

# Hiding of Phase-Based Stereo Disparity for Ghost-Free Viewing Without Glasses

TAIKI FUKIAGE, TAKAHIRO KAWABE, and SHIN'YA NISHIDA, NTT Communication Science Laboratories

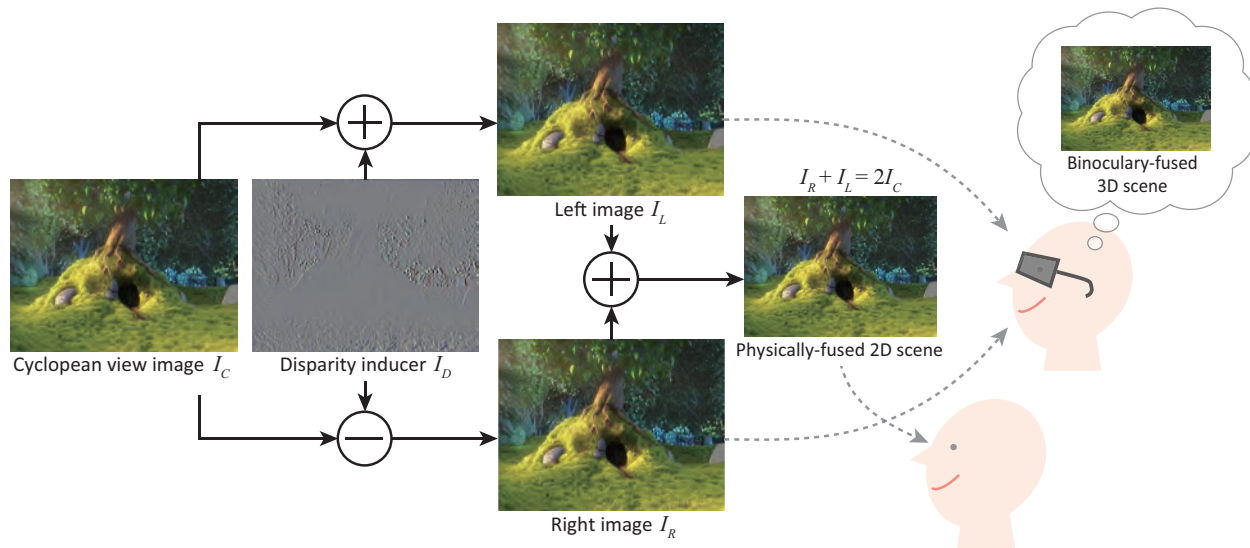


Fig. 1. A schematic of how Hidden Stereo cancels stereo ghosts for viewers without stereo glasses. Hidden Stereo generates a stereo pair by adding disparity-inducer patterns to a 2D image. The disparity-inducer patterns are identical except for the contrast polarity. Physical fusion of the stereo pair cancels out the disparity-inducer components, which makes only the original 2D image visible to viewers without stereo glasses. (The input image is taken from "Big Buck Bunny" ©by Blender Foundation, [www.bigbuckbunny.org](http://www.bigbuckbunny.org))

When a conventional stereoscopic display is viewed without stereo glasses, image blurs, or 'ghosts', are visible due to the fusion of stereo image pairs. This artifact severely degrades 2D image quality, making it difficult to simultaneously present clear 2D and 3D contents. To overcome this limitation (backward incompatibility), here we propose a novel method to synthesize ghost-free stereoscopic images. Our method gives binocular disparity to a 2D image, and drives human binocular disparity detectors, by the addition of a quadrature-phase pattern that induces spatial subband phase shifts. The disparity-inducer patterns added to the left and right images are identical except for the contrast polarity. Physical fusion of the two images cancels out the disparity-inducer components and makes only the original 2D pattern visible to viewers without glasses. Unlike previous solutions, our method perfectly excludes stereo ghosts without using special hardware. A simple algorithm can transform 3D contents from the conventional stereo format into ours. Furthermore, our method can alter the depth impression of a real object without its being noticed by naked-eye viewers by means of light projection of the disparity-inducer components onto the object's surface. Psychophysical evaluations have confirmed the practical utility of our method.

All authors have equal contributions. This work is supported by Grants-in-Aid for Scientific Research on Innovative Areas (15H05915) from JSPS/MEXT Japan. Author's addresses: T. Fukiage and T. Kawabe and S. Nishida, NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa, 243-0198, Japan.

© 2017 ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Graphics*, <https://doi.org/http://dx.doi.org/10.1145/3072959.3073672>.

CCS Concepts: • **Computing methodologies** → **Perception**; **Image processing**;

Additional Key Words and Phrases: stereoscopy, spatial phase shift, backward compatible, perception

## ACM Reference format:

Taiki Fukiage, Takahiro Kawabe, and Shin'ya Nishida. 2017. Hiding of Phase-Based Stereo Disparity for Ghost-Free Viewing Without Glasses. *ACM Trans. Graph.* 36, 4, Article 147 (July 2017), 17 pages. DOI: <http://dx.doi.org/10.1145/3072959.3073672>

## 1 INTRODUCTION

In standard 3D stereoscopic displays, an image for the left eye is superimposed on that for the right eye. For the viewers wearing shutter or polarized stereo glasses, the left and right images are separately presented to the corresponding eyes, and viewers with good stereopsis can enjoy the 3D contents with natural vergence angles. For viewers who unfortunately watch the display without stereo glasses, however, the superposition of the left and right images produces uncomfortable image blurs and bleedings, which are called 'ghosts' in this paper. Stereoscopic ghosts, caused by the binocular disparity between the stereo image pair, seriously limit the utility of the standard 3D display. In particular, ghosts make the display unsuitable in situations where the same screen is viewed by a heterogeneous group of people, including those who are not supplied with stereo glasses or who have poor stereopsis. To overcome

this limitation, which is referred to as the backward-compatibility problem [Didyk et al. 2011], several attempts have been made to make displays that simultaneously present proper 3D contents to viewers with glasses and ghost-free 2D contents to viewers without them [Didyk et al. 2011, 2012; Fujimura et al. 2012; Lang et al. 2010; Scher et al. 2013]. The previous solutions, however, have considerable shortcomings with regard to image quality and/or compatibility with standard hardware (see Related work).

Here we propose a novel method to synthesize completely ghost-free stereoscopic images. Our technique, named Hidden Stereo, gives binocular disparity to a 2D image by the addition of a quadrature-phase pattern that induces spatial subband phase shifts (see Figure 1). The patterns added to the left and right images, which we call disparity inducers, are identical except that they are opposite in contrast polarity. Physical fusion of the two images cancels out the disparity-inducer components and brings the image back to the original 2D pattern. This is how binocular disparity is hidden to viewers without glasses but remains visible to viewers with them. Although the synthesized stereo images are slightly different from those obtained in the real 3D world, scientific knowledge about human stereopsis and binocular fusion suggests that Hidden Stereo can properly drive the viewer's binocular vision. In addition to our basic algorithm that synthesizes a ghost-free stereo pair from a 2D image and its disparity map, we will also describe a simple way to transform images from a conventional stereo pair to ours without explicitly estimating the disparity map. Since our technique excludes stereo ghosts at the stage of stereo image synthesis, it is perfectly compatible with standard stereo display hardware. In addition, since our technique is based on the addition of disparity inducers, when combined with spatial augmented reality (projection mapping), it can add to real objects such depth structures that are visible only to viewers with glasses.

In exchange for the ability to present perfect ghost-free 2D images to viewers without glasses, Hidden Stereo has a slightly limited ability to present 3D images compared with the conventional position-shift method. Specifically, there is an upper limit in the disparity magnitude given to the image. Adding a large disparity could affect the dynamic range and induce perceptual artifacts. We will discuss how to practically overcome these limitations and demonstrate the utility of our method by psychophysical evaluation experiments.

In the next section, we briefly describe previous attempts to overcome stereo ghosts. In the third section, we explain the algorithms of our proposed method. In the fourth section, we evaluate the performance of our algorithm. In the fifth section, we discuss the utilities and limitations of our method. In the Appendix, we briefly review the mechanisms of human binocular vision related to our method.

## 2 RELATED WORKS

### 2.1 Stereoscopic methods producing ghosts

When conventional stereoscopic images are presented with standard 3D displays with a temporal multiplexing system or with a polarized 3D system, viewers without glasses see ghosts. With

the temporal multiplexing system, stereoscopic images are alternatively presented, but as long as the alternation rate is high enough, they are perceptually mixed due to a sluggish early visual response [de Lange Dzn 1958] and trajectory-dependent signal integration [Terao et al. 2010; Watanabe and Nishida 2007] of the human visual system (HVS). With the polarized system such as dual-projector polarized displays used in 3D theaters, stereoscopic images are spatiotemporally mixed and become inseparable for polarity-insensitive human viewers. Hidden Stereo is able to remove ghosts in both cases.

### 2.2 Image cancellation

A previous solution to simultaneously present 3D contents and ghost-free 2D contents with a single display is to make a special image combination that cancels out, say, the right (R) image component and leaves only the left (L) image.

The 2x3D display [Fujimura et al. 2012] presents an L-R image with one projector and an R image with another projector. Without glasses, viewers see the superimposed image, which is  $L=(L-R)+R$ . With glasses, they see the superimposed L image with the naked left eye and only the R image with the right eye through a polarized glass. This method, unlike ours, can present arbitrary images to the left and right eyes. However, it significantly sacrifices the dynamic range and makes the L image much brighter in intensity and much lower in contrast than the R image. In addition, since the images from the two projectors differ from an ordinary stereo image pair, this method is incompatible with the standard stereo display system.

3D+2D TV [Scher et al. 2013] uses a third image to perceptually cancel out one of the stereo pair. In addition to separate L and R images in each frame, a third image, invisible through glasses, is added. In the combined view seen by viewers without glasses, this method cancels the R image, leaving only the L image. Like the 2x3D display, the problem is low contrast of the resulting 2D (L) image, since the cancellation of the R component inevitably adds a uniform gray field. The L contrast can be increased by darkening the R image, but this could cause another artifact caused by the Pulfrich effect [Pulfrich 1922], in which the difference in the luminance level of a moving stimuli is interpreted as binocular disparity by the HVS [Qian and Freeman 2009; Read and Cumming 2005]. 3D+2D TV needs a special display and glasses.

### 2.3 Depth compression

With the standard stereo system, an effective way to attenuate ghosts is to compress the magnitude of binocular disparity. Microstereopsis is a conventional approach in this direction, which reduces the distance between stereo cameras shooting a real 3D scene [Siegel and Nagata 2000]. More recent studies use the depth image based rendering technique (DIBR), which gives binocular disparities to an image by shifting pixels on the basis of the associated depth map [Mark et al. 1997; Zitnick et al. 2004]. Several attempts have been made to compress the depth map while keeping perceived depth as intact as possible [Kellnhofer et al. 2016; Lang et al. 2010].

Depth compression by itself is not a method for perfectly removing ghosts, but backward compatible stereo [Didyk et al. 2011,

2012] attempts to render ghosts as invisible as possible to human viewers. Based on a perceptual model for binocular disparity, this technique compresses disparity while keeping a specified minimum magnitude of perceived disparity. Compression magnitude is spatial-frequency dependent. Considering the low sensitivity of the HVS to very low- and high-frequency components of disparity modulation, the backward compatible stereo selectively reduces these components more than the middle-frequency components. Justification of the low-frequency reduction came from the Craik-O'Brien-Cornsweet illusion in the depth dimension [Anstis et al. 1978; Brookes and Stevens 1989], similar to the illusion in the brightness domain [Cornsweet 2012; Kingdom and Moulden 1988; Purves et al. 1999]. The vision-based depth compression, adopted by the backward compatible stereo, is potentially useful for many purposes, including our Hidden Stereo technique.

Instead of using DIBR, Didyk et al. [2013] used an image-based technique to render multi-view images for autostereoscopic 3D displays from a single stereo pair. They suggested that the same algorithm can be used for simple disparity manipulations of a stereo pair. The basic algorithm they used is a phase-based technique originally developed to magnify invisible motions [Wadhwa et al. 2013]. Our Hidden Stereo also manipulates binocular disparity based on image-based phase shifts, but our algorithm is different from theirs: our algorithm is related more to Eulerian video magnification [Wu et al. 2012].

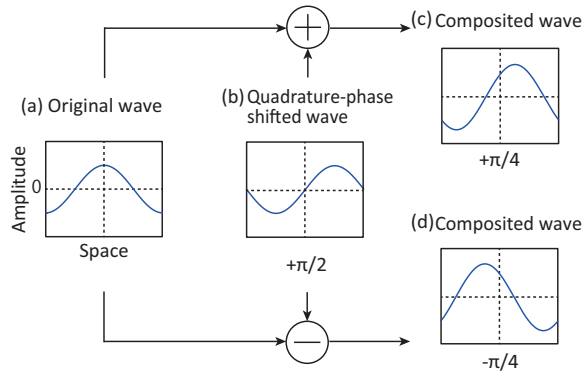


Fig. 2. The basic mechanism to produce a relative phase-shift in Hidden Stereo. Here the image is assumed to be a vertical grating whose intensity profile is a sine wave. (a) An original wave. (b) A wave shifted rightwards in a quadrature phase ( $\pi/2$ ). (c) Addition of (b) to (a) results in a wave whose phase shift from the original wave is  $+\pi/4$ . (d) Subtraction of (b) from (a) results in a wave whose phase shift is  $-\pi/4$ . Considering (c) and (d) as a stereo pair, there is a relative phase shift (binocular disparity) of  $\pi/2$ . Fusion of the pair [addition of (c) and (d)] cancels out the effects of (b) and makes the image profile the same as (a) except for amplitude scale factors. The shift size in (c) and (d) can be controlled by changing the amplitude of (b).

### 3 PROPOSED METHOD

Figure 1 shows a schematic of how our method cancels stereo ghosts for viewers without glasses while providing disparity for viewers with glasses. The key concept of our method is that we generate

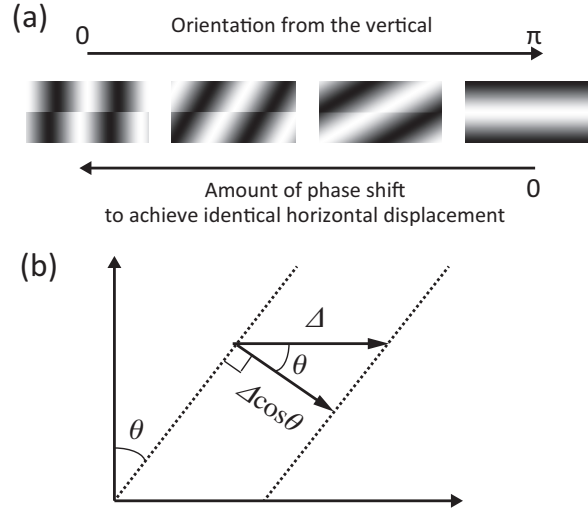


Fig. 3. The relationship between phase-shift size and orientation. (a) The amount of phase shift needed to achieve the identical horizontal displacement decreases as the orientation becomes closer to the horizontal. (b) Provided that phase of a vertical grating has to be shifted by  $\Delta$ , a grating oriented by  $\theta$  has to be shifted by  $\Delta \cos \theta$  to achieve the same horizontal displacement.

a stereo image pair such that the linear sum of the stereo images becomes equivalent to an image from the cyclopean view i.e., an image viewed by a virtual eye placed midway between the left and right eye positions. (In this paper, we use the term 'cyclopean' to refer to this physical image, as well as to a perceptual image created in the viewer's brain by fusing two stereo images.) Left and right stereo images ( $I_L$  and  $I_R$ ) are synthesized by linearly adding a disparity inducer ( $I_D$ ) and the contrast-inverted version of the same disparity inducer ( $-I_D$ ), respectively, to a cyclopean-view image  $I_C$  as

$$I_L = I_C + I_D \quad (1)$$

$$I_R = I_C - I_D. \quad (2)$$

Consequently, viewers without glasses can observe the original cyclopean-view image as  $I_L + I_R = 2I_C$ .

On the other hand, viewers with glasses can enjoy a binocularly fused image with depth impression induced by the disparity inducer. The binocularly fused image seen by the viewers (perceptual cyclopean image) should be close to the original image as long as perceptual binocular integration can be approximated by a linear summation of the left and right images. See Appendix A.2 for the relevant mechanisms of human binocular fusion.

The major difference between our method and previous cancellation methods [Fujimura et al. 2012; Scher et al. 2013] is that only the disparity inducer embedded in a stereo image pair is cancelled out in our case, while one of the images of the stereo image pair is entirely cancelled out in the previous cases. Thus, in our method, the reduction in the dynamic range of stereo images is small. In addition, unlike the other previous methods, our method has perfect

compatibility with conventional stereo presentation systems using shutter/polarization glasses. In the following subsections, we first describe the mathematical formulations underlying our method. Then, we show how we generate the disparity inducer in general image cases.

### 3.1 Disparity manipulation by adding a quadrature-phase component

Our method synthesizes a stereo image by linearly adding a disparity inducer. To realize this process, we manipulate binocular disparity based on a local phase shift. Unlike a similar previous approach [Didyk et al. 2013], we do not directly change the phase, but shift it by adding a quadrature-phase ( $\pi/2$ ) shifted component (Figure 2), since the linearity of the operation is essential to our purpose.

For the sake of simplicity, let us begin with a simple 1D case where we want to displace a vertical grating whose intensity profile is a sinusoidal wave with frequency  $\omega$  by the size of disparity,  $d$ . When a quadrature-phase-shifted component is added to the original wave, the resultant composited wave is obtained as follows:

$$\sin \omega x + A \sin \left( \omega x + \frac{\pi}{2} \right) = \sqrt{1 + A^2} \sin (\omega x + \phi), \quad (3)$$

where  $\tan \phi = A$ . Thus, we can control the amount of the phase shift of the composited wave by adjusting the amplitude of the quadrature-phase-shifted wave,  $A$ .

Importantly, we can generate the second composite wave that is displaced by the same amount in the opposite direction just by subtracting the same quadrature-phase-shifted wave from the original one:

$$\sin \omega x - A \sin \left( \omega x + \frac{\pi}{2} \right) = \sqrt{1 + A^2} \sin (\omega x - \phi). \quad (4)$$

Thus, to obtain the disparity  $d$ , we need to shift the phase of the original wave for each direction in such a way that the size of each phase shift  $\phi$  becomes equivalent to  $d/2$  in the horizontal displacement size. Since the size of the phase shift increases as spatial frequency increases,

$$\phi = \frac{\omega d}{2}. \quad (5)$$

Consequently, the amplitude of the quadrature-phase-shifted wave (or a disparity inducer for the 1D sinusoidal wave) is determined to be:

$$A = \tan \frac{\omega d}{2}. \quad (6)$$

Next, let us consider a 2D case where the target pattern is a sinusoidal grating with frequency  $\omega$ , which is oriented  $\theta$  from the vertical. In this case, the amount of phase shift needed to achieve the identical horizontal displacement  $d$  decreases as the orientation  $\theta$  becomes closer to the horizontal (see Figure 3). Therefore, the amplitude  $A$  of the quadrature-phase-shifted grating needed to achieve the horizontal disparity  $d$  can be rewritten as:

$$A = \tan \frac{\omega d |\cos \theta|}{2}, \quad (7)$$

where the range of  $\theta$  is  $[0, \pi]$ .

Obviously, the disparity range that this approach can achieve is limited. In theory, the maximum possible displacement is less than a half of the wavelength, which means that the higher the spatial frequency of the target pattern is, the less disparity our method can produce for that pattern. In Section 3.2.2, we describe how we handle this limitation in practical cases.

### 3.2 Extension to general image cases

In general, any arbitrary image can be represented as a combination of sinusoids by a 2D Fourier transform. Thus, the disparity manipulation described above can be easily applied to arbitrary images as long as the disparity for each sinusoidal component is spatially uniform. However, when we want to add a structured map of binocular disparity, we need spatially localized information about the spatial frequency and orientation of an input image. To analyze this local structure information, we use the steerable pyramid [Portila and Simoncelli 2000; Simoncelli and Freeman 1995], which decomposes an image into subband images, each representing local responses tuned to different spatial frequency and orientation bands. Specifically, we use the complex steerable pyramid to obtain an analytic version of the filter responses, whose imaginary counterpart represents responses for a quadrature-phase shifted version of an input image. It should be noted that the HVS analyzes binocular disparity using a bank of multi-scale, orientation-tuned sensors similar to the steerable pyramid (see Appendix A.2).

**3.2.1 Disparity manipulation.** Figure 4 shows an overview of the process for generating a Hidden Stereo pair from a cyclopean-view image. Here we assume that a disparity map  $D(\mathbf{x})$  and a cyclopean view image  $I_C(\mathbf{x})$  are given.  $D(\mathbf{x})$  represents the disparity size for each spatial location  $\mathbf{x}$ , with positive and negative values indicating crossed and uncrossed disparity, respectively. We first build a complex steerable pyramid of  $I_C(\mathbf{x})$  by applying a series of filters  $\Psi$  to the discrete Fourier transform of  $I_C(\mathbf{x})$  as

$$\tilde{S}_{C,i,j} = \tilde{I}_C \Psi_{i,j}, \quad (8)$$

where  $\tilde{I}_C$  is the discrete Fourier transform of  $I_C$ , and  $\Psi_{i,j}$  is a steerable filter tuned to the  $i$ th frequency and  $j$ th orientation band. Taking the imaginary part from the resulting complex steerable pyramid  $S_C$ , we get quadrature-phase shifted responses of the pyramid  $S_{C'}$ :

$$S_{C',i,j}(\mathbf{x}) = \begin{cases} \text{Im}[S_{C,i,j}(\mathbf{x})] & \text{if } \theta_j \leq \frac{\pi}{2} \\ -\text{Im}[S_{C,i,j}(\mathbf{x})] & \text{if } \theta_j > \frac{\pi}{2} \end{cases}. \quad (9)$$

Here,  $\theta_j$  denotes the peak orientation of the  $j$ th orientation band and ranges from 0 to  $\pi$  (angle from the vertical). To obtain responses consistently shifted in the same direction, we have to switch the sign of the imaginary responses at the horizontal angle as in Eq. 9. Alternatively, one can more directly obtain the imaginary filter responses by decomposing the sine-phase filters of the complex steerable pyramid.

Then, we multiply each of the subbands  $S_{C'}$  by the weight function  $A$ :

$$\hat{S}_{C',i,j}(\mathbf{x}) = A_{i,j}(\mathbf{x}) S_{C',i,j}(\mathbf{x}), \quad (10)$$

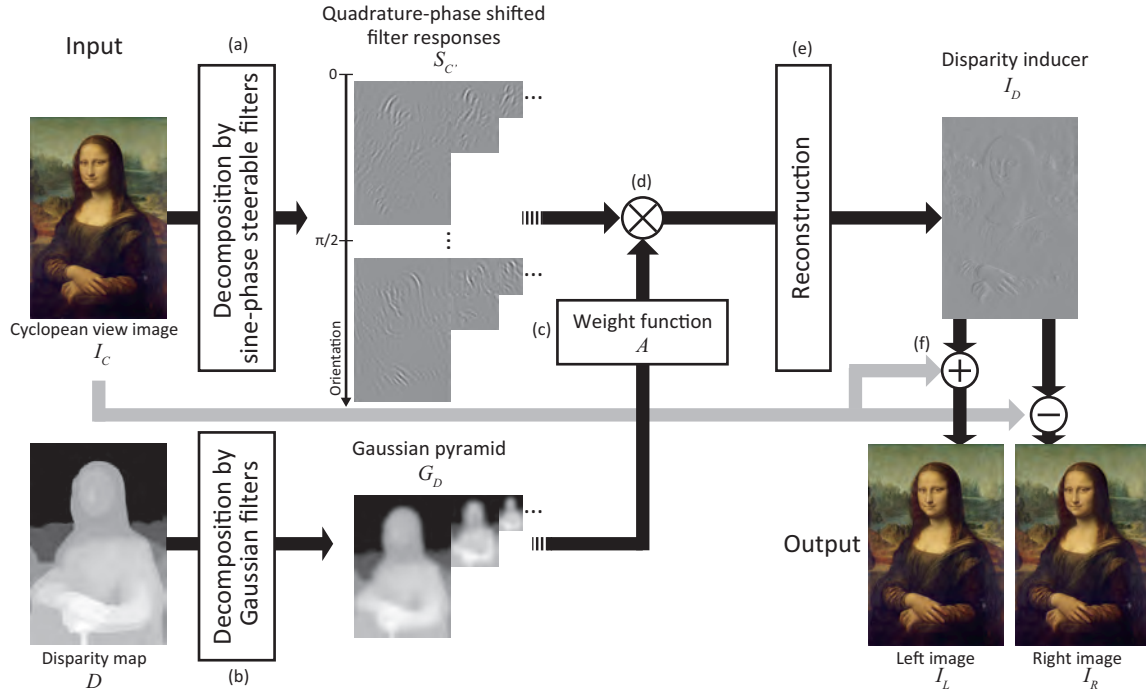


Fig. 4. Process overview to generate a Hidden Stereo pair from a cyclopean view image (*Mona Lisa* by Leonardo da Vinci) and a disparity map. (a) The system first decomposes the input cyclopean-view image  $I_C$  into the steerable pyramid and obtains quadrature-phase shifted version of filter responses,  $S_{C'}$ . (b) The input disparity map  $D$  is decomposed into Gaussian pyramid,  $G_D$ . (c) The weight  $A$  for each of the quadrature-phase responses is computed based on a disparity size in the Gaussian pyramid (Eq. 11). (d) Each of the quadrature-phase responses is then weighted by  $A$ . (e) The disparity inducer  $I_D$  is obtained by reconstructing the weighted filter responses. (f) Finally, the left and right stereo images ( $I_L$  and  $I_R$ ) are generated by the addition/ subtraction of the disparity inducer to/from the input cyclopean-view image, respectively.

where

$$A_{i,j}(\mathbf{x}) = \tan \frac{G_{Di}(\mathbf{x})\omega_i |\cos \theta_j|}{2}. \quad (11)$$

Here,  $G_{Di}(\mathbf{x})$  denotes a Gaussian pyramid of the disparity map  $D(\mathbf{x})$ , which contains spatial frequency bands less than and equal to the  $i$ th level of the steerable pyramid.  $\omega_i$  denotes the peak frequency of the  $i$ th frequency band. The reason to use a blurred disparity map,  $G_D$ , instead of  $D$  is that the spatial map of disparity cannot contain information finer than the subband of the intensity distribution used to compute the disparity map from binocular correlations [Banks et al. 2004].

Finally, the disparity inducer  $I_D(\mathbf{x})$  is obtained by reconstructing the weighted quadrature-phase pyramid  $\hat{S}_{C'}$  as

$$\tilde{I}_D = \sum_{i,j} \Psi_{i,j} \hat{S}_{C',j}, \quad (12)$$

where  $\tilde{I}_D$  and  $\hat{S}_{C'}$  are the discrete Fourier transforms of  $I_D$  and  $\hat{S}_{C'}$ , respectively.

The disparity inducer can be computed separately for RGB color channels. Alternatively, considering the minor role of color information in human stereopsis [Howard 2012a; Kingdom and Simmons 1996], one can compute the disparity inducer only in the luminance channel.

**3.2.2 Maximum disparity bound.** In theory, the maximum displacement that can be achieved by our method is less than a half of the wavelength. However, the amplitude of the disparity inducer becomes infinitely large near this limit, which means that the contrast of the original pattern (including noises) is significantly amplified as the disparity comes close to this limit. Therefore, we practically limit the maximum phase shift after composition to  $\pm\pi/4$ . Under this limit, the disparity bound can be described as follows:

$$|d| < \frac{\pi}{2\omega_i |\cos \theta_j|} \quad (13)$$

To give a disparity map consistent across subband filters at each location, one has to compress the entire disparity map such that the displacement size of the highest frequency vertical component satisfies Eq. 13. However, this is a rather conservative approach given that the HVS has independent disparity detection mechanisms, each tuned to various ranges of orientations and spatial frequencies (see Appendix A). Therefore, we decided to independently clip each level of the Gaussian pyramid  $G_D$  according to Eq. 13. Namely, we substitute the following  $\hat{G}_{Di,j}$  instead of  $G_{Di}$  into Eq. 11.

$$\hat{G}_{Di,j} = \begin{cases} \frac{\pi}{2\omega_i |\cos \theta_j|} & \text{if } G_{Di} > \frac{\pi}{2\omega_i |\cos \theta_j|} \\ -\frac{\pi}{2\omega_i |\cos \theta_j|} & \text{if } G_{Di} < -\frac{\pi}{2\omega_i |\cos \theta_j|} \\ G_{Di} & \text{otherwise} \end{cases} \quad (14)$$

When the disparity size is different among orientations and/or scales, reconstructed results could suffer from depth underestimation and visible artifacts. Nevertheless, as we will show in Section 4.2, our method can produce perceptually pleasing results even in these cases as long as the disparity size is not very large. [Wu et al. 2012] has proposed a similar solution for Eulerian video magnification.

**3.2.3 Handling dynamic range reduction.** Since we limit the maximum disparity such that the weights of the quadrature-phase subbands (i.e.,  $A$ ) do not exceed one, the contrast of the disparity inducer ( $I_D$ ) does not become significantly high. That said, if a large disparity is added on high-contrast regions in the original image ( $I_C$ ), the intensities of the resulting stereo images ( $I_L$ ,  $I_R$ ) often exceed the original dynamic range. For retaining the entire intensity variation under the display's dynamic range, a straightforward method is to linearly compress stereo images  $I_L$  and  $I_R$ . However, this inevitably lowers the contrast of the original image contents. Moreover, if the stereo content is a sequence of image frames, all of the frames have to be compressed based on all of their maximum/minimum intensity values to avoid visual flicker.

As a simple solution to this issue, in this work, we clip the disparity inducer wherever the intensities in the resulting stereo images exceed predefined bounds  $[b_l, b_u]$ . Namely, we first calculate the maximum deviation from the upper and the lower bounds as

$$M_U(\mathbf{x}) = \max(\max(I_R(\mathbf{x}) - b_u, 0), \max(I_L(\mathbf{x}) - b_u, 0)), \quad (15)$$

$$M_L(\mathbf{x}) = \min(\min(I_R(\mathbf{x}) - b_l, 0), \min(I_L(\mathbf{x}) - b_l, 0)). \quad (16)$$

Then, we subtract these deviated values from the disparity inducer:

$$\hat{I}_D(\mathbf{x}) = I_D(\mathbf{x}) - M_U(\mathbf{x}) - M_L(\mathbf{x}). \quad (17)$$

Finally, we obtain a modified stereo image pair by

$$\hat{I}_L(\mathbf{x}) = I_C(\mathbf{x}) + \hat{I}_D(\mathbf{x}), \quad (18)$$

$$\hat{I}_R(\mathbf{x}) = I_C(\mathbf{x}) - \hat{I}_D(\mathbf{x}). \quad (19)$$

The resulting stereo pair is assured to be within the range  $[b_l, b_u]$ . Although this operation sacrifices the disparity size of the clipped areas, the influence on the perceived depth quality is usually minor or even unnoticeable since the clipped areas are mostly overlapped with the high-frequency component where the lower-frequency (and/or diagonal) component can support the original disparity size. The evaluation in the user study (Section 4.3) also demonstrated that clipping disparity inducer does not cause any practical problems at least in the cases we tested.

**3.2.4 Transforming existing stereo contents.** There are huge amounts of 3D content in the conventional stereo format. The most straightforward way to transform them into our ghost-free format, is to compute a disparity map  $D(\mathbf{x})$  from a given stereo pair.

Then, one can generate a disparity inducer in the same way as described above using one of the stereo images as  $I_C(\mathbf{x})$ . However, here we present a more concise way, in which we compute disparity as phase difference between a given stereo pair and calculate the weight for the quadrature-phase shifted image directly from the phase difference.

Figure 5 shows an overview of the process for transforming a standard stereo pair into the Hidden Stereo format. Let  $I_L'$  and  $I_R'$  represent an input stereo pair. Since we use the left image as a cyclopean view image,  $I_C = I_L'$ . First, we compute the phase difference between  $I_L'$  and  $I_R'$ , following the approach presented in [Didyk et al. 2013]. We decompose an input stereo pair into complex steerable pyramids, giving  $S_{L',i,j}(\mathbf{x})$  and  $S_{R',i,j}(\mathbf{x})$ . Then, we compute the phase difference  $\delta_{i,j}(\mathbf{x})$  between each of the corresponding complex coefficient pairs. As was done in [Didyk et al. 2013], we correct the phase difference to twice the phase difference in the level below as

$$\delta_{i,j}(\mathbf{x}) = 2\delta_{i-1,j}(\mathbf{x}) \quad (20)$$

whenever the absolute phase difference in the level below exceeds  $\pi/2$ .

A Hidden Stereo pair is obtained by shifting the phase of a cyclopean view image by  $\delta/2$  in each direction. However, phase difference  $\delta/2$  may be too large to be expressed in our method. In such cases, one may scale the phase difference by multiplying a scaling factor  $\alpha$  (or use a more elegant method of disparity compression). Then, each of the phase differences is clipped by the maximum disparity bounds as

$$\hat{\delta}_{i,j}(\mathbf{x}) = \begin{cases} \pi/4 & \text{if } \alpha\delta_{i,j}(\mathbf{x})/2 > \pi/4 \\ -\pi/4 & \text{if } \alpha\delta_{i,j}(\mathbf{x})/2 < -\pi/4 \\ \alpha\delta_{i,j}(\mathbf{x})/2 & \text{otherwise} \end{cases} \quad (21)$$

This time, the disparity bounds are directly defined in the phase angle.

Next, we generate a quadrature-phase-shifted version of the steerable pyramid of a cyclopean-view image,  $S_{C'}$ , by taking the imaginary part from  $S_L$ . Instead of reconstructing a cyclopean image from the two input images, we simply use the left input image as a cyclopean-view image. Then, we multiply each of the subbands  $S_{C'}$  by the following weight function  $A$ :

$$A_{i,j}(\mathbf{x}) = \tan\{\hat{\delta}_{i,j}(\mathbf{x})\}. \quad (22)$$

Finally, the disparity inducer  $I_D(\mathbf{x})$  is obtained by reconstructing the weighted phase-shifted subbands in the same way as described in Section 3.2.

## 4 EXPERIMENTS

In this section, we first describe the details about the implementation of our algorithm. Then, we report our psychophysical evaluations of Hidden Stereo with regard to apparent depth and 3D image quality for viewers with glasses and 2D image quality for viewers without glasses. We also present an example application to a light projection system that manipulates a real object's depth without degrading its appearance for viewers without glasses.

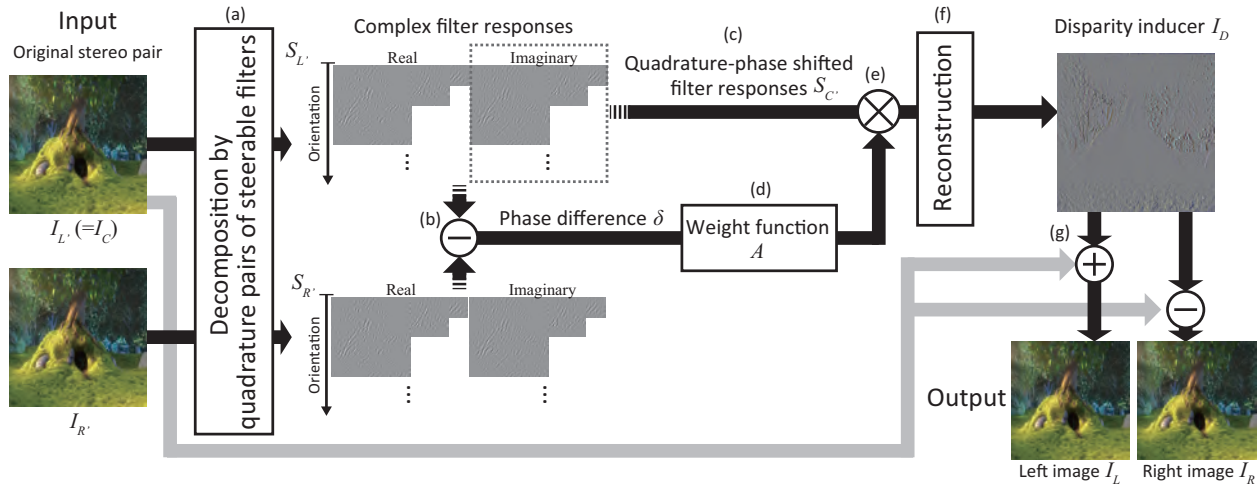


Fig. 5. Process overview to generate a Hidden Stereo pair from the conventional stereo format. Here, we show an example where we use the left input image  $I_L$  as a cyclopean-view image  $I_C$ . (a) The system first decomposes the input stereo pair into complex steerable pyramid representations. (b) Then, we compute the phase difference  $\delta$  between each of the corresponding complex coefficient pairs. (c) From the imaginary responses of the complex steerable pyramid of the left image, we obtain the quadrature-phase shifted responses of the cyclopean-view image. (d) The weight for each of the quadrature-phase responses is computed based on the phase difference  $\delta$  (Eq. 22). (e) Each of the quadrature-phase responses is then weighted by  $A$ . (f) The disparity inducer  $I_D$  is obtained by reconstructing the weighted filter responses. (g) Finally, the left and right stereo images ( $I_L$  and  $I_R$ ) are generated by the addition/ subtraction of the disparity inducer to/from the cyclopean-view image, respectively. (The input stereo pair is taken from "Big Buck Bunny" ©by Blender Foundation).

#### 4.1 Implementation details

We generated the results using a laptop computer with a quad-core CPU and 16 GB RAM. The algorithm was implemented in a non-optimized Python code and run in a single thread. We used eight orientations in the complex steerable pyramid, though just two or four orientations still give good results. To generate the results below, we computed the disparity inducer independently for each of the RGB channels. However, one can obtain perceptually comparable results by processing the luminance channel alone after converting an image from RGB to YUV (other formats such as XYZ or YIQ are also possible). We took this approach when manipulating depth by light projection (Section 4.4). Under this condition, our algorithm takes about 16 seconds per channel to transform a standard stereo image pair (960×540 pixels in resolution) into a Hidden Stereo pair. When a disparity map is given, the computation time is reduced to 11 seconds. Since the major performance overhead in our algorithm is in building and reconstructing the steerable pyramid, some other approaches [Didyk et al. 2013; Wadhwa et al. 2014] that have accelerated the same process would similarly benefit the performance of our method.

#### 4.2 Perceptual experiment 1: Depth magnitude

As described in Section 3.2.2, Hidden Stereo limits the maximum disparity independently for each subband. Therefore, the stereo images made by our method do not provide geometrically correct disparity, especially when an input disparity map contains large values. We assume that the disparity detection mechanisms in the HVS compute disparity in their preferred frequency and orientation range in parallel and resolve the inconsistency in the integration stage (see Appendix A). However, it is not known whether

the perceived depth magnitudes in such inconsistent conditions are comparable to those in consistent (geometrically correct) conditions.

Another possible concern in Hidden Stereo is that the addition of disparity inducers ( $I_D$  and  $-I_D$ ) often causes artifacts due to inconsistent phase shifts as well as inconsistent contrast amplification among the subbands of the resulting stereo images. If the image difference produced by the artifacts is large, the cyclopean image perceived by viewers with stereo glasses may be accompanied by binocular luster (see Appendix A.2), which could potentially contribute to the visual discomfort of the 3D display [Kooi and Toet 2004].

To address these issues, we conducted two perceptual experiments, one using simple and controlled stimuli for evaluation of the basic characteristics of our method (reported in this section), and the other using natural 3D scenes for evaluation of the practical utility of our method (reported in Section 4.3).

**Methods.** The first experiment estimated the perceived depth magnitude of Hidden Stereo as a function of the specified disparity by way of depth matching with standard stereo images, and then estimated the binocular luster impression by way of rating. A part of the images used in the experiment are shown in Figure 6. We used four natural texture images from [Olmos and Kingdom 2004] as cyclopean-view images. The reason we used natural textures (instead of more simple random-dot patterns) for evaluation of our method was that natural images have scale-invariant image statistics. Horizontal sinusoidal gratings were used as disparity maps. The vertical intensity profile of the disparity maps can be written as

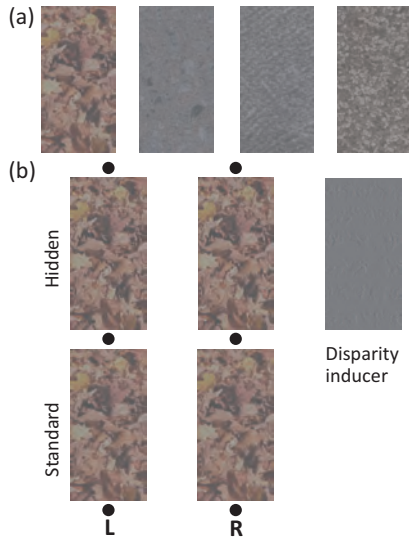


Fig. 6. The stimuli used in the perceptual evaluation in Section 4.2. (a) Four textures used in the experiment as a cyclopean-view image. The textures were taken from McGill Calibrated Colour Image Database [Olmos and Kingdom 2004]. (b) Examples of the stereo pairs presented in the experiment. A sinusoidal undulation was given to the textures by Hidden Stereo (referred to as "Hidden") and by pixel warping (referred to as "Standard"). The disparity magnitude in these examples is eight pixels. Here, the left and right images are shown side by side for parallel fusion. However, one can also view an equivalent depth impression by cross fusion (only the phase of the depth undulation is reversed). A disparity inducer embedded in the Hidden Stereo pair is shown on the right side.

$$D(y) = d \sin(2\pi f + \varphi), \quad (23)$$

where  $d$  denotes disparity magnitude and  $\varphi$  is a random constant. The spatial frequency  $f$  was 0.46 cpd (cycles per degree), which is close to the optimal frequency for disparity detection by the HVS [Bradshaw and Rogers 1999]. Since the sinusoidal depth map was irrelevant to the texture pattern, we could evaluate the effects of binocular disparity independent of other monocular depth cues (which would be difficult with natural scene pictures as we used in the next experiment). Another merit of the sinusoidal depth structure was that there was no ambiguity for the observers to specify the regions of maximum depth. There were five conditions in the disparity magnitude of Hidden Stereo images: 2, 4, 8, 16 and 32 pixels, with one pixel corresponding to 0.76 min in visual angle. The Hidden Stereo pairs were generated as described in Section 3.2.1. We limited the maximum disparity for each subband as mentioned in Section 3.2.2. To test the basic property of our method, we did not clip the disparity inducer as in Section 3.2.2, but linearly compressed the intensities of the left and right image based on the maximum/minimum intensity of all the stimuli such that they could be presented in the display's dynamic range. As a result, the texture patterns were presented at relatively low contrasts. The standard stereo pairs were generated by pixel-warping the cyclopean-view images with linear interpolation according to the disparity map.

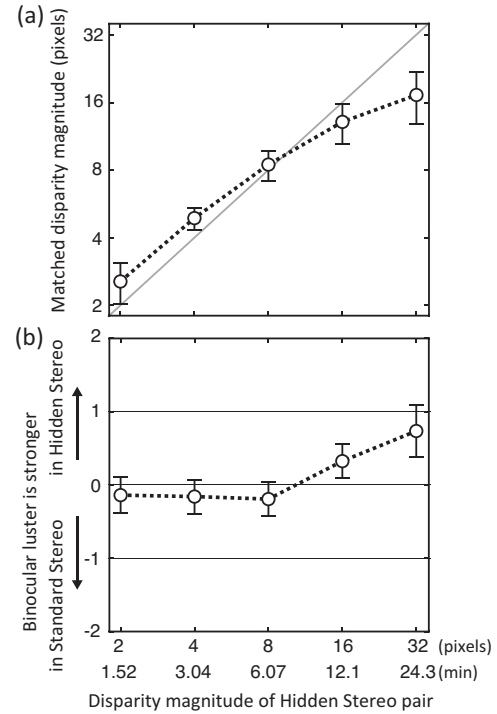


Fig. 7. The psychophysical data showing the effective disparity range of Hidden Stereo. The error bars show  $\pm 95\%$  confidence intervals. (a) The perceived depth magnitude of the Hidden Stereo image measured by matching the disparity magnitude with the standard stereo image. (b) Ratings of relative binocular luster impression as a function of the disparity magnitude of the Hidden Stereo image. The positive scores indicate that stronger luster was perceived in the Hidden Stereo image, and the negative scores indicate that stronger luster was perceived in the standard stereo image.

16 naïve human observers with normal visual acuity (11 females and 5 males, aged from 20 to 40) participated in the experiment. In each trial, an observer with stereo glasses viewed two images presented by a stereo projector (Vivitek Q7 Plus QUMI, 1280 x 720 pixels, 120Hz) on the left and right sides of the screen, where the left was the stereo pair generated by our method and the right was the standard stereo pair. Each image spanned 10.5×21.0 cm (256×512 pixels) on the screen. The viewing distance was 190 cm. During the experiment, the observer was asked to manipulate, by pressing keys, the disparity in the standard stereo image pair in such a way that the perceived depth magnitude of the standard stereo image matched that of the Hidden Stereo image. After being satisfied with their matching, the observers had to compare the impression of 'luster/glare' between the two images and rate the relative strength of the binocular luster on the following scale: -2 (the left image is obviously more lustrous), -1 (the left image is slightly more lustrous), 0 (almost the same), 1 (the right image is slightly more lustrous), 2 (the right image is obviously more lustrous). For the second task, the observers were instructed to check if there were any unnaturally lustrous or glared areas in the images, and compare their strength between the two images.

**Results.** The experimental results are shown in Figure 7. Figure 7a shows the perceived depth magnitude of the Hidden Stereo image. The horizontal axis indicates the disparity magnitude of the Hidden Stereo image, while the vertical axis indicates the matched disparity magnitude of the standard stereo image. The data were averaged across subjects as well as texture types. The result shows that Hidden Stereo can produce perceived depth magnitude comparable to the standard stereo at least when the disparity magnitude is equal to or smaller than eight pixels, which corresponds to 6.08 min in visual angle in the condition we tested. Beyond this range, the perceived depth tends to level off. Figure 7b shows the rated scores of relative binocular luster impressions as a function of the disparity magnitude of the Hidden Stereo image. Here, the positive scores indicate that the binocular luster was higher in the Hidden Stereo image, and the negative scores indicate that it was higher in the standard stereo image. The data showed that the binocular luster tended to appear stronger in the Hidden Stereo image when the given disparity magnitude exceeded eight pixels. In summary, the first experiment suggested that the effective disparity limit of Hidden Stereo is around 6 min in visual angle.

#### 4.3 Perceptual experiment 2: Image quality

Next, to see whether our method can be applied for practical uses, we conducted a user study in which stereo image pairs of natural 3D scenes were used as the stimuli.

**Methods.** As the experimental stimuli, we took three scenes named *Lego*, *Flower*, and *Cards* from “The (New) Stanford Light Field Archive” (<http://lightfield.stanford.edu>). The image sizes of the stimuli (in pixels) were  $1280 \times 960$  (*Lego*),  $900 \times 1080$  (*Flower*), and  $1024 \times 1024$  (*Cards*). A set of four different stereo pairs was chosen for each scene. Within each set, the horizontal position of the right camera was changed while the position of the left camera was fixed. The disparity size of each stereo pair is shown in Table 1. The disparity size was computed as the maximum (95 percentile) of the absolute value of disparity from the screen.

Using these stereo pairs as inputs, we generated Hidden Stereo pairs in the way described in Section 3.2.4 with a scaling factor  $\alpha = 1$ . This time, we clipped the disparity inducers as mentioned in Section 3.2.3 such that the original image contrast (dynamic range) was kept intact (i.e., we set  $b_l = 0$  and  $b_u = 1$ ). Figures 11-13 shows the resulting images with the second and the fourth (largest) disparity for each scene. In Table 1, we show PSNR and SSIM [Wang et al. 2004] of the obtained Hidden Stereo images as objective measures of the image quality. To calculate these objective measures, we used camera images from the corresponding view points<sup>1</sup> in the light field dataset as reference images. We calculated the objective measures both for the left and right images, and averaged them to obtain the single values shown in Table 1. For cells with missing values, the corresponding reference images were not available in the light field datasets.

20 naïve participants (11 females and 9 males, aged from 29 to 45) took part in this second experiment. All participants had normal

Table 1. Objective Image quality scores of Hidden Stereo images

| Scene         | Disparity level | Disparity size (pixels) | Disparity size (min) | PSNR  | SSIM  |
|---------------|-----------------|-------------------------|----------------------|-------|-------|
| <i>Lego</i>   | 1               | 4.0                     | 4.21                 | 35.9  | 0.938 |
|               | 2               | 5.9                     | 6.24                 | -     | -     |
|               | 3               | 7.9                     | 8.34                 | 31.3  | 0.907 |
|               | 4               | 12.0                    | 12.6                 | 28.7  | 0.876 |
| <i>Flower</i> | 1               | 3.0                     | 3.15                 | 30.9  | 0.938 |
|               | 2               | 4.9                     | 5.13                 | -     | -     |
|               | 3               | 6.1                     | 6.37                 | 27.7  | 0.887 |
|               | 4               | 9.3                     | 9.72                 | 26.5  | 0.857 |
| <i>Cards</i>  | 1               | 3.1                     | 3.29                 | -     | -     |
|               | 2               | 7.1                     | 7.43                 | 24.3  | 0.868 |
|               | 3               | 10.1                    | 10.6                 | -     | -     |
|               | 4               | 13.9                    | 14.7                 | 19.56 | 0.684 |

visual acuity, and were able to discriminate the depth direction of random-dot stereogram with 2.1 min disparity. In the experiment, the stimuli were presented on a 3D LCD display (ASUS ROG SWIFT PG248Q, 24 inch,  $1920 \times 1080$  pixels, 120Hz) with stereo glasses (NVIDIA 3D Vision 2). The display gamma was linearized by using a colorimeter. The viewing distance was 90 cm.

The user study consisted of two separate sessions, a 3D session with stereo glasses and a 2D session without stereo glasses. The experimental procedure was based on the double-stimulus impairment scale method [BT.500-13 2012].

In the 3D session, the observers wore the stereo glasses, and evaluated the impairment in the image quality along with that in the depth impression of the Hidden Stereo image (test image) in comparison to the original standard stereo image (reference image). In addition to the stimulus conditions shown in Table 1, the session included control conditions, where the standard stereo image (with the minimum disparity for each scene) was presented as a test stimulus (i.e., the same standard stereo image was repeatedly presented both as reference and test). Thus, there were in total 15 conditions (4 disparity levels  $\times$  3 scenes + 3 controls). In each trial, the reference image was displayed for 5 seconds, then the test stimulus for 5 seconds. This was repeated twice. Between every two stimuli, a blank gray screen was inserted for 2 seconds. After the stimulus presentation, the observer was asked to evaluate the impairment in perceived image quality (which we expected to include image glare and luster) of the test image relative to the reference image on the following scale: 1 (very annoying), 2 (annoying), 3 (slightly annoying), 4 (perceptible, but not annoying), and 5 (imperceptible)<sup>2</sup>. During the judgment period, the task and the meaning of each number were clearly indicated on the monitor screen. After answering the questionnaire for the image quality, the observer was requested to evaluate the impairment in perceived depth structure using the same five-grade impairment scale. Each observer evaluated all of the 15 conditions presented in a pseudo-random order. For familiarization of the task and establishment of a proper rating

<sup>1</sup>Given that the horizontal camera positions (L,R) of the input standard stereo pair is (0,1), our method generates synthesized views from (-0.5,0.5). Thus, we used camera images from (-0.5,0.5) as references, not the input images.

<sup>2</sup>In order to avoid overestimation of performance of Hidden Stereo, we took a procedure that conservatively assumed that the 3D image quality of Hidden Stereo could not be better than that of the reference standard stereo. However, the 3D image quality can be better for Hidden Stereo in such a case where the crosstalk impairs that of the standard stereo, as we discuss in Section 5.

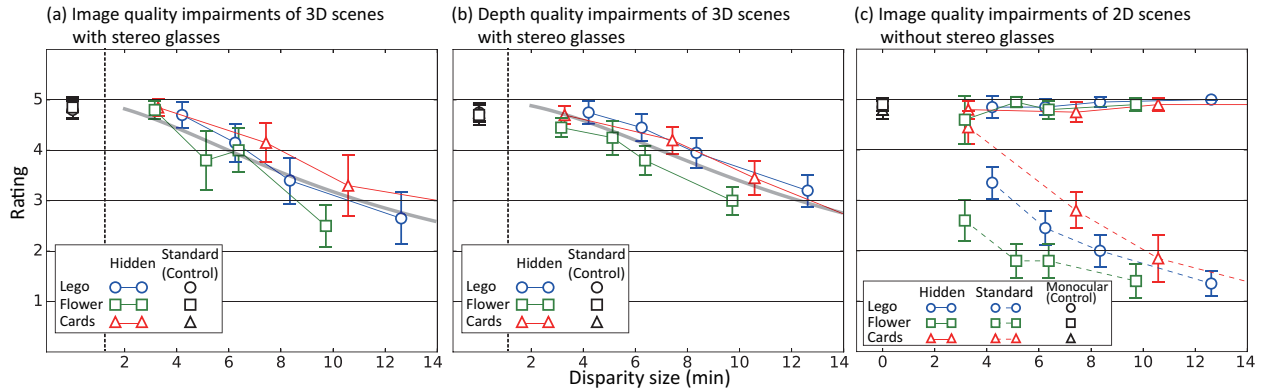


Fig. 8. Subjective quality impairments as a function of disparity sizes when natural scene images were used as test stimuli. The images used in the experiment are shown in Figures 11-13. The error bars show  $\pm 95\%$  confidence intervals. (a) Ratings for the image quality of perceived cyclopean image viewed with stereo glasses. (b) Ratings for the quality of perceived depth impressions viewed with stereo glasses. (c) Ratings for the image quality viewed without stereo glasses.

scale, prior to every session, each observer performed a training session using images of a scene that were not used in the main experiment, and then previewed the whole image set that would appear in the following main session.

In the 2D session, the observers evaluated the impairment of the image quality of the stereoscopically presented stimuli without wearing glasses. The reference stimulus was binocular presentation of the same monocular image (the left image of a standard stereo pair was presented to both eyes), while the test stimulus was either the Hidden Stereo image or the standard stereo image. In control conditions, the same monocular image was presented both as reference and test. There were in total 27 conditions (i.e., 4 disparity levels  $\times$  3 scenes  $\times$  2 (Hidden and standard) + 3 controls). Only the impairment in image quality was evaluated. Otherwise, the experimental procedure was the same as in the 3D session.

Each observer ran the 3D and 2D sessions successively, with the order of sessions being counterbalanced across observers.

**Results.** The results of the 3D session, i.e., the averaged scores of the image quality and the depth quality, are shown in Figure 8a and 8b, as a function of the disparity size. The errorbars show the 95 % confidence intervals. The thick gray curve in each plot indicates a non-symmetrical logistic function fit to the data. The Pearson correlation coefficients after fitting the logistic function were 0.90 for the image quality data, and 0.93 for the depth quality data. When the fourth grade score ("perceptible, but not annoying") is used as the tolerable threshold, both the image quality evaluation and the depth quality evaluation suggest that the disparity sizes up to about 6 min are the effective range of our Hidden Stereo method. That is, when the maximum absolute disparity of an image is less than 6 min, 3D presentation with Hidden Stereo has acceptable perceptual quality. The limitation is close to the value estimated in the first experiment in Section 4.2.

In comparison with the disparity size, the objective measures of image impairment (PSNR and SSIM) were not good predictors of the subjective rating scores. The Pearson correlation coefficients after fitting the logistic function were 0.55 for the image quality scores vs PSNR, 0.76 for the depth quality scores vs PSNR, 0.67 for

the image quality scores vs SSIM, and 0.85 for the depth quality scores vs SSIM.

The results of the 2D session are shown in Figure 8c. Except for the smallest disparity condition of the scene "Cards", the image quality of the standard stereo images was significantly worse than that of the monocular images, and was well below the fourth grade score (tolerable threshold). On the other hand, we did not find any significant difference in the image quality impairment scores between the Hidden Stereo images and the monocular images regardless of the disparity size and scene.

Taken together, the results of our experiments suggest that Hidden Stereo is practically useful under the condition where most of disparities in an image are less than 6 min. Under this condition, for viewers with stereo glasses, the 3D image quality of Hidden Stereo is almost comparable to that of the standard stereo, and for viewers without glasses, the 2D image quality of Hidden Stereo is as good as that of monocular image presentation, and better than that of the standard stereo. Beyond the 6-min limit, however, the 3D image quality of Hidden Stereo could be significantly worse than the standard stereo.

However, the results of the 3D session also suggest that disparity size in terms of visual angle is not the only determinant of the 3D image quality of Hidden Stereo, since the rating value plotted against disparity does not perfectly overlap among the three scene conditions. Further investigation is necessary to achieve a good control over the image and depth quality obtained by our Hidden Stereo method.

#### 4.4 Manipulation of real object's depth by light projection

An important aspect of Hidden Stereo is that the disparity inducer can be independently presented on a 2D image content. Here we show one such example, where patterns containing only the disparity inducer are projected on the surface of a printed picture using a 3D projector. Figure 9 shows a schematic depiction of the setup. The projector was a commercially available 3D projector with active shutter glasses. Using a printed picture as a cyclopean view image  $I_C$ , we generated a disparity inducer  $I_D$  in the way described in

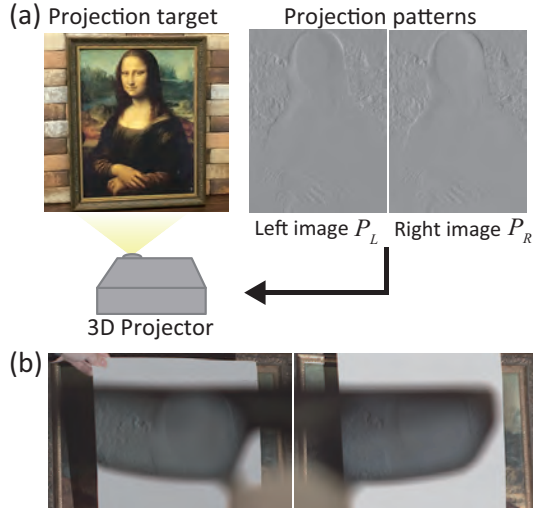


Fig. 9. Light projection system to give depth impressions to a real object. (a) The disparity-inducer patterns are presented on the target painting (*Mona Lisa* by Leonardo da Vinci) from a 3D projector. (b) The projection patterns visible only through stereo glasses. Since the patterns are identical except for contrast polarity, they are fused into a neutral gray pattern when viewed without stereo glasses (see the regions outside of the glasses in the images)

Section 3.2 with a manually designed disparity map. The projected left and right patterns ( $P_L$  and  $P_R$ ) were generated by

$$P_L(\mathbf{x}) = W(\mathbf{x})I_D(\mathbf{x}) + 0.5, \quad (24)$$

$$P_R(\mathbf{x}) = -W(\mathbf{x})I_D(\mathbf{x}) + 0.5, \quad (25)$$

where  $W(\mathbf{x})$  is a weight function that compensates for the projector's light attenuation according to the surface albedo of the projection target. As for the details about the geometric and photometric calibration, please refer to [Bimber et al. 2007]. We only computed the disparity inducer in the luminance channel, and thus the projection patterns  $P_L$  and  $P_R$  contain luminance alone. The disparity inducer was clipped such that  $W(\mathbf{x})I_D(\mathbf{x})$  did not exceed the bound  $[-0.5, 0.5]$ . Therefore, the projected patterns are fused into a neutral gray pattern when viewed by naked eyes. Although there are differences in colors and resolutions between a real picture and the projected patterns, the HVS resolves those small dissociations in a way similar to Deformation Lamps [Kawabe et al. 2016], where a motion impression is added to a static object by projecting luminance motion signals. Consequently, viewers with shutter glasses perceive the natural depth impression added by the light projection, while viewers without glasses can still enjoy the original appearance of the picture. The light projection system proposed here can be used, for example, in an art museum to present additional depth impressions onto existing paintings without causing any conflicts with observers who want to appreciate their original appearances.

## 5 LIMITATIONS AND FUTURE WORKS

The major limitations of Hidden Stereo emerge when stereoscopic images contain large disparity. First, the maximum disparity it can add is a quarter cycle of the subband frequency. The upper limit of disparity, in terms of image pixels or visual angle, decreases as the subband scale becomes finer. Second, with an increase in disparity, the contrast of the disparity inducer image increases. This reduces the effective dynamic range of the image and makes the perceptual artifact (binocular luster) more visible to viewers. The artifacts are clearly visible in Figure 10 wherein the disparity size is about three times larger than the disparity level four conditions of Table 1. According to our user study, Hidden Stereo can produce stereoscopic images with acceptable 3D image/depth qualities as long as the maximum binocular disparity (from the horopter) does not exceed  $\sim 6$  min, at least under the image and viewing conditions we used. The 6-min disparity limit implies the reproducible depth range, with the inter-pupillary distance of 63 mm, to be [493, 507] mm at the viewing distance of 500 mm, [973, 1029] mm at 1000 mm, [1895, 2117] mm at 2000 mm, and [2770, 3272] mm at 3000 mm. For comparison, the horizontal disparity within which human observers can binocularly fuse monocular patterns without vergence eye movements (Panum's fusional radius) is  $\sim 10$  min for spatially fine patterns presented in the fovea [Howard 2012a; Schor et al. 1984]. Although this is the same order as the limit of our method, the fusion limit extends as the spatial frequency is lowered or the retinal eccentricity is increased, and the upper depth perception limit is generally higher than the fusion limit. Concerning 3D viewing with vergence, the comfortable disparity range within which the vergence-accommodation conflict is tolerable is 1-2 deg [Lambooy et al. 2009; Shibata et al. 2011], and thus is an order of magnitude larger than the limit of our method. It should be also noted that image re-alignment with vergence cannot perfectly overcome binocular image mismatches produced with Hidden Stereo. Our technique is not intended for large 3D disparities that would induce a strong vergence.

Due to the limitations in effective disparity, for translation of standard stereo images to our ghost-free format, one may have to consider depth compression in many cases. While we have only tested simple linear compression, the utility of Hidden Stereo would be greatly improved by combining it with the perception-based disparity compression [Didyk et al. 2011, 2012; Masia et al. 2013]. It should be noted that depth compression *per se* is functionally useful, since it contributes to comfortable viewing of 3D stereo images. [Didyk et al. 2012; Kellnhofer et al. 2016]. Where all the image disparities are compressed into the Panum's fusion range, the vergence-accommodation conflict [Lambooy et al. 2009; Shibata et al. 2011] is not a problem anymore. For natural scenes containing many depth cues, inconsistency in the magnitude, and sometimes the sign, of binocular disparity with other cues, as well as with the grand truth, does not have a critical influence on human depth perception [Didyk et al. 2011; Gregory 1970; Landy et al. 1995].

Having said that, to increase the utility of Hidden Stereo, we think it worthwhile to pursue a method to enlarge the maximum perceived depth. One possible direction is to enhance the contribution of coarse subbands to depth perception, and another is

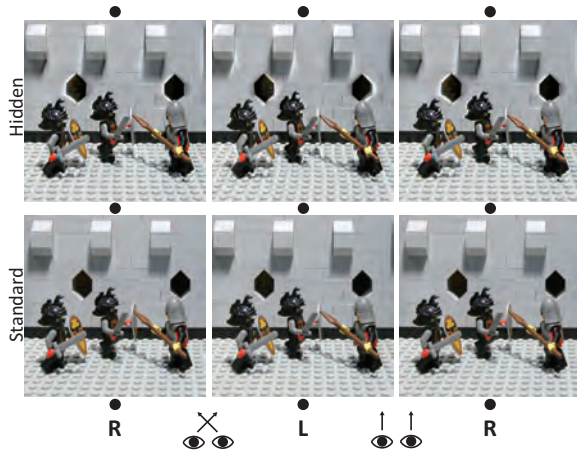


Fig. 10. Hidden Stereo pair obtained by transforming a stereo pair with huge disparities (the maximum disparity is  $\sim 36$  pixels in the original resolution of  $1024 \times 1024$  pixels). In this case, the significant binocular luster is visible when the two stereo images are binocularly fused. The bottom row shows the original standard stereo pair taken from "The (New) Stanford Light Field Archive" (<http://lightfield.stanford.edu>). The image pairs in the left and middle columns are for cross fusion, and those in the middle and right columns are for parallel fusion. (The images in the left and right columns are identical.)

to suppress the contribution of fine subbands. The stereo images produced by Hidden Stereo are different from standard stereo images in several respects, including monocular occlusion. Standard stereo images often contain image regions that are occluded to one eye, and visible only to the other eye. The HVS is able to use such monocular regions as a depth cue [Nakayama and Shimojo 1990]. Nevertheless, Hidden Stereo does not properly handle this occlusion cue. Although the contribution of monocular occlusion is likely to be minor for images with small disparities suited to Hidden Stereo, we have not perfectly understood the perceptual consequence of the lack of monocular occlusion. Likewise, since Hidden Stereo "cheats" human vision in an unprecedented way, the current understanding of human binocular processing is insufficient to precisely predict how the depth and cyclopean image (including binocular lusters) are perceived for specific Hidden Stereo images. Further evaluation under a wide range of conditions will be necessary.

Our method provides a simple way to linearly decompose a stereo image pair into a single (cyclopean) image and a depth pattern (disparity inducer). If the disparity inducer is given, it is easy to change any 2D content into 3D, as we demonstrated in Section 4.4. It remains to be studied whether this decomposition contributes to other applications. One promising direction is effective image data compression. It should be also noted that this image decomposition works effectively only for pairs of slightly shifted images, not for arbitrary image pairs.

Crosstalk, the imperfect separation of the left- and right-eye images, produces perception of ghosts when standard stereo images are viewed with stereo glasses. Hidden Stereo effectively excludes

the stereo ghosts of this type in addition to those seen without glasses. Specifically, crosstalk only reduces the effective contrast of the disparity inducer pattern. As a result, the perceived depth is compressed in proportion to the magnitude of crosstalk, but the perceived cyclopean image remains clear. Given that crosstalk has been regarded as the main display-related perceptual factor degrading image quality and causing visual discomfort [Lambooij et al. 2009], this is a clear advantage of Hidden Stereo over the standard method even for the purpose of genuine stereoscopic presentation. Since crosstalk is also the key problem for autostereoscopic displays, one may be able to develop the notion of disparity cancellation to improve their image qualities.

The present algorithm adds only horizontal disparities to a cyclopean image, but it is easy to add vertical disparities in the same manner. Vertical disparities play significant roles in proper viewing of stereoscopic displays by human observers [Held and Banks 2008].

The current implementation of our method assumes a linear relationship between the image pixel value and the final image intensity. We therefore linearized the monitor response in advance in our user experiments. However, the display response is generally non-linear, with the typical gamma value ( $\gamma$ ) being 2.2. If one applies our algorithm without considering gamma correction, the positive and negative disparity inducers will not be perfectly cancelled out. A simple way to avoid this artifact is: (1) Transform the input cyclopean image by  $I_{outC} = I_{signalC}^\gamma$ , where  $I_{out}$  is the expected monitor output and  $I_{signal}$  is the image pixel value. (2) Generate the disparity inducer  $I_D$  from  $I_{outC}$  and obtain the right and the left images by  $I_{outL} = I_{outC} - I_D$  and  $I_{outR} = I_{outC} + I_D$ , respectively. (3) Transform the left and right images by  $I_{signalL} = I_{outL}^{1/\gamma}$  and  $I_{signalR} = I_{outR}^{1/\gamma}$ . By presenting  $I_{signalL}$  and  $I_{signalR}$  through the display, the perfect cancellation of  $I_D$  will be achieved.

Finally, it is challenging for Hidden Stereo to produce ghost-free images for one of the popular stereo presentation systems, anaglyph 3D. Since the disparity inducer components are opposite in sign between the left and right images, when two images in different colors (e.g., red and cyan) are superimposed, the disparity inducer components become reddish or cyanish depending on the direction of the disparity. This problem is hard to solve, since decomposing multi-sensor images into luminance (average) and color (difference) components is computationally similar to decomposing stereo images into a cyclopean image and a disparity pattern (see [Adelson and Bergen 1991] for related discussion). However, an advanced wavelength-multiplexing anaglyph [Jorke et al. 2009] may provide a solution.

## 6 CONCLUSIONS

This paper proposed a novel method to generate perfectly ghost-free stereoscopic images. Our method gives binocular disparity to a 2D image by the addition of a quadrature-phase pattern that induces spatial subband phase shifts. Physical fusion of the two images cancels out the disparity-inducer components. This makes only the original 2D pattern visible to viewers without glasses. At the same time, viewers with glasses are able to enjoy 3D contents without significant loss of image quality. As far as we know, this is the

first method to present perfectly backward compatible 3D using standard stereo hardware.

## REFERENCES

- Edward. H. Adelson and James. R. Bergen. 1991. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*. MIT Press, 3–20.
- Stuart Anstis and Alan Ho. 1998. Nonlinear combination of luminance excursions during flicker, simultaneous contrast, afterimages and binocular fusion. *Vision Research* 38, 4 (Feb. 1998), 523–539. DOI: [https://doi.org/10.1016/S0042-6989\(97\)00167-3](https://doi.org/10.1016/S0042-6989(97)00167-3)
- Stuart M. Anstis, Ian P. Howard, and Brian Rogers. 1978. A Craik-O'Brien-Cornsweet illusion for visual depth. *Vision Research* 18, 2 (1978), 213–217.
- Martin S. Banks, Sergei Gepshtein, and Michael S. Landy. 2004. Why is spatial stereoresolution so low? *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 24, 9 (March 2004), 2077–2089. DOI: <https://doi.org/10.1523/JNEUROSCI.3852-02.2004>
- Oliver Bimber, Daisuke Iwai, Gordon Wetzstein, and Anselm Grundhöfer. 2007. The Visual Computing of Projector-Camera Systems. In *Eurographics 2007 - State of the Art Reports*.
- Randolph Blake and Hugh Wilson. 2011. Binocular vision. *Vision Research* 51, 7 (April 2011), 754–770. DOI: <https://doi.org/10.1016/j.visres.2010.10.009>
- Mark F. Bradshaw and Brian J. Rogers. 1999. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research* 39, 18 (Sept. 1999), 3049–3056. DOI: [https://doi.org/10.1016/S0042-6989\(99\)00015-2](https://doi.org/10.1016/S0042-6989(99)00015-2)
- Allen Brookes and Kent A. Stevens. 1989. The Analogy between Stereo Depth and Brightness. *Perception* 18, 5 (Oct. 1989), 601–614.
- ITU-R Rec. BT.500-13. 2012. *Methodology for the subjective assessment of the quality of television pictures*.
- Tom Cornsweet. 2012. *Visual Perception*. Academic Press.
- Bruce G. Cumming and Gregory C. DeAngelis. 2001. The Physiology of Stereopsis. *Annual Review of Neuroscience* 24 (2001), 203–238.
- H. de Lange Dzn. 1958. Research into the Dynamic Nature of the Human Fovea → Cortex Systems with Intermittent and Modulated Light I Attenuation Characteristics with White and Colored Light. *Journal of the Optical Society of America* 48, 11 (Nov. 1958), 777.
- Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. 2011. A Perceptual Model for Disparity. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2011, Vancouver)* 30, 4 (2011).
- Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. 2012. Apparent stereo: the Cornsweet illusion can enhance perceived depth. In *Human Vision and Electronic Imaging XVII, IS&T/SPIE's Symposium on Electronic Imaging*. Burlingame, CA, 1–12.
- Piotr Didyk, Pitchaya Sitthi-Amorn, William Freeman, Frédo Durand, and Wojciech Matusik. 2013. Joint View Expansion and Filtering for Automultiscopic 3D Displays. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia 2013, Hong Kong)* 32, 6 (2013).
- Jian Ding, Stanley A. Klein, and Dennis M. Levi. 2013. Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. *Journal of Vision* 13, 2 (2013), 1–37. DOI: <https://doi.org/10.1167/13.2.13>
- Jian Ding and George Sperling. 2006. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4 (Jan. 2006), 1141–1146. DOI: <https://doi.org/10.1073/pnas.0509629103>
- Monika A. Formankiewicz and J. D. Mollon. 2009. The psychophysics of detecting binocular discrepancies of luminance. *Vision Research* 49, 15 (July 2009), 1929–1938. DOI: <https://doi.org/10.1016/j.visres.2009.05.001>
- Wataru Fujimura, Yukua Koide, Robert Songer, Takahiro Hayakawa, Akihiko Shirai, and Kazuhisa Yanaka. 2012. 2x3D: Real Time Shader for Simultaneous 2D/3D Hybrid Theater. In *SIGGRAPH Asia 2012 Emerging Technologies (SA '12)*. ACM, 1:1–1:2. DOI: <https://doi.org/10.1145/2407707.2407708>
- Mark A. Georgeson, Stuart A. Wallis, Tim S. Meese, and Daniel H. Baker. 2016. Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research* 129 (Dec. 2016), 98–118. DOI: <https://doi.org/10.1016/j.visres.2016.08.001>
- Richard L. Gregory. 1970. *Intelligent Eye*. Littlehampton Book Services Ltd, London.
- Robert T. Held and Martin S. Banks. 2008. Misperceptions in Stereoscopic Displays: A Vision Science Perspective. *ACM Transactions on Graphics* (2008), 23–32.
- Sid Henriksen, Seiji Tanabe, and Bruce Cumming. 2016. Disparity processing in primary visual cortex. *Philosophical Transactions of the Royal Society B* 371, 1697 (2016).
- Ian P. Howard. 2012a. *Perceiving in Depth Volume 1 Basic Mechanisms*. Oxford University Press.
- Ian P. Howard. 2012b. *Perceiving in Depth Volume 2 Stereoscopic Vision*. Oxford University Press.
- Helmut Jorke, Arnold. Simon, and Markus. Fritz. 2009. Advanced stereo projection using interference filters. *Journal of the Society for Information Display* 17, 5 (2009), 407–410.
- Takahiro Kawabe, Taiki Fukiage, Masataka Sawayama, and Shin'ya Nishida. 2016. Deformation Lamps: A Projection Technique to Make Static Objects Perceptually Dynamic. *ACM Trans. Appl. Percept.* 13, 2 (March 2016), 10:1–10:17. DOI: <https://doi.org/10.1145/2874358>
- Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M. Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. 2016. GazeStereo3D: Seamless Disparity Manipulations. *ACM Trans. Graph.* 35, 4 (July 2016), 68:1–68:13.
- Frederick. A. Kingdom and Bernard. Moulden. 1988. Border effects on brightness: a review of findings, models and issues. *Spatial Vision* 3, 4 (1988), 225–262.
- Frederick. A. Kingdom and David. R. Simmons. 1996. Stereoacuity and colour contrast. *Vision Research* 36, 9 (1996), 1311–1319.
- Frank L. Kooi and Alexander Toet. 2004. Visual comfort of binocular and 3D displays. *Displays* 25, 2–3 (Aug. 2004), 99–108. DOI: <https://doi.org/10.1016/j.displa.2004.07.004>
- Marc Lambooi, Wijnand IJsselstein, Marten Fortuin, and Ingrid Heynderickx. 2009. Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology* 53, 3 (2009), 1–14.
- Michael S. Landy, Laurence T. Maloney, Elizabeth B. Johnston, and Mark Young. 1995. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research* 35, 3 (Feb. 1995), 389–412. DOI: [https://doi.org/10.1016/0042-6989\(94\)00176-M](https://doi.org/10.1016/0042-6989(94)00176-M)
- Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. 2010. Nonlinear Disparity Mapping for Stereoscopic 3D. In *ACM SIGGRAPH 2010 Papers (SIGGRAPH '10)*. ACM, New York, NY, USA, 75:1–75:10.
- Hwan S. Lee and Allan C. Dobbins. 2006. Perceiving surfaces in depth beyond the fusion limit of their elements. *Perception* 35, 1 (2006), 31–39.
- Gordon E. Legge. 1984. Binocular contrast summation—II. Quadratic summation. *Vision Research* 24, 4 (1984), 385–394.
- Gordon E. Legge and Gary S. Rubin. 1981. Binocular interactions in suprathreshold contrast perception. *Perception & Psychophysics* 30, 1 (Jan. 1981), 49–61. DOI: <https://doi.org/10.3758/BF03206136>
- Ira Ludwig, Wolfgang Pieper, and Harald Lachnit. 2007. Temporal integration of monocular images separated in time: Stereopsis, stereoacuity, and binocular luster. *Perception & Psychophysics* 69, 1 (2007), 92–102. DOI: <https://doi.org/10.3758/BF03194456>
- William R. Mark, Leonard McMillan, and Gary Bishop. 1997. Post-rendering 3D Warping. In *Proceedings of 1997 Symposium on Interactive 3D Graphics*. ACM, New York, NY, USA, 7–16.
- Belen Masia, Gordon Wetzstein, Carlos Aliaga, Ramesh Raskar, and Diego Gutierrez. 2013. Display adaptive 3D content remapping. *Computer and Graphics* 37, 8 (2013), 983–996.
- Rainer Mausfeld, Gunnar Wendt, and Jürgen Golz. 2014. Lustrous Material Appearances: Internal and External Constraints on Triggering Conditions for Binocular Lustre. *i-Perception* 5, 1 (Feb. 2014), 1–19. DOI: <https://doi.org/10.1068/i0603>
- Tim S. Meese, Mark A. Georgeson, and Daniel H. Baker. 2006. Binocular contrast vision at and above threshold. *Journal of Vision* 6, 11 (Oct. 2006), 1224–1243. DOI: <https://doi.org/10.1167/6.11.7>
- Ken Nakayama and Shinsuke Shimojo. 1990. Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research* 30, 11 (1990), 1811–1825.
- Izumi. Ohzawa, Gregory. C. DeAngelis, and Ralph. D. Freeman. 1990. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science (New York, N.Y.)* 249, 4972 (Aug. 1990), 1037–1041.
- Adriana Olmos and Frederick A. A. Kingdom. 2004. A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33 (2004), 1463–1473.
- D. A. Palmer. 1961. Measurement of the Horizontal Extent of Panum's Area by a Method of Constant Stimuli. *Optica Acta* 8 (1961), 151–159. DOI: <https://doi.org/10.1080/713826374>
- Javier Portila and Eero. P. Simoncelli. 2000. A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *International Journal of Computer Vision* 40, 1 (2000), 49–70.
- Carl Pulfrich. 1922. Die Stereoskopie im Dienste der isochromen und heterochromen Photometrie. *Naturwissenschaften* 10, 35 (1922), 751–761.
- Dale Purves, Amita Shimpi, and R. Beau Lotto. 1999. An Empirical Explanation of the Cornsweet Effect. *The Journal of Neuroscience* 19, 19 (Oct. 1999), 8542–8551.
- Ning. Qian and Ralph. D. Freeman. 2009. Pulfrich phenomena are coded effectively by a joint motion-disparity process. *Journal of Vision* 9, 5 (May 2009), 1–16. DOI: <https://doi.org/10.1167/9.5.24>
- Jenny C. A. Read and Bruce G. Cumming. 2005. All Pulfrich-like illusions can be explained without joint encoding of motion and disparity. *Journal of Vision* 5, 11 (2005), 901–927.
- Steven Scher, Jing Liu, Rajan Vaish, Prabath Gunawardane, and James Davis. 2013. 3D+2DTV: 3D Displays with No Ghosting for Viewers Without Glasses. *ACM Transactions on Graphics* 32, 3 (2013), 21:1–10.

- Clifton Schor, Ivan Wood, and Jane Ogawa. 1984. Binocular sensory fusion is limited by spatial resolution. *Vision Research* 24, 7 (1984), 661–665.
- Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision* 11, 8 (2011), 1–29.
- Mel Siegel and Shojiro Nagata. 2000. Just Enough Reality: Comfortable 3D Viewing via Microstereopsis. *IEEE Transactions on Circuits and Systems for Video Technology* 10, 3 (2000), 387–396.
- Eero. P. Simoncelli and William T. Freeman. 1995. The Steerable Pyramid: A Flexible Architecture for Multi-scale Derivative Computation. In *Proceedings of the 2nd IEEE International Conference on Image Processing*, Vol. 3. Washington, DC, USA, 444–447.
- Masahiko Terao, Junji Watanabe, Akihiro Yagi, and Shin'ya Nishida. 2010. Smooth Pursuit Eye Movements Improve Temporal Resolution for Color Perception. *PLoS ONE* 5, 6 (June 2010), e11214. DOI: <https://doi.org/10.1371/journal.pone.0011214>
- Neal Wadhwa, Michael Rubinstein, Frédo Durand, and W. T. Freeman. 2013. Phase-Based Video Motion Processing. *ACM Transactions on Graphics* 32, 4 (2013), 1–9.
- Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. 2014. Riesz Pyramids for Fast Phase-Based Video Magnification. (2014).
- Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero. P. Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.
- Junji Watanabe and Shin'ya Nishida. 2007. Veridical perception of moving colors by trajectory integration of input signals. *Journal of Vision* 7, 11 (2007).
- Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttg, Frédo Durand, and William Freeman. 2012. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics* 31, 4 (July 2012), 65:1–8.
- C. Lawrence Zitnick, Sing Bing Kang, Matthew Uyttendaele, Simon Winder, and Richard Szeliski. 2004. High-quality Video View Interpolation Using a Layered Representation. In *ACM SIGGRAPH 2004 Papers (SIGGRAPH '04)*. ACM, New York, NY, USA, 600–608.

## A HUMAN BINOCULAR VISION

A conventional stereoscopic image pair is geometrically correct in that it consists of two images of the same scene viewed from slightly different vantage points. The stereo images made by our method do not strictly follow this principle. In consideration of human brain processing, however, geometrical correctness is not a critical condition for a stereo image pair to produce binocular depth perception. In order to justify an apparent anomaly of our method, here we briefly explain two mechanisms of human binocular processing relevant to Hidden Stereo. One is stereopsis, depth computation from binocular disparity. The other is binocular combination of two retinal images into a single cyclopean image.

### A.1 Stereopsis

The disparity energy detector [Ohzawa et al. 1990] is the current standard model of human binocular disparity detection [Blake and Wilson 2011; Cumming and DeAngelis 2001; Henriksen et al. 2016; Howard 2012a]. The frontend of this detector is a pair of oriented Gabor-shaped receptive fields for the two eyes. (A Gabor function is a sinusoid multiplied by a Gaussian envelope.) The monocular receptive field for one eye is spatially shifted in phase from that for the other eye. An idealized model assumes a quadrature-phase ( $\pi/2$ ) shift, although in reality the phase shift may not be limited to  $\pi/2$ , and/or it may be accompanied by a position shift of the receptive field [Henriksen et al. 2016]. Disparity energy is the squared sum of the outputs of the two monocular receptive fields. This response reflects binocular cross-correlation at the spatial location, spatial frequency (scale), and orientation falling within the preferred range of the detector. The size of the window (i.e., receptive field) within which the cross-correlation is computed determines the spatial resolution of depth processing [Banks et al. 2004]. As the preferred spatial

frequency is lowered and the window size is increased, the spatial resolution becomes worse. A bank of detectors measures local disparity energies in parallel for a range of scales and orientations. How these disparity-energy measurements are integrated into the final depth perception remains to be investigated, yet it has been suggested that spatial alignments of the peak of the correlation across the scale could be evidence of true disparity [Henriksen et al. 2016].

It should be noted that a disparity energy detector responds to disparity whenever there is a phase shift in the preferred range of spatial frequency. Driving the first stage disparity processing does not require a shift of the whole pattern, i.e., a coherent shift at all scales. This is a basic idea underlying the disparity manipulation in Hidden Stereo. However, it is difficult to predict the perceived depth magnitude with this technique. Therefore, we examined this issue in Section 4.2.

### A.2 Binocular combination

Computation of a cyclopean image from a binocular combination is not directly linked with stereoscopic processing, since binocular depth is perceived with disparities well above the perceptual fusion limit [Lee and Dobbins 2006; Palmer 1961]. The rule of binocular combination, specifically supra-threshold image contrast summation, depends on the magnitude of the discrepancy between the images on the two retinæ [Legge and Rubin 1981]. First, when the two images are fairly similar to each other, the combined image can be approximated by a linear average. Second, as the image difference is increased, the image with the stronger contrast contributes more to the cyclopean image. Quadratic summation was proposed as a rule to describe this non-linear integration [Legge 1984], but more recent studies have proposed more elegant and complicated mechanical models to account for a broader range of binocular combination phenomena [Ding et al. 2013; Ding and Sperling 2006; Georgeson et al. 2016; Meese et al. 2006]. Third, when the two images are very different and incompatible with each other (e.g., luminance increment vs. luminance decrement), binocular rivalry produces a shiny, lustrous appearance of the incompatible area, which is known as binocular luster [Anstis and Ho 1998; Formankiewicz and Mollon 2009; Howard 2012b; Ludwig et al. 2007; Mausfeld et al. 2014].

As long as the image difference between a stereo image pair made by Hidden Stereo is small enough, we can expect the cyclopean image seen with glasses to be close to a linear average of the two images, which is the same as the original 2D image. On the other hand, when the addition of a large disparity causes a large image difference, the quality of the cyclopean image may be impaired by binocular luster. It is difficult to precisely predict "what happens when" only from the reported psychophysical data obtained using simple artificial stimuli such as a sinusoidal grating. We therefore tested this issue in Section 4.2.

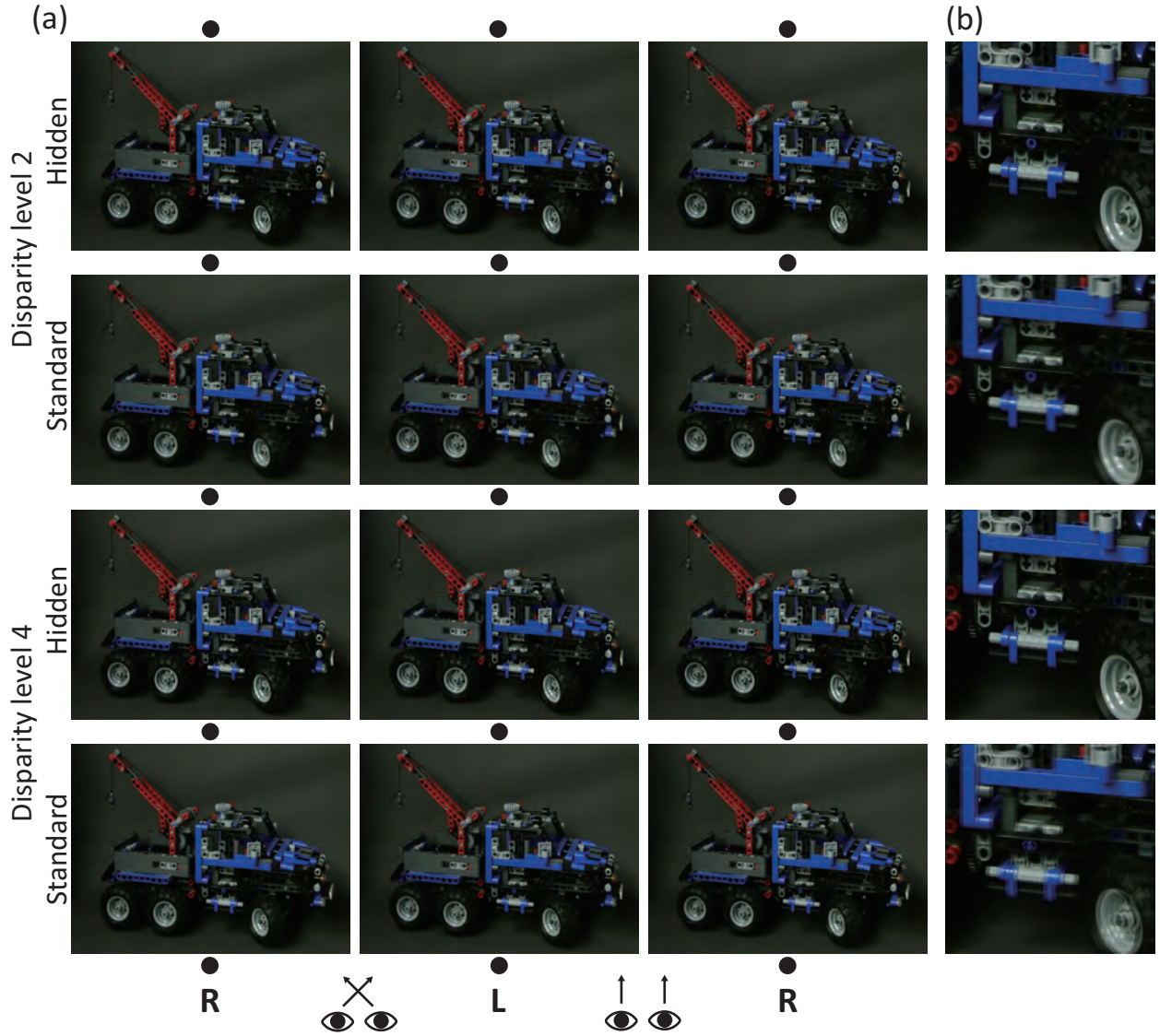


Fig. 11. Comparison of image quality to the conventional standard stereo images for *Lego* scene (from "The (New) Stanford Light Field Archive," <http://lightfield.stanford.edu>). The images shown here were also used in the user study in Section 4.3. The images in odd rows are Hidden Stereo images (Hidden) and those in even rows are the standard stereo images (Standard) used to generate the corresponding Hidden Stereo images. The upper-half rows and the lower-half rows show stereo images in the disparity level 2 and those in the disparity level 4 in Table 1, respectively. (a) Comparison of 3D scenes. The image pairs in the left and middle columns are for cross fusion, and those in the middle and right columns are for parallel fusion. (The images in the left and right columns are identical.) The images are seen in the same visual angle as in the experiment when viewed from a distance 2.5 times larger than the image width. (b) Comparison of 2D scenes viewed without stereo glasses. Here, we presented synthetic results generated by averaging the left and right images. For the standard stereo images, the physically fused images show stereo ghosts and blurs. On the other hand, for the Hidden Stereo pair, the fusing process just returns the image pair back to one of the input stereo images, and thus never produces ghosts or blurs. Please refer to supplemental materials for more results in large image size.

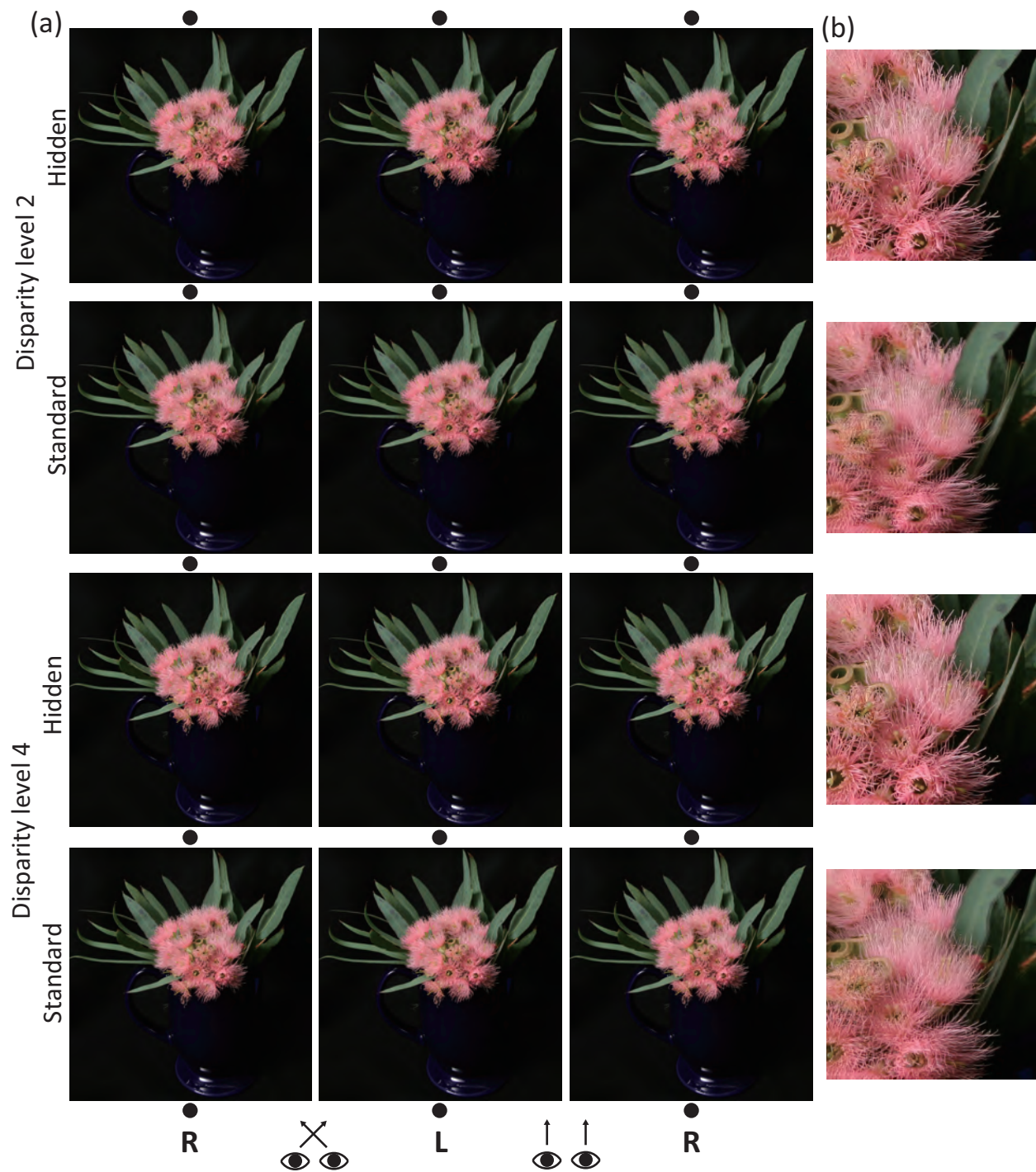


Fig. 12. Comparison of image quality to the conventional standard stereo images for *Flower* scene (from "The (New) Stanford Light Field Archive," <http://lightfield.stanford.edu>). (a) Comparison of 3D scenes. The images are seen in the same visual angle as in the experiment when viewed from a distance 3.6 times larger than the image width. (b) Comparison of 2D scenes viewed without stereo glasses.

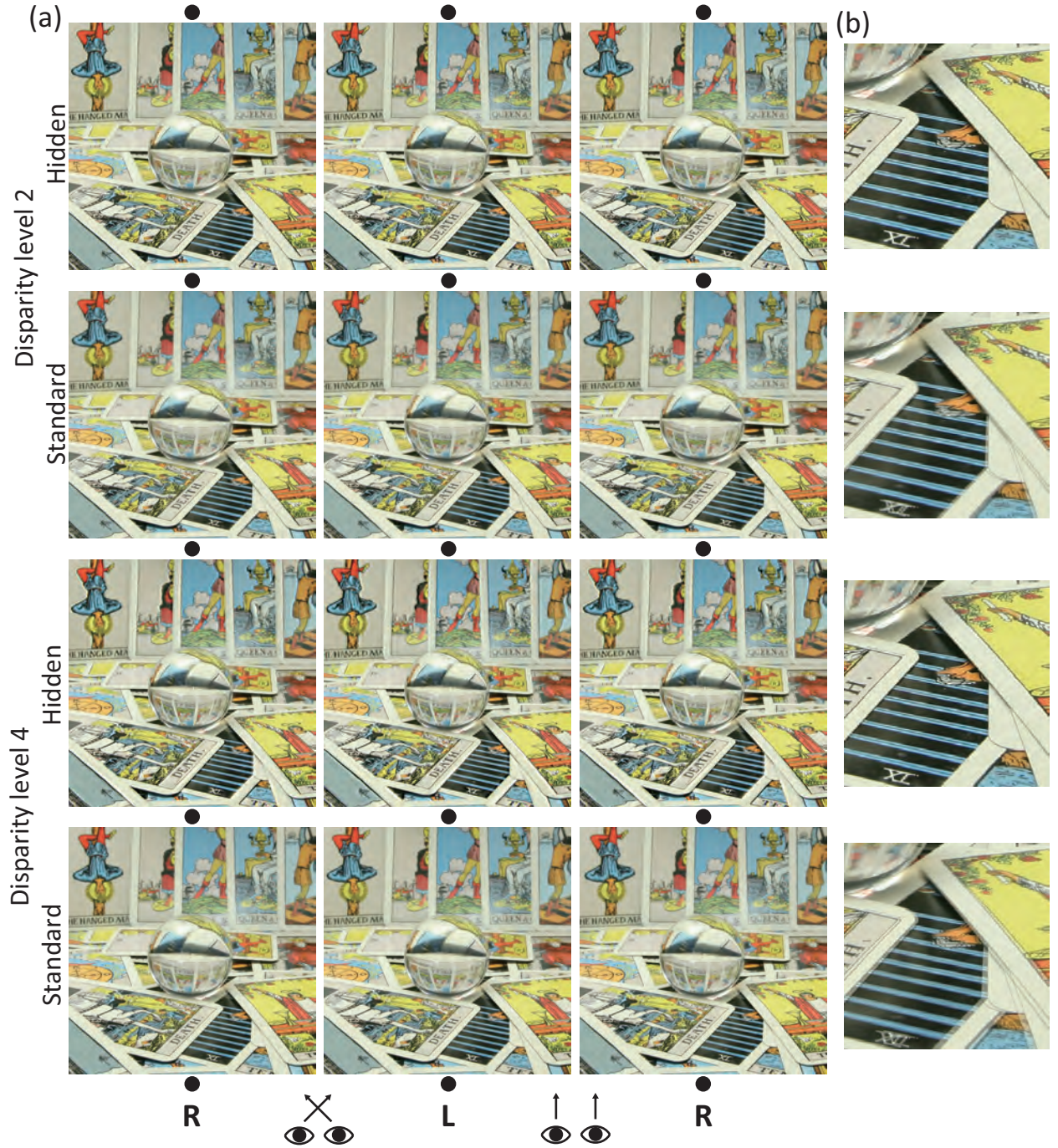


Fig. 13. Comparison of image quality to the conventional standard stereo images for *Cards* scene (from "The (New) Stanford Light Field Archive," <http://lightfield.stanford.edu>). (a) Comparison of 3D scenes. The images are seen in the same visual angle as in the experiment when viewed from a distance 3.2 times larger than the image width. (b) Comparison of 2D scenes viewed without stereo glasses.