. J. Bex, "Simulated disparity and peripheral blur interact during binocular fusion," *J. Vis.*, vol. 14, no. 8, p. 13, Jul. 2014.

[17] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Trans. Graph.*, vol. 29, no. 4, pp. 1–10, Jul. 2010.

[18] B. Masia, G. Wetzstein, C. Aliaga, R. Raskar, and D. Gutierrez, "Display adaptive 3D content remapping," *Comput. Graph.*, vol. 37, no. 8, pp. 983–996, Dec. 2013.

[19] P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "Adaptive image-space stereo view synthesis," in *Proc. VMV*, 2010, pp. 299–306.

[20] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand, and A. Matusik, "Joint view expansion and filtering for automultiscopic 3D displays," *ACM Trans. Graph.*, vol. 32, no. 6, p. 221, Nov. 2013.

[21] P. Kellnhofer, P. Didyk, S.-P. Wang, P. Sitthi-Amorn, W. Freeman, F. Durand, and W. Matusik, "3DTV at home: Eulerian-Lagrangian stereo-to-multiview conversion," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–13, Jul. 2017.

[22] R. Hovden, Y. Jiang, H. L. Xin, and L. F. Kourkoutis, "Periodic artifact reduction in Fourier transforms of full field atomic resolution images," *Microsc. Microanal.*, vol. 21, no. 2, pp. 436–441, Apr. 2015.

[23] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, "Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors," *Science*, vol. 249, no. 4972, pp. 1037–1041, Aug. 1990.

[24] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.

[25] *Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide Screen 16:9 Aspect Ratios*, Standard BT.601-7, ITU-R Recommendations, 2011.

[26] S. C. Rawlings and T. Shipley, "Stereoscopic acuity and horizontal angular distance from fixation," *J. Opt. Soc. Amer.*, vol. 59, no. 8, pp. 991–993, Aug. 1969.

[27] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a quality metric for dense light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 58–67.

[28] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *Pattern Recognition* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2014, pp. 31–42.

[29] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[30] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *Computer Vision—ACCV 2016*. Cham, Switzerland: Springer, 2017, pp. 19–34.

[31] E. Arabadzhiyska, O. T. Tursun, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Saccade landing position prediction for gaze-contingent rendering," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–12, Jul. 2017.

[32] *Methodologies for the Subjective Assessment of the Quality of Television Images*, Standard BT.500-14, ITU-R Recommendations, 2019.

[33] M. Siegel and S. Nagata, "Just enough reality: Comfortable 3-D viewing via microstereopsis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 387–396, Apr. 2000.

[34] R. S. Allison and L. M. Wilcox, "Perceptual tolerance to stereoscopic 3D image distortion," *ACM Trans. Appl. Perception*, vol. 12, no. 3, pp. 1–20, Jul. 2015.

TAIKI FUKIAGE received the Ph.D. degree in interdisciplinary information studies from The University of Tokyo, in 2015. He joined NTT Communication Science Laboratories, in 2015, where he studies media technologies based on scientific knowledge about visual perception. He is currently a Distinguished Researcher at the Human Information Science Laboratory, Sensory Representation Group, NTT Communication Science Laboratories. He is a member of the Vision Sciences Society and the Vision Society of Japan.

SHIN'YA NISHIDA is currently a Professor at the Graduate School of Informatics, Kyoto University, and a Visiting Senior Distinguished Scientist at NTT Communication Science Laboratories. He is an Honorary Professor of Nottingham University, U.K. He is an expert in human vision (e.g., motion perception, material perception), haptics, multisensory integration, and visual media technology. He is a Council Member of the Science Council of Japan and a Project Leader of Deep SHITSUKAN. He is on the Editorial Boards of *Journal of Vision* and *Multisensory Research*, and on the Board of Directors of the Vision Sciences Society.