

GazeStereo3D: Seamless Disparity Manipulations

Petr Kellnhofer^{1,2} Piotr Didyk^{2,3} Karol Myszkowski² Mohamed M. Hefeeda⁴ Hans-Peter Seidel² Wojciech Matusik¹

¹MIT CSAIL ²MPI Informatik ³Saarland University, MMCI ⁴Qatar Computing Research Institute

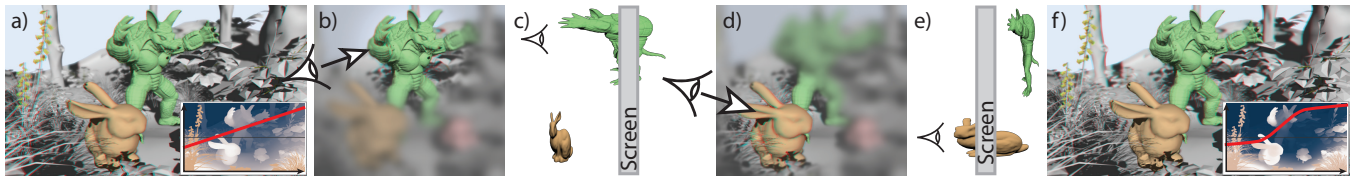


Figure 1: Beyond common disparity mapping (a) our approach adapts to attended regions such as the Armadillo (b) or Bunny (d) and modifies disparity in that region to enhance depth and reduce discomfort from the accommodation-vergence conflict (c,e). To that end we build non-linear remapping curves and use our novel perceptual model to ensure a seamless transition between them (f).

Abstract

Producing a high quality stereoscopic impression on current displays is a challenging task. The content has to be carefully prepared in order to maintain visual comfort, which typically affects the quality of depth reproduction. In this work, we show that this problem can be significantly alleviated when the eye fixation regions can be roughly estimated. We propose a new method for stereoscopic depth adjustment that utilizes eye tracking or other gaze prediction information. The key idea that distinguishes our approach from the previous work is to apply gradual depth adjustments at the eye fixation stage, so that they remain unnoticeable. To this end, we measure the limits imposed on the speed of disparity changes in various depth adjustment scenarios, and formulate a new model that can guide such seamless stereoscopic content processing. Based on this model, we propose a real-time controller that applies local manipulations to stereoscopic content to find the optimum between depth reproduction and visual comfort. We show that the controller is mostly immune to the limitations of low-cost eye tracking solutions. We also demonstrate benefits of our model in off-line applications, such as stereoscopic movie production, where skillful directors can reliably guide and predict viewers' attention or where attended image regions are identified during eye tracking sessions. We validate both our model and the controller in a series of user experiments. They show significant improvements in depth perception without sacrificing the visual quality when our techniques are applied.

Keywords: stereoscopic 3D, gaze tracking, gaze-contingent display, eye tracking, retargeting, remapping

Concepts: •Computing methodologies → Perception; Image processing;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. SIGGRAPH '16 Technical Paper., July 24 - 28, 2016, Anaheim, CA, ISBN: 978-1-4503-4279-7/16/07 DOI: <http://dx.doi.org/10.1145/2897824.2925866>

1 Introduction

A mental image of the surrounding world is built in the human visual system (HVS) by performing a sequence of saccadic eye movements and fixating at a sparse set of locations. This enables resolving fine spatial details in the fovea region of the retina, where the density of photoreceptors is highest, while it gradually reduces towards the retina periphery [Banks et al. 1991]. Gaze-contingent displays use eye tracking technology to monitor the fixation location and to conform the quality of depiction to the variable photoreceptor density. This way a more efficient control of the level of detail in geometric models [Murphy and Duchowski 2001], the image resolution in foveated rendering [Guenter et al. 2012], the state of luminance adaptation in tone mapping [Jacobs et al. 2015], the amount of blur in depth-of-field effects [Duchowski et al. 2014a], and the level of video compression [Geisler and Perry 1998] can be achieved, to name just a few key applications. When the fixation is shifted to another location the image content must be adjusted accordingly, and saccadic suppression [McConkie and Loschky 2002; Loschky and Wolverson 2007], i.e., cutting off conscious registration of the blurred retinal signal due to fast saccadic eye motion (up to 1000 deg/s), is employed to hide the visibility of such adjustments. This imposes stringent requirements on the overall latency in the rendering system, as well as on the precision and sampling rate of eye tracking, which is used for the next fixation prediction.

The strategy of gaze-driven content manipulation is in particular interesting in the context of stereoscopic displays [Duchowski et al. 2014a]. Due to the well-known accommodation-vergence conflict [Hoffman et al. 2008; Zilly et al. 2011; Shibata et al. 2011] the range of disparities that can be shown on such screens is limited. To alleviate the problem, stereoscopic content has to be carefully prepared and manipulated [Lang et al. 2010; Didyk et al. 2011; Oskam et al. 2011]. This usually includes an aggressive compression of the depth range that can be presented on the screen, and, as a result, flattens the entire scene.

To address this problem, we propose gaze-contingent disparity processing that preserves depth information as much as possible around the fixation point and compresses it everywhere else. The key idea behind these manipulations is that they are performed at the eye fixation stage and remain imperceptible. We conduct a series of psychophysical experiments and observe that the HVS is insensitive to relatively fast, but smoothly performed depth manipulations, such as local depth range changes, or bringing the fixation point to the screen surface during the actual fixation. Based on the experimental outcome, we build a model that predicts the speed with which

depth manipulations can be performed without introducing visible temporal instabilities. Such sub-threshold manipulations allow us to hide any latency issues of the eye tracker device, which makes our approach applicable even for low-cost devices. We investigate a number of applications for both real-time disparity optimization as well as offline content preprocessing. The contributions of this work are:

- a perceptual model that predicts the visibility of disparity manipulations during fixation,
- a metric of depth manipulation perceptibility,
- a real-time controller for adjusting stereoscopic content based on eye tracker data,
- a perceptual validation of the controller in terms of its seamless operation and local depth enhancement, and
- offline saliency-based solutions for disparity manipulation and scene cut optimization that improve viewing comfort.

2 Background

In this section we briefly discuss the dynamic aspects of the stereovision with special emphasis on the HVS limitations in perceiving depth changes. We consider two cases that fully determine the HVS operation modes in the response to such depth changes. When a target slowly changes its position in depth and screen space, the eye vergence is combined with a smooth pursuit eye motion to fuse the left and right images and maintain the target in the foveal region. When fixation switches between two targets that significantly differ in the depth and screen location, vergence must be combined with saccadic eye motion to achieve the same goal. In the former case, the speed of motion-in-depth (MID) is the key factor that determines the visibility of resulting depth changes (Sec. 2.1) and activates different eye vergence mechanisms (Sec. 2.2). In the latter case, the saccadic suppression is the dominant factor in hiding depth changes (Sec. 2.3).

2.1 Motion in depth

It has been suggested that the HVS sensitivity to the motion in depth (MID) speed expressed as vergence angle differentials is constant with distance to screen, and it follows Weber’s Law with 10% speed change being the just-noticeable difference [Portfors-Yeomans and Regan 1996]. This fraction gets higher for complex scenes due to the presence of monocular cues [Harris and Watamaniuk 1995]. The sensitivity to the MID is poorly correlated with the sensitivity to frontoparallel motion, but it is well correlated with the static disparity sensitivity under different base disparity and defocus conditions [Cumming 1995]. The HVS sensitivity to temporal disparity changes seems to be relatively low [Kane et al. 2014]. We exploit this property in our seamless remapping model.

A unique target with distinctive frontoparallel motion is known to have a very strong “pop-out” effect. However, no such effect was observed when disparity, as the only cue, induced MID of stereoscopically observed targets [Harris et al. 1998].

In this work, we complement these findings by measuring the HVS sensitivity for the speed of a scene depth range change as well as the speed of smoothly shifting the fixation point towards the accommodation plane (the screen plane) as relevant for stereoscopic displays.

2.2 Eye vergence

Eye vergence is performed through a fast and possibly imprecise transient (trigger) mechanism that is activated for larger depth changes as well as a slower and precise sustained (fusion-lock) mechanism

that compensates for the remaining fusion inaccuracies [Semmlow et al. 1986]. Slower depth changes with the ramp velocity below 1.4 deg/s can be fully processed by the sustained mechanism, while the motoric eye vergence (transient or sustained mechanisms) might not be even required for small depth changes that are within Panum’s fusional area, in which case sensoric fusion in the brain might be sufficient. For stereoscopic displays the eye vergence is excessively dragged towards the screen plane and at the screen depth the vergence error is smallest [Duchowski et al. 2014a]. This may be caused by the accommodation-vergence cross-link when the incorrect focus cue at the screen plane shifts the vergence towards the screen [Kim et al. 2014].

In this work, we control the speed of depth manipulations, so that only the sustained mechanism and the sensoric fusion are activated, which minimizes intensified efforts of the oculomotor system. At the same time, the eye vergence is kept as close to the screen plane as possible, which reduces the vergence error, the vergence-accommodation conflict, the frame violation effect [Zilly et al. 2011], and crosstalking between the left and right eye images [Shibata et al. 2011], and improves the viewing comfort and quality [Peli et al. 2001; Hanhart and Ebrahimi 2014].

2.3 Saccadic suppression

The HVS cuts off the sensory information during fast saccadic eye motion to avoid the perception of blurred retinal signal, which is referred as the saccadic suppression. The duration of actual saccadic eye motion depends on its angular extent and falls into the range of 20–200 ms. Each saccade is preceded by a preparatory stage with a latency of 200 ms, where new sensory information is cut off 80 ms prior to the eye motion initialization to the saccade completion [Becker and Juergens 1975]. McConkie and Loschky [2002] have shown that a switch from a significant blur to a sharp image can be detected by the observer even 5 ms after the end of a saccade. However, the tolerable delay can grow to up to 60 ms [Loschky and Wolverton 2007] for more subtle blur changes as in multi-resolution techniques [Guenter et al. 2012; Geisler and Perry 1998].

The eye vergence is a relatively slow process that for stereoscopic displays takes in total about 300–800 ms [Templin et al. 2014]. The actual time depends on the initial disparity and the vergence motion direction. In general, the motion towards the screen plane is faster than in the opposite direction with the maximum velocity of about 20 deg/s. The latency prior to the eyeball vergence initialization amounts to 180–250 ms [Semmlow and Wetzel 1979; Krishnan et al. 1973]. This effectively means that the eye vergence motion is typically continued after the saccade completion and approx. 100 ms are needed before new sensory information can actively guide the vergence to the target depth. In this respect, our seamless disparity manipulation shows some similarities as it may induce eye vergence motion after the saccade completion.

3 Previous work

In this section, we discuss the previous work on general disparity manipulations (Sec. 3.1) as well as the disparity processing techniques that are driven by gaze direction (Sec. 3.2). We also provide a brief overview of other gaze-driven applications (Sec. 3.3), and discuss the suitability of the saccadic suppression (Sec. 2.3) for hiding disparity manipulations.

3.1 Disparity manipulation

Disparity range compression is one of the most common disparity manipulations [Shibata et al. 2011; Zilly et al. 2011; Hoffman et al.

2008; Lang et al. 2010; Didyk et al. 2011] employed to avoid the accommodation-vergence conflict. Since this task shares many similarities with luminance range compression, standard tone mapping operators [Reinhard et al. 2010] can easily be adapted for disparity processing [Lang et al. 2010]. However, special care should be taken to avoid depth reversals when using local operators which for luminance could be explicitly used to expand the local contrast. Didyk et al. [2011] proposed a perceptual model for disparity that mimics the disparity processing done by the HVS and applies this for disparity manipulations. In the context of automultiscopic displays, the problem of extreme depth compression was addressed [Didyk et al. 2012; Masia et al. 2013; Chapiro et al. 2014]. These techniques aim at fitting disparities into a very shallow range taking care that the crucial disparity details are preserved. Such techniques can be also driven by additional saliency information. In the context of real-time solutions, simple but efficient methods include baseline and convergence manipulations [Jones et al. 2001; Oskam et al. 2011]. The common feature of all these techniques is maintaining good image quality in all regions regardless of the observer’s current gaze direction. In our work, we go beyond that and try to improve perceived quality based on the available gaze information. Additionally, we directly address the visibility problem of disparity manipulation.

3.2 Gaze-driven disparity manipulation

While the existing disparity manipulation techniques are successful on their own, their immediate adaptation for gaze-contingent setups is not an obvious task. In this section, we summarize the relatively few efforts addressing this problem, which are almost exclusively focused on shifting the whole disparity range with respect to the display plane (Fig. 2b).

Fisker et al. [2013] investigate the gaze-driven disparity adjustment towards the screen plane in an informal experiment and report promising results in terms of visual comfort and perceived depth. Bernhard et al. [2014] perform a full-scale experiment where an abrupt disparity adjustment to the screen plane is compared to the static disparity case in terms of the vergence time. In this case, shorter timings are desirable for reducing the viewing fatigue. They find that the abrupt disparity adjustment often increases the vergence time, and suggest that stereo fusion might require some readjustments in such conditions. This might be because the vergence facilitation effects due to peripheral vision processing is invalidated [Cisarik and Harwerth 2005]. It is worth noting that in that experiment, the actual saccade is triggered by the color change of a test square, whose destination is precisely known, and the disparity adjustment can be performed with a minimal latency. In practical applications, the problems reported by Bernhard et al. can be aggravated, and the visibility of the disparity change is more difficult to hide. For this reason, all disparity manipulations that we propose are performed in a seamless manner so that the stereo fusion is never disrupted and all latency issues are irrelevant.

The work of Peli et al. [2001] is conceptually close to our idea, and the authors measure the probability of detecting motion in depth for the fixated object when its disparity is shifted to zero at various speeds. The experimental method, however, limits the free exploration of the scene by the observer who is directly instructed to look at particular target locations. A similar setup with a virtual hand as a target controlled by a hand-tracking device is explored for virtual reality applications with head mounted displays [Sherstyuk et al. 2012]. Chamaret et al. [2010] uses a visual attention model to predict the new region-of-interest (RoI) to gradually reduce its disparity to zero. They experimentally derive the maximum disparity change that remains unnoticeable as 1.5 pixel steps. Hanhart and Ebrahimi [2014] extend this work by employing an eye tracker for determining the RoI. They assume a disparity change of 1 pixel

per frame without the frame-rate notion, and obtain favorable user judgments of such disparity manipulations when compared to the static disparity case.

We extend the work of Peli et al. [2001] by systematically measuring the just-noticeable speed of disparity changes for different initial disparity values. Our measurements are performed for continuous disparity shifts (Fig. 2b) rather than the discrete steps where the accumulated effect of disparity manipulation is not considered [Chamaret et al. 2010]. For the first time, we perform similar measurements for changes in the disparity range (Fig. 2c), and build a model that integrates both scaling and shifting of disparities. Based on the model we propose a novel gaze-contingent disparity mapping that enables seamless and continuous disparity manipulations, which significantly enhances the perceived depth.

3.3 Other gaze-driven applications

Gaze location tracking has also been used in other applications. In foveated rendering [Gunter et al. 2012], the efficiency of image generation can be improved by maintaining high image resolution only around the gaze location. The authors reported that seamless rendering can be achieved with a latency below 40 ms. In a similar way, chrominance complexity [Liu and Hua 2008] and level of detail [Murphy and Duchowski 2001] can also be gradually degraded with distance from the gaze location. Besides improving the rendering performance, the gaze location has been used to improve the image quality and the viewer experience. In the context of tone mapping, the luminance range can be used more effectively by reducing image contrast in regions that correspond to the peripheral vision [Jacobs et al. 2015]. Gaze-contingent depth-of-field effects have been modeled to improve the rendering realism [Mantiuk et al. 2011], reduce the vergence-accommodation conflict [Duchowski et al. 2014b], or enhance the depth impression [Vinnikov and Allison 2014]. Although all these techniques lead to either reduced costs in rendering or a better image reproduction, they often require the frame update to be strictly within the saccadic suppression period. In all cases the users express dissatisfaction if there is a noticeable lag due to the insufficient performance of the eye tracker or the rendering.

All these practical results clearly indicate that the use of saccadic suppression to hide the content change for the new fixation is very sensitive to the type of performed changes, the eye tracking precision, and the overall system latency. In gaze-contingent disparity manipulations, abrupt depth changes must be completed within the saccadic suppression, as new sensory information acquired afterwards actually guides the eye vergence motion (Sec. 2.3). On the other hand, the saccade must be effectively completed for a precise determination of the target depth, which leaves little room for fully informed scene depth manipulation. While there exist methods for predicting the saccade landing position based on some initial saccade direction and velocity measurements, e.g., using a ballistic model [Komogortsev and Khan 2008], any inaccuracies in this respect could be a serious hindrance for any gaze-driven disparity manipulation effort. For all those reasons, in this work we advocate seamless depth manipulations, when the new fixation is established.

4 Overview

In this work, we propose a new technique for manipulating stereoscopic content that accounts for the gaze information. To enhance perceived depth our method expands its range around the fixation location and reduces it in unattended regions that do not contribute significantly to depth perception. Additionally, objects around the fixation location are moved towards the screen to reduce artifacts such as visual discomfort (stereoscopic displays or virtual reality systems) or reduced spatial resolution (multi-view/lightfield displays).

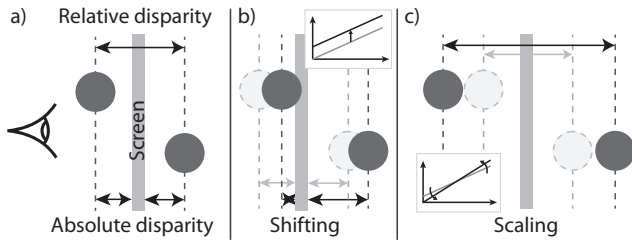


Figure 2: (a) Two objects as displayed in depth on a stereoscopic screen with their absolute disparity from the screen and the relative disparity between them. (b) Shifting of disparity moves both objects jointly, and thus changes absolute but preserves relative disparities. (c) Scaling of disparity changes mutual distances between objects and therefore both absolute and relative disparities.

The main challenge here is to apply manipulations that adapt to rapid changes in fixations on the fly. We identify the following requirements guiding our design:

- depth manipulations should be performed with a speed nearly imperceptible to the observer so that the manipulations do not interfere with artistic designs,
- as the fixation point can change unexpectedly, it should always be possible to quickly recover to a neutral depth that provides acceptable quality across the entire image.

To address these requirements, we first study the sensitivity of the HVS to the temporal disparity changes (Sec. 5). As most disparity manipulations can be approximated by local scaling and shifting of depth (Fig. 2), we limit our study to these two manipulations. Based on the data obtained in the perceptual experiment, we next demonstrate how the visibility of temporal disparity manipulations can be predicted (Sec. 6). We use the resulting visible disparity change predictor to derive a sequence of disparity mapping curves, so that the target disparity can be achieved seamlessly in a minimal number of discrete steps (effectively frames) for any input disparity map (Sec. 7). This enables a number of applications for such formulated seamless disparity manipulation (Sec. 8). Besides the main real-time application, in which eye tracking data is available (Sec. 8.1), we demonstrate a few scenarios where gaze information can be either provided beforehand or predicted (Sec. 8.2–8.3). Furthermore, we propose a metric that predicts the visibility of any disparity manipulation for all possible gaze directions (Sec. 8.4).

5 Sensitivity to disparity manipulations

In order to determine how fast disparity shift and scaling can be applied before an observer notices changes, we conducted two separate threshold estimation experiments that were guided by the QUEST procedure [Watson and Pelli 1983].

5.1 Experiment 1: Disparity shifting

The goal of the first experiment was to determine the minimum speed at which a continuous shift of disparity becomes visible to an observer.

Stimuli Each stimulus consisted of a flat, circular patch that was textured using a high number of easily visible dots (random dot stereogram – RDS). The size of an individual patch spanned 18 deg. To investigate the impact of the initial disparity, we considered 7 different starting disparities $d_s \in \{20, 10, 5, 0, -5, -10, -20 \text{ arcmin}\}$ that were measured with respect to the screen depth. An example of stimuli used in our experiments is presented in Fig. 3a.

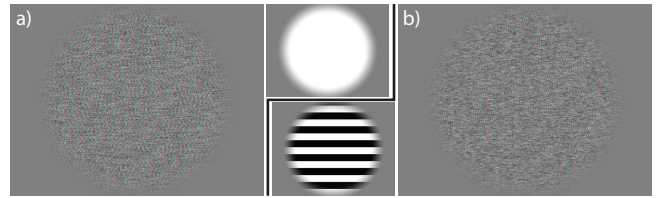


Figure 3: The random dot stereograms used in our experiments with disparity patterns in the middle. (a) Flat stimuli for Experiment 1. (b) Spatial corrugation for Experiment 2.

Task In order to measure the speed threshold, a two-alternative forced choice (2AFC) staircase procedure was used. At each trial a participant was shown two stimuli in randomized, time-sequential order. One of them was static while the other was moving in depth with constant velocity v_d . The direction of the motion was chosen to move the stimulus towards the screen as this is a likely scenario in a retargeting application. Each of the stimuli was shown for a period of 2.3 seconds, which was followed by 500 ms of a blank screen. The participant verged at the center of the stimulus and followed it as it moved in depth. The task was to decide which of the two stimuli contained motion or other temporal distortions and indicate the answer using arrow keys. The velocity of the moving stimuli was adjusted using the QUEST procedure. We chose to stop the staircase procedure when the standard deviation of the estimated threshold became smaller than 6.3% of the initial estimate. The range of v_d considered by the procedure was set between 1 and 60 arcmin/s, which was determined in a pilot experiment conducted on five subjects.

Equipment In both experiments, the stimuli were presented using the NVIDIA 3D Vision active shutter glasses on a 27" Asus VG278HE display with a resolution of 1920×1080 pixels, at a viewing distance of 80 cm under normal, controlled office lighting. We avoided depth distortion due to the time-sequential presentation by excluding any frontoparallel motion [Hoffman et al. 2011].

Participants 14 participants (2 F, 12 M, 23 to 27 years old) took part in both our experiments. All of them had normal or corrected-to-normal vision and passed a stereo-blindness test by describing the content of several RDS images. Each of them completed threshold estimation procedures for all d_s in a random order. The subjects were naïve with respect to the purpose of the experiment. The average duration of the whole experiment was one hour. Participants were offered a break and they could resume the experiment on the next day.

Results The results of the experiments are presented in Fig. 4a. We observed a large variance of stereo sensitivity between subjects as expected for a general population [Coutant and Westheimer 1993]. We decided for a general model although personalization would be an option. While our initial hypothesis was that the speed threshold depends on the initial disparity, an additional analysis of variance did not show any effect ($F(6,72) = 0.42, p = 0.42$). This verified that the initial vergence does not influence the sensitivity significantly, and therefore, we model the threshold as a constant. Due to the significant variance in performance of individual users (ANOVA: $F(13,65) = 4.07, p < 0.001$), we used the median of all values as an estimate of the sensitivity threshold (the dashed line in Fig. 4a). Consequently, we model the disparity change thresholds as a constant:

$$v_b = c_0, \quad (1)$$

where $c_0 = 17.64$ arcmin/s.

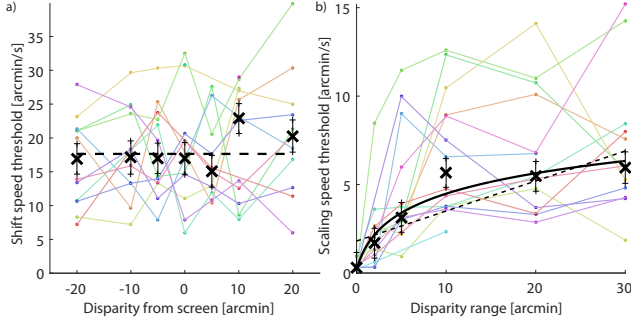


Figure 4: Results of Experiments 1 and 2 and our fitted model (a,b). Colors encode individual subjects; black crosses are median values across all subjects. (a) Thresholds as a function of the disparity from the screen (Experiment 1) with global median values shown as dashed line. (b) Thresholds as a function of the disparity range (Experiment 2) with both linear and logarithmic fit to median values.

5.2 Experiment 2: Disparity scaling

The goal of the second experiment was to measure how quickly the scene disparity range can be scaled before the temporal changes become visible.

Stimuli Similarly to the previous experiment, here we used a patch textured with a high contrast dot pattern. As we seek a speed threshold for disparity scaling, we considered a patch with a square wave disparity corrugation (Fig. 3b). To make our model conservative, we chose corrugation frequency to be 0.3 cpd as the HVS reaches its peak sensitivity for such a signal [Bradshaw and Rogers 1999]. We also used a square wave instead of sinusoidal one as it generalizes better for step functions [Kane et al. 2014] which successfully capture our manipulations that mostly occur between object edges. Because the sensitivity to disparity greatly depends on the amplitude of the disparity corrugation [Didyk et al. 2011], we consider different initial disparity ranges/amplitudes $d_a \in \{0, 2, 5, 10, 20, 30$ arcmin}. The values were chosen so that they do not result in a diplopia [Tyler 1975; Didyk et al. 2011]. The disparity corrugation was always centered around the screen plane, i.e., the average disparity of the patch is zero.

Task The procedure was similar to the previous experiment, with the exception that instead of the motion introduced to the entire patch, we introduced scaling to the disparity of the patch as a change of peak-to-trough amplitude over time. The maximum velocity that was considered by the 2AFC staircase procedure was set to 20 arcmin/s, and it was determined in a pilot experiment to be clearly visible. At such a speed diplopia could be reached during the exposure time of 2.3 seconds, but in practice, participants usually reported temporal change before this happened. Each participant performed one staircase procedure for each value of d_a in a randomized order.

Results The results of the experiments are presented in Fig. 4b. We observed a significant effect of the initial disparity range on the scaling speed threshold ($F(5,72) = 10.88$, $p < 0.001$) with a growing yet saturating tendency. The thresholds for disparity scaling are generally lower than for shifting. This is expected as disparity perception is driven mostly by the relative, not absolute, changes of depth. As a result, the sensitivity of the HVS to the relative disparity changes is much higher [Brookes and Stevens 1989]. The variance

between users is again significant ($F(13,64) = 2.14$, $p < 0.05$). Similarly as in the previous experiment, we used the median as an estimate of the thresholds (black crosses in Fig. 4b) to which we fit an analytic function. Because a linear function yields low DoF-adjusted $R^2 = 0.50$ and does not adequately describe the saturating shape visible in the data (dashed line in Fig. 4b), we use a logarithmic function which is known to be adequate for describing many mechanisms of the HVS. As a result, we model the disparity range change thresholds as a function of the disparity magnitude:

$$v_g(s) = c_1 + c_2 \cdot \log(s + 1), \quad (2)$$

where s is the disparity range size in arcmin and $c_1 = 0.1992$ and $c_2 = 1.787$ are the fitting parameters with DoF-adjusted $R^2 = 0.89$.

6 Visible disparity change predictor

Our disparity manipulation sensitivity model from the previous section predicts visibility of disparity changes for simple stimuli. To predict visibility of disparity manipulations for complex images, we define a predictor \mathcal{V} that for a given original disparity map $D_o : \mathbb{R}^2 \rightarrow \mathbb{R}$, two disparity mapping curves $d, d' : \mathbb{R} \rightarrow \mathbb{R}$, and a time $t : \mathbb{R}^+$ predicts whether the transition between the two curves in time t leads to disparity changes that are faster than the thresholds in Eq. 1 and Eq. 2. Formally, we define the predictor as:

$$\mathcal{V}(D_o, d, d', t) = \begin{cases} 1 & \text{if the transition is visible,} \\ 0 & \text{otherwise.} \end{cases}$$

In order to compute $\mathcal{V}(D_o, d, d', t)$, we have to check whether there is a location where either absolute or relative disparity (see Fig. 2a) changes become visible. The first case occurs if there exists a location \mathbf{x} for which the absolute disparity change is faster than the allowed speed in Eq. 1, i.e.,

$$\exists_{\mathbf{x} \in \mathbb{R}^2} \frac{|D'(\mathbf{x}) - D(\mathbf{x})|}{t} > v_b, \quad (3)$$

where $D'(\mathbf{x}) = d'(D_o(\mathbf{x}))$ and $D(\mathbf{x}) = d(D_o(\mathbf{x}))$. The second case occurs if there exist two locations \mathbf{x}, \mathbf{y} such that the relative disparity between them changes too fast (see Eq. 2), i.e.,

$$\exists_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^2} \frac{|\Delta D'(\mathbf{x}, \mathbf{y}) - \Delta D(\mathbf{x}, \mathbf{y})|}{t} > v_g(\Delta D(\mathbf{x}, \mathbf{y})), \quad (4)$$

where $\Delta D'(\mathbf{x}, \mathbf{y}) = D'(\mathbf{x}) - D'(\mathbf{y})$ and $\Delta D(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}) - D(\mathbf{y})$. With these two criteria, we can formulate our predictor as:

$$\mathcal{V}(D_o, d, d', t) = \begin{cases} 1 & \text{neither Eq. 3 nor Eq. 4 holds} \\ 0 & \text{otherwise.} \end{cases}$$

This definition holds for small values of t , as the relative disparity thresholds are a function of disparity magnitude (Eq. 2), which changes when different disparity mappings are applied. In our work, we assume that it is sufficient if t is equal to the period of one frame.

7 Seamless transition to target disparity

Our visibility prediction can be used to design a seamless transition between two disparity mapping curves d and d' . If the two disparity mappings are similar enough and $\mathcal{V}(D_o, d, d', t) = 0$ for t equal to the period of one frame, the transition can be done in one frame. However, this might not be the case if more aggressive disparity

manipulations are desired. In such cases, it is necessary to spread the transition over a longer period of time to maintain the speed of changing disparities below the threshold values. To this end, we have to construct a sequence of new disparity mapping curves that will be applied sequentially in consecutive frames. At the same time, we want to keep the transition time as short as possible. More formally, for a given original disparity map D_o , and two disparity mapping curves d and d' , we want to find a shortest sequence of disparity mapping curves $d_i : 0 \leq i \leq n$, one for each frame i , such that $d_0 = d$, $d_n = d'$, and $\forall_{1 \leq i \leq n} \mathcal{V}(D_o, d_{i-1}, d_i, t_i) = 0$. To make the construction possible, we assume that each curve d_i is an interpolation between d and d' . Consequently, we define each curve d_i using corresponding interpolation weights w_i as:

$$d_i(x) = (1 - w_i) \cdot d(x) + w_i \cdot d'(x). \quad (5)$$

It can be shown (see Appendix) that for this definition of intermediate curves the optimal weights defining the fastest seamless transition can be obtained as follows:

$$w_0 = 0, w_n = 1, w_i = w_{i-1} + \Delta w_i \quad (6)$$

$$\Delta w_i = \min_{\mathbf{x}, \mathbf{y} \in \mathbb{R}^2} \left(\frac{v_b \cdot t}{|D'(\mathbf{x}) - D(\mathbf{x})|}, \frac{v_g(\Delta D_{i-1}(\mathbf{x}, \mathbf{y})) \cdot t}{|\Delta D'(\mathbf{x}, \mathbf{y}) - \Delta D(\mathbf{x}, \mathbf{y})|} \right), \quad (7)$$

where t is the time of one frame. While different parametrizations of the transitions curves are possible, ours leads to a simple yet effective solution.

In order to construct the whole transition, we need to iterate Eq. 7 starting with $w_0 = 0$ until we reach $w_n = 1$. This is, however, computationally expensive as the evaluation of Eq. 7 requires iterating over all pixel pairs \mathbf{x} and \mathbf{y} , which leads to a quadratic complexity with respect to the number of pixels in the image. Instead, we propose a more efficient way of evaluating this equation by discretizing disparity maps into M values, so that there are only M^2 possible disparity pairs that we have to consider. If M is sufficiently large this will not create any accuracy issues. Assuming that the disparity range does not exceed -100 to 100 pixels, $M = 512$ results in errors not greater than $1/5$ of a pixel size. Consequently, we define an array H of size M such that $H[i] = 1$ if the disparity map D contains values between $\min(D) + i \cdot |\max(D) - \min(D)|/M$ and $\min(D) + (i + 1) \cdot |\max(D) - \min(D)|/M$, and $H[i] = 0$ otherwise. Later, to evaluate Eq. 7, we consider all indices $i, j < M$ such that $H[i] = H[j] = 1$, and we refer to the corresponding values of disparities p_i and p_j .

8 Applications

Disparity manipulations are often performed by stereographers who use them as a storytelling tool. At the same time, additional disparity manipulations are applied to reduce the visual discomfort or to find the best trade-off between the image quality and depth reproduction. We argue that the second type of manipulation should be performed in a seamless and invisible way, so it does not interfere with artists' intentions. In this section, we present applications of our model in different scenarios where such manipulations are crucial.

8.1 Real-time gaze-driven retargeting

In this section, we propose a real-time disparity manipulation technique that adjusts disparity information in the stereoscopic content taking into account gaze information. Our key insight is that depth information has to be accurate only around the fixation location and it can be significantly compressed in the periphery where depth perception is limited [Rawlings and Shipley 1969]. An additional

improvement can be achieved by bringing the attended part of the image close to the screen [Peli et al. 2001; Hanhart and Ebrahimi 2014]. We make use of our technique for creating seamless transitions between different disparity mappings to assure that our manipulations do not introduce objectionable temporal artifacts and are robust to sudden gaze changes. An additional feature of our solution is that because the temporal changes to disparities are seamless, the technique is immune to latency issues of the eye trackers.

At every frame, our technique takes as an input the original disparity map $D_o(\mathbf{x})$ together with the current disparity mapping function d_p , and the gaze location \mathbf{g} provided by the eye tracking system. Then it proceeds in three steps (Fig. 5). First, it constructs a candidate mapping curve $d_c : \mathbb{R} \rightarrow \mathbb{R}$ which is optimal for the current frame. Next, it restricts d_c to $d_t : \mathbb{R} \rightarrow \mathbb{R}$ such that a quick recovery to a neutral linear mapping d_l in case of saccade is possible. As the last step, the current disparity mapping $d_p : \mathbb{R} \rightarrow \mathbb{R}$ is updated to $d : \mathbb{R} \rightarrow \mathbb{R}$ which is a single step of the seamless transition from Sec. 7. The mapping d is then applied to the image.

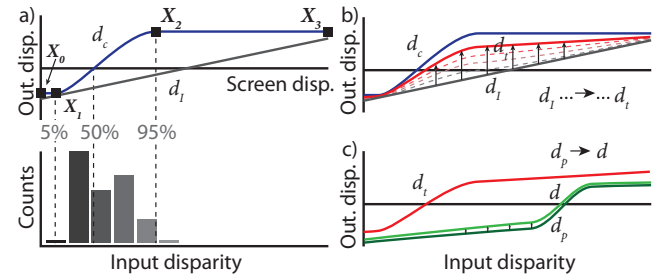


Figure 5: Construction of the disparity mapping for real-time retargeting. (a) A gaze-weighted disparity histogram is used to determine horizontal locations of the inner control points of the candidate curve d_c and the vertical offset necessary to minimize disparity from the screen in the gaze region. (b) An iterative transition algorithm finds the maximum transition from a neutral linear mapping d_l towards the candidate d_c in the defined time limit T_l as a target curve d_t . (c) The same algorithm is employed to update the previous curve d_p towards the target curve d_t and obtain a new mapping d to be used for the current frame.

Target curve construction To get the target curve d_t , we first build a candidate curve d_c parametrized by four control points $\mathbf{X}_{i \in 0..3} = [x_i, y_i]$ as presented in Fig. 5a. The two outer points \mathbf{X}_0 and \mathbf{X}_3 restrict the entire scene to the comfort range $[r_{c,0}, r_{c,1}]$ of the displayable disparity:

$$\mathbf{X}_0 = [\min(D_o), r_{c,0}]$$

$$\mathbf{X}_3 = [\max(D_o), r_{c,1}].$$

The two inner points \mathbf{X}_1 and \mathbf{X}_2 are responsible for the depth expansion around the gaze location. Therefore, their positions should span the range of disparities present around the fixation point. We define x -coordinates of \mathbf{X}_1 and \mathbf{X}_2 as the 5th (p_{05}) and 95th (p_{95}) percentile of the disparities around the gaze location. The percentiles are computed based on a histogram of D_o . To restrict its computation to the attended region and avoid temporal instabilities, we compute it as a weighted histogram, i.e., each disparity $D_o(\mathbf{x})$ contributes to the histogram according to the Gaussian $G_g(\mathbf{x}) = G(\|\mathbf{x} - \mathbf{g}\|, \sigma)$. Formally, we define the histogram H_G as:

$$H_G[i] = \sum_{\mathbf{x} \in R(i)} G_g(\mathbf{x}), \quad i = 0, 1 \dots M_G, \quad (8)$$

such that:

$$R(i) = \{x : \min(D_o) + i \cdot z \leq D_o(x) < \min(D_o) + (i + 1) \cdot z\}, \\ z = |\max(D_o) - \min(D_o)|/M_G.$$

The process of choosing the control points is presented in Fig. 5b. For the purpose of this paper we chose σ to be 2.5 deg, as the stereoacuity significantly declines with the retinal eccentricity beyond this point [Rawlings and Shipley 1969], and the histogram size $M_G = 512$.

Initially, the inner segment of the curve d_t is constructed to map the disparities of the attended region to the entire available disparity range, i.e., $\mathbf{X}_1 = [p_{05}, r_{c,0}]$ and $\mathbf{X}_2 = [p_{95}, r_{c,1}]$. This forces the rest of the curve to be flat, but can also lead to scaling relative disparities beyond their original values. To prevent this, we limit the expansion between X_1 and X_2 by restricting the slope of the curve to 1. We achieve this by shifting the two control points toward each other with respect to the midpoint between them. Consequently, we define the control points X_1 and X_2 as:

$$\mathbf{X}_1 = \left[p_{05}, \max(r_{c,0}, r_{c,0} + \frac{(r_{c,1} - r_{c,0}) - (p_{95} - p_{05})}{2}) \right] \\ \mathbf{X}_2 = \left[p_{95}, \min(r_{c,1}, r_{c,1} - \frac{(r_{c,1} - r_{c,0}) - (p_{95} - p_{05})}{2}) \right].$$

To bring the attended region close to the screen depth, we force the 50th (p_{50}) percentile of the disparities around the gaze location to map to 0. We achieve this by shifting all control points by p_{50} . The final control points are defined as:

$$\mathbf{X}'_i = \mathbf{X}_i - [0, p_{50}], \quad \text{for } i = 0 \dots 3. \quad (9)$$

To compute a smooth curve by the control points, we interpolate values between them using piecewise cubic Hermite interpolation and store the outcome in a discretized version using 256 bins.

Quick recovery guarantee Depending on the depth variation in the scene and the gaze location, the disparity mapping curve d_c may correspond to very drastic changes in depth. This is undesired because we want to maintain a good depth quality even after a sudden gaze change. We solve this problem by refining d_c in such a way that using our seamless transition strategy we can recover from it within a predefined time period T_l . To guarantee this quick recovery, we derive the final target curve d_t by constructing a seamless transition from an identity disparity mapping d_I (Fig. 5a) to the candidate mapping d_c according to Eq. 7, and defining d_t as the mapping that is achieved at time T_l (Fig. 5b).

Seamless transition Although d_t is built in every frame, in order to prevent sudden disparity mapping changes, it cannot be directly used. Instead, in each frame we execute a single step towards this curve. To this end, we use Eq. 7 to compute a single step of a transition between previous disparity mapping d_p and d_t (Fig. 5c). Finally, we use the resulting curve d to generate a stereo image presented to the user (see Fig. 6 and Fig. 9) using the image warping technique of Didyk et al. [2010].

8.2 Seamless disparity mapping in preprocessing

When the gaze location is not available, e.g., during post-production, our strategies can benefit from additional information about regions that are likely to be attended. For example, in movie production, it is common that attended image regions are known and purposely steered by a director. In other cases, a pre-viewing may be used to

gather such data. In this paper, we define this information as the probability distributions of gaze locations $S_k : \mathbb{R}^2 \rightarrow \mathbb{R}$ for each key frame $k \in [1, N]$. For the purpose of this paper, we estimate S_k using an image based saliency estimator proposed by Zhang et al. [2013]. We also assumed that the entire video sequence is available so we can optimize the disparity mapping curves across the whole content. The key idea of this method is to compute per-frame optimal disparity mapping curves (Fig. 5a), and then optimize them so the transitions between them (Fig. 5c) are seamless according to our model.

For every key frame k we build a desired mapping curve \hat{d}_k (Fig. 7a). To this end, we follow the suggestion of Lang et al. [2010] and first build a histogram of disparity $H_w(D_k)$ similarly to Eq. 8 but with $G_g(x)$ replaced by the saliency map S_k . We also compute a standard histogram $H(D_k)$ using the same formula but with a constant weight $1/N_D$, where N_D is the number of pixels in D_k . To account for different sizes of salient regions across S_k , we normalize $H_w(D_k)$ by $H(D_k)$ prior to deriving the mapping curve u_k as a cumulative sum:

$$u_k[i] = \sum_{j=0..i} \frac{H_w[j]}{H[j]}.$$

This mapping attributes a larger disparity range to salient regions. We then derive \hat{d}_k by scaling u_k to the displayable range $[r_{c,0}, r_{c,1}]$ and shifting it to the screen to minimize the expected disparity from the screen estimated as the 50th (p_{50}) percentile of $H_w(D_k)$:

$$\hat{d}_k = u_k \cdot (r_{c,1} - r_{c,0}) + r_{c,0} - p_{50}.$$

There is no guarantee that the series of \hat{d}_k results in seamless manipulations, as drastic changes between neighboring frames can occur. We address this problem by finding a sequence of curves such that it provides seamless disparity changes. To this end, we jointly optimize all curves d_k (Fig. 7b) according to the following strategy:

$$\begin{aligned} \text{minimize} \quad & E = |d_k - \hat{d}_k| \\ \text{subject to} \quad & \forall_{k \in [2, N]} \mathcal{V}(D_k, d_{k-1}, d_k, t_k - t_{k-1}) = 0 \\ & \forall_{k \in [1, N-1]} \mathcal{V}(D_k, d_k, d_{k+1}, t_{k+1} - t_k) = 0, \end{aligned}$$

where t_k is the time stamp of the k -th key frame.

We solve this problem iteratively. We initialize $d_{k,0} = \hat{d}_k$. In the i -th step we compute the new candidate curve d'_k as:

$$d'_k = (1 - \alpha)d_z + \alpha \cdot \hat{d}_k \\ d_z = \frac{d_{k-1,i-1} + d_{k+1,i-1}}{2},$$

where $\alpha \in [0, 1]$ is obtained by solving a convex 1D problem:

$$\begin{aligned} \text{maximize} \quad & \alpha \\ \text{subject to} \quad & \mathcal{V}(D_k, d_{k-1,i}, d'_k, t_k - t_{k-1}) = 0 \\ & \mathcal{V}(D_k, d'_k, d_{k+1,i}, t_{k+1} - t_k) = 0 \end{aligned}$$

using bisection. The mapping curve $d_{k,i}$ is then updated as:

$$d_{k,i} = (1 - \beta)d_{k,i-1} + \beta \cdot d'_k,$$

where $\beta = 0.1$ is a step size parameter. We stop the solver when $\max_k |d_{k,i} - d_{k,i-1}| < \epsilon$. We use $\epsilon = 0.1 \arccos$, which is achieved in less than 2 seconds for 50 key frames after ~ 100 iterations of our fast GPU solver running on a Quadro K2000M laptop GPU. If sparse key frames are used, then Eq. 5 is used to compute transitions between the intermediate frames. Samples from our results are presented in Fig. 7, and we refer readers to the supplemental video for a full demonstration.

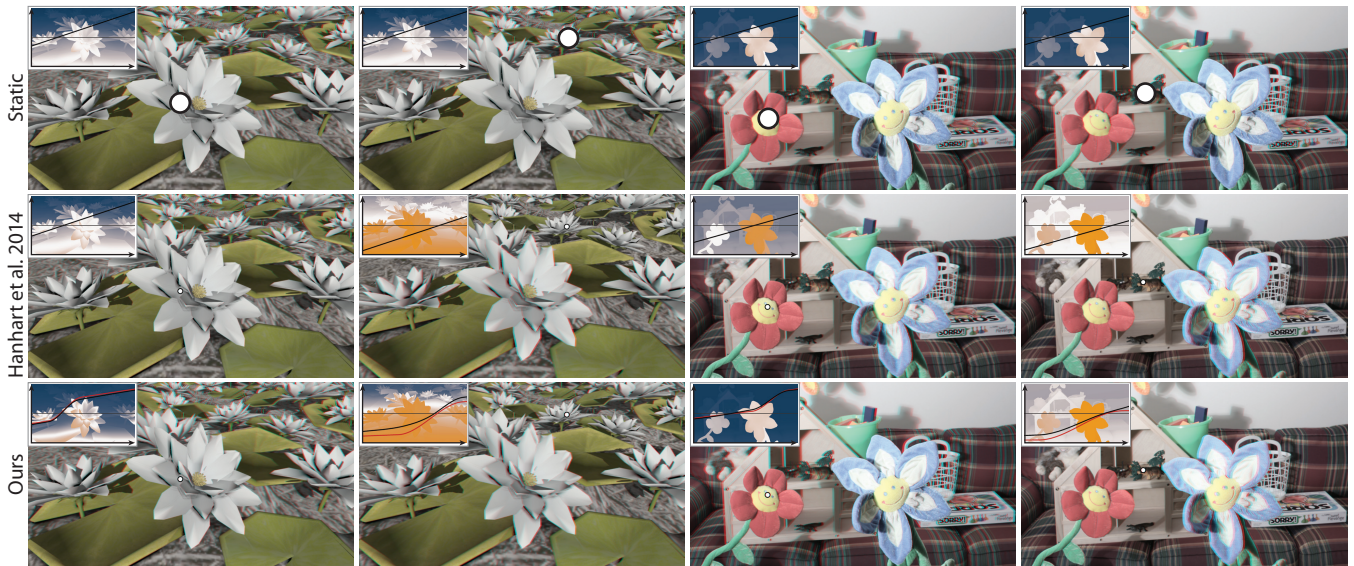


Figure 6: Comparison of our mapping (3rd row) with a static mapping (1st row) and the method of Hanhart and Ebrahimi [2014] (2nd row) as applied to our rendered image (left) and the image Flowers from the Middlebury dataset [Scharstein et al. 2014] for two different gaze locations (white dots). A disparity image with a mapping curve is shown in the insets. Crossed disparity is coded orange, uncrossed blue and screen disparity white. For our method the black curve is the rendered mapping d and the red curve is the target mapping d_t .

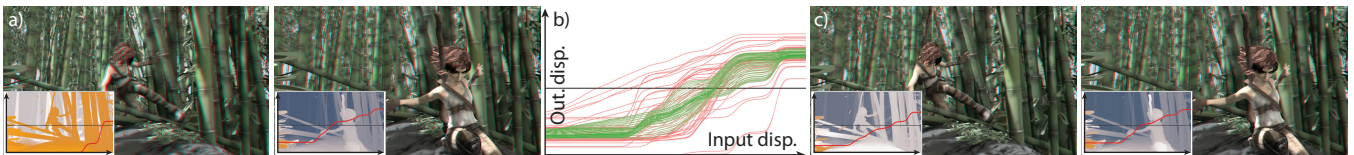


Figure 7: Our seamless transition applied to a per-frame saliency-based remapping [Lang et al. 2010]. (a) Two frames from the per-frame remapped sequence. (b) All per-frame (red) and our seamless transition (green) curves. (c) Our results. The Sintel video and disparity information are courtesy of the Blender Foundation and Butler et al. [2012], respectively.

8.3 Scene cut optimization

Templin et al. [2014] proposed an optimization for disparity transitions introduced by video cuts. They argued that minimizing the disparity difference at a cut reduces the eye vergence times and thereby improves the perceived image quality and scene understanding. To achieve this goal, disparity has to be retargeted on one or both sides of the cut and smooth transitions are required to blend to the original disparity. However, no precise model for such transitions was provided.

The seamless transition model for disparity mapping in Sec. 7 is well suited for this task. We optimize the cut by shifting disparities on both sides of the cut, which can be represented using a linear curve with a bias (see Fig. 2b). For simplicity, we assume that the time between subsequent cuts is always long enough to fit the entire mapping transition. Then, we can optimize each cut independently.

We first use the model of Templin et al. [2014] to find the optimal bias h_o of the pixel disparity maps D_c and D_{c+1} on both sides of the cut at frame c . We follow their suggestion and solve the problem by minimizing:

$$h_o = \arg \min_h \sum_{\mathbf{x}} S(\mathbf{x}) V \left(D_c(\mathbf{x}) - \frac{h}{2}, D_{c+1}(\mathbf{x}) + \frac{h}{2} \right),$$

where $S : \mathbb{R}^2 \rightarrow \mathbb{R}$ is equivalent to the attention probability map S_k from Sec. 8.2 for the frame c . We use a uniform estimate in our

examples. Function $V(a_0, a_1)$ stands for the vergence time model at the cut, where a_0 and a_1 denote the initial and target disparities [Templin et al. 2014]:

$$V(a_0, a_1) = \begin{cases} 0.04a_0 - 2.3a_1 + 405.02 & \text{if } a_0 < a_1 \\ -1.96a_0 + 3.49a_1 + 308.72 & \text{if } a_0 \geq a_1 \end{cases}.$$

Linear mappings d_c and d_{c+1} are then built for each of the two cut frames with respective disparity shifts $h_c = -\frac{h}{2}$ and $h_{c+1} = \frac{h}{2}$ (Fig. 2b). For every other frame i with time stamp t_i , we use our transition model (Eq. 7) to derive the corresponding mappings d_i as a transition to the original mapping d_0 :

$$\begin{cases} \text{from } d_c \text{ to } d_0 & \text{if } i \leq c \\ \text{from } d_{c+1} \text{ to } d_0 & \text{if } i > c \end{cases}$$

for the duration $T_i = |t_i - (t_c + t_{c+1})/2|$. We refer readers to the supplemental video for an example of the resulting mapping.

8.4 Visibility visualization

In stereo content production when no assumptions can be made about the attended image region, our predictor of disparity change visibility (Sec. 6) can be used directly as a metric for the evaluation of a disparity mapping.

As an input we assume either two disparity mapping curves d and d' from two different frames, or the same disparity frame mapped

by two different unknown curves as D and D' . The condition of the same frame can be relaxed if the distribution of physical depth in the scene does not change significantly over time. Additionally, we know the time span T between both inputs.

If only the mapped disparities D and D' are given, we construct the best approximation of the mapping curves between them rather than the mappings from the potentially unavailable original D_o (Sec. 6). The first curve d describes an identity mapping D to itself. The second curve d' describes a transition from D to D' and is constructed using a cumulative histogram, where each value from D' is accumulated to the bin corresponding to the value of D , and finally normalized by the number of accumulated values. The variance of values accumulated in each bin increases with a deviation from the global mapping assumption. The bins without any samples are filled by linear interpolation.

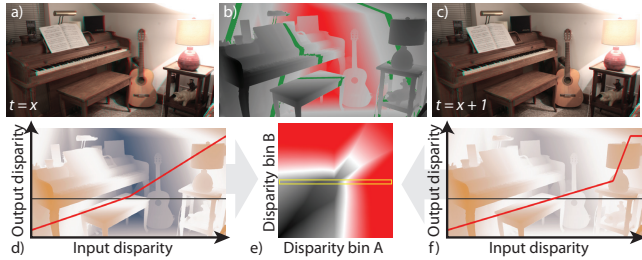


Figure 8: Output of our metric for 2 mappings as if transitioned over an interval of 1 second. (a, c) Boundary images. (d, f) Boundary disparity maps with their mapping curves. (e) The visibility matrix $Q(\mathbf{x})$ for every absolute and relative disparity. Red values ($Q > 1$) represent visible distortions. (b) Visualization of a single row of the matrix (yellow rectangle) as a distortion from each disparity pixel (red) with respect to the reference disparity (green) corresponding to the given row.

Now we can use the predictor $\mathcal{V}(D, d, d', T)$ to determine the visibility of a transition from d to d' in a binary way. To get the prediction in a continuous form, we can use our transition formula in Eq. 7 to compute the time T_c needed for a transition from d to d' as $T_c = n \cdot t$, where n is the number of discrete steps required. This allows us to formulate the metric score Q as the time needed relative to the time available:

$$Q = \frac{T_c}{T}.$$

The value units can be interpreted as just-noticeable differences (JNDs) and values lower than one can be considered imperceptible by the user, while values significantly larger can cause visible temporal artifacts as the depth is being transformed from one mapping to another.

We also have an option to evaluate the metric for every absolute and relative disparity separately. This way, each pair of disparity values mapped by d and d' defines two linear mapping functions for which Q can be computed the same way. Enumerating all such pairs leads to a matrix representation $Q(\mathbf{x})$ and allows for a detailed inspection of the mapping properties and guiding the user towards the source of distortions. See Fig. 8 for an example.

9 Evaluation

We evaluated the performance of our perceptual model and the disparity manipulation technique in a user experiment. To this end, we compared the depth impression and the temporal stability of our gaze-contingent disparity retargeting (Sec. 8.1) to three potential

alternatives: first, a traditional static mapping which does not adapt to the gaze location in any way; second, an immediate shift of depth which brings the depth in the gaze location to the screen without temporal considerations (similar to [Bernhard et al. 2014]); and finally, the method of Hanhart and Ebrahimi [2014] as discussed in Sec. 3.2. Our model was derived for simple stimuli. To test its performance on complex images that contain more complex depth cues, we tested three variants of our method with different multipliers for the speed thresholds in Eqs. 1 and 2. We chose multipliers 1, 2 and 4.

Stimuli The techniques were compared using both captured and CG content. 4 stereoscopic images from the Middlebury dataset 2014 [Scharstein et al. 2014] and 2 from our own rendering were used as stimuli (Fig. 6 and Fig. 9).

Task We compared the three variants of our method with each other as well as with all alternative methods in a 2AFC experiment. At each trial a participant was shown the question and then the two stimuli in randomized, time-sequential order. Both contained the same content but with the disparity mapped in two different ways. Each of the stimuli was shown for a period of 10 seconds, which was followed by 800 ms of a blank screen. The participant answered one of the following questions:

- Which demo has more depth?
- Which demo is more stable?

The user could choose to repeat the sequence at will.

Equipment The stimuli were presented using the polarized glasses technology on a 24" Zalman ZM-M240W display with a resolution of 1920×1080 pixels, at a viewing distance of 80 cm under normal, controlled office lighting. The display technology was chosen not to interfere with the eye tracker Tobii EyeX that was used for the gaze-adaptive mapping. A chin rest was employed to improve the tracking performance and to prevent the participant from moving away from the optimal viewing angle.

Participants 14 participants (2 F, 12 M, 23 to 27 years old) took part in the study. All of them had normal or corrected-to-normal vision and passed a stereo-blindness test by describing content of several RDS images. The subjects were naïve to the purpose of the experiment.

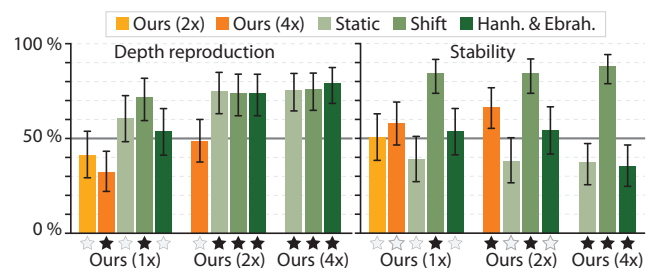


Figure 10: Results of our validation study for both the depth reproduction and stability questions. Each group of bars compares a variant of our method (multipliers 1, 2 and 4) against the other variants (warm colors) and competitor methods (green colors). 50% is a chance level. A value above 50% encodes participants' preference of the bottom label variant over the color-coded method. The error bars are confidence intervals. A significance in a binomial test is marked by a full star.

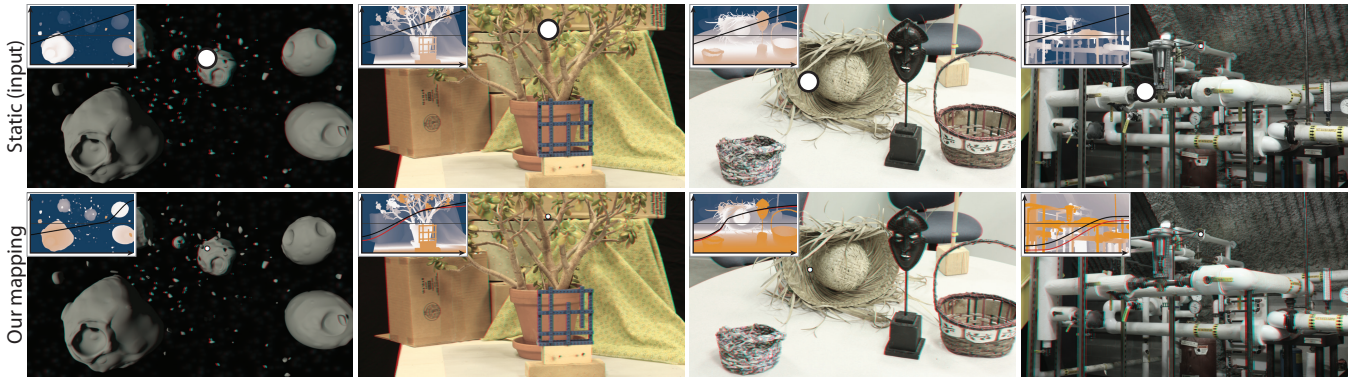



Figure 9:  The stimuli used in our validation experiment (see also Fig. 6). From left: Our CG rendering and 3 images from the Middlebury dataset [Scharstein et al. 2014].

Results We have evaluated relative preference for each compared pair and each question (Fig. 10). Our method achieves a significantly stronger depth reproduction than a simple disparity shift (71.4%, binomial test $p < 0.01$) with the threshold multiplier 1, and than both a static mapping (75.0%, $p < 0.01$) and the method of Hanhart and Ebrahimi (73.9%, $p < 0.01$) for the multiplier 2. There was no significant difference between the depth reproduction of our method with the multiplier 1 and 2. This shows that this comparison of depth was a difficult task for users and required substantial disparity differences to be accomplished above the chance levels. There was significantly less depth reported for the multiplier 1 than for 4 (31.7%, $p < 0.01$); therefore, using a larger multiplier generally results in greater perceived depth, as expected.

Our method is significantly more stable than an immediate shift to the screen for the multipliers 1 (84.3%, $p < 0.01$), 2 (84.3%, $p < 0.01$) and even 4 (87.7%, $p < 0.01$). This illustrates that the latency of current eye tracking systems make performing modifications during the saccadic suppression difficult. This further supports our choice of relying on seamless disparity processing at the fixation. There was no significant difference in stability with respect to a static mapping and the method of Hanhart and Ebrahimi except for the highest multiplier 4 (35.8%, $p < 0.05$ and 35.0%, $p < 0.05$ respectively). The trend towards lower stability reports in a comparison to the static mapping visible for the lower multipliers is expected, as a presence of any visible difference between two stimuli will likely lead to a statistically significant difference in answers after a sufficient amount of trials. The discrepancy between close-to-chance results for the comparison of our multipliers 1 and 2 and the method of Hanhart and Ebrahimi, and on the other hand significant difference for the multiplier 4, suggests that the actual stability for the two lower multipliers is good.

The results show that our method can deliver more depth without sacrificing stability. The statistically higher stability of the multiplier 2 compared to 4 (66.7%, $p < 0.01$) and at the same time insignificantly but consistently higher depth reproduction than the multiplier 1, confirms that the multiplier 2 is a better choice for a complex stereo content. This is in agreement with previous observations about thresholds measured on artificial stimuli and their validity for realistic images, e.g., when measuring the perceivable color differences in CIELAB and CIELUV [Reinhard et al. 2010] or disparity differences [Didyk et al. 2011]. Further, our experiments show that the choice of the stimuli for the model construction (Fig. 3) generalizes for complex images, as the manipulations stay seamless when multipliers 1 and 2 are used, but become quickly visible when multiplier 4 is considered.

10 Limitations

Our perceptual model accounts only for disparity changes around fixation location; it does not account for peripheral sensitivity to motion. Although in our experiments we did not observe any problems, it might be interesting to investigate peripheral vision in the future, especially for wide-angle VR systems.

The “pop-out” effect, which brings scene objects in front of the screen, is often used as a storytelling tool. Our technique preserves it for quick temporal disparity changes, but the effect may diminish after the re-adaptation. This might only be a concern for standard stereoscopic displays. In autostereoscopic displays a significant “pop-out” effect is usually avoided as it leads to aliasing problems [Zwicker et al. 2006]. In VR displays, the “pop-out” does not exist as there is no notion of “in front of the screen”.

Our techniques rely on several methods that may introduce additional artifacts. In particular, a poor estimation of visual saliency may lead to suboptimal results in our preprocessing application (Sec. 8.2). This is a general limitation of saliency-based manipulations, which can be improved by a director’s supervision or a pre-screening. The image warping technique used for generating our results can create monocular artifacts in disoccluded areas, if the disparity scaling is too large [Didyk et al. 2010]. This together with cross-talk or aliasing during large shifts can potentially introduce artifacts perceived as additional 2D cues which can further affect the visibility of our disparity manipulations.

11 Conclusions and future work

Gaze-contingent displays are gaining in popularity in various applications. Since such displays rely on the saccadic suppression to hide any required image changes from the user, their success strongly depends on the overall latency of the rendering system. In this work, we are interested in stereoscopic content authoring, which involves disparity manipulation, where the tolerability for the latency issues is very low. Our key insight is that near-threshold disparity changes can be efficiently performed at the eye fixation without being noticed by the user. This effectively makes the latency issues irrelevant. To this end, we measured the HVS sensitivity to disparity changes and formalize it as a metric. We employed the metric to guide the derivation of seamless transitions between frames in our gaze-contingent disparity retargeting. In this way, we improved the perceived depth significantly, while greatly reducing the requirements imposed on the eye tracker accuracy and latency. We also presented other applications of our metric in saliency-based disparity manipulations and scene cut optimization.

The benefits of our method extend beyond standard stereoscopic displays. New glasses-free 3D displays such as parallax-barrier or lightfield displays support only a relatively shallow depth range [Masia et al. 2013]. As a result, the visual quality quickly degrades for objects that are further away from the screen plane. Head-mounted displays have also recently gained a lot of attention and including eye tracking in these devices is a natural next step. We believe that our method can provide a substantial quality improvement in all these cases. Gaze-driven techniques targeting specific display devices that use our model are an exciting avenue for future work.

Acknowledgements

We would like to thank Junaid Ali, Thomas Leimkühler, Alexandre Kaspar, Krzysztof Templin, Louise van den Heuvel, Tobias Ritschel, and the anonymous subjects who took part in our perceptual studies. This work was partially supported by the Fraunhofer and Max Planck cooperation program within the framework of the German pact for research and innovation (PFI).

References

- BANKS, M., SEKULER, A., AND ANDERSON, S. 1991. Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *J Opt Soc Am A* 8, 11, 1775–87.
- BECKER, W., AND JUERGENS, R. 1975. Saccadic reactions to double-step stimuli: Evidence for model feedback and continuous information uptake. In *Basic Mechanisms of Ocular Motility and their Clinical Implications*, 519–527.
- BERNHARD, M., DELL’MOUR, C., HECHER, M., STAVRAKIS, E., AND WIMMER, M. 2014. The effects of fast disparity adjustment in gaze-controlled stereoscopic applications. In *Proc. Symp. on Eye Tracking Research and Appl. (ETRA)*, 111–118.
- BRADSHAW, M. F., AND ROGERS, B. J. 1999. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Res.* 39, 18, 3049–56.
- BROOKES, A., AND STEVENS, K. A. 1989. The analogy between stereo depth and brightness. *Perception* 18, 5, 601–614.
- BUTLER, D. J., WULFF, J., STANLEY, G. B., AND BLACK, M. J. 2012. A naturalistic open source movie for optical flow evaluation. In *European Conf. on Computer Vision (ECCV)*, Springer-Verlag, A. Fitzgibbon et al. (Eds.), Ed., Part IV, LNCS 7577, 611–625.
- CHAMARET, C., GODEFFROY, S., LOPEZ, P., AND LE MEUR, O. 2010. Adaptive 3D rendering based on region-of-interest. In *Proc. SPIE vol. 7524*, 0V–1–12.
- CHAPIRO, A., HEINZLE, S., AYDN, T. O., POULAKOS, S., ZWICKER, M., SMOLIC, A., AND GROSS, M. 2014. Optimizing stereo-to-multiview conversion for autostereoscopic displays. *Computer Graphics Forum* 33, 2, 63–72.
- CISARIK, P. M., AND HARWERTH, R. S. 2005. Stereoscopic depth magnitude estimation: Effects of stimulus spatial frequency and eccentricity. *Behavioural Brain Research* 160, 1, 88–98.
- COUTANT, B. E., AND WESTHEIMER, G. 1993. Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics* 13, 1, 3–7.
- CUMMING, B. G. 1995. The relationship between stereoacuity and stereomotion thresholds. *Perception* 24, 1, 105–114.
- DIDYK, P., RITSCHHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Adaptive image-space stereo view synthesis. In *Vision, Modeling and Visualization Workshop*, 299–306.
- DIDYK, P., RITSCHHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Trans. Graph. (Proc. SIGGRAPH)* 30, 4, 96.
- DIDYK, P., RITSCHHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. (Proc. SIGGRAPH)* 31, 6, 184.
- DUCHOWSKI, A. T., HOUSE, D. H., GESTRING, J., CONGDON, R., ŚWIRSKI, L., DODGSON, N. A., KREJTZ, K., AND KREJTZ, I. 2014. Comparing estimated gaze depth in virtual and physical environments. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*, 103–110.
- DUCHOWSKI, A. T., HOUSE, D. H., GESTRING, J., WANG, R. I., KREJTZ, K., KREJTZ, I., MANTIUK, R., AND BAZYLUK, B. 2014. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proc. ACM Symp. on Appl. Perc. (SAP)*, 39–46.
- FISKER, M., GRAM, K., THOMSEN, K. K., VASILAROU, D., AND KRAUS, M. 2013. Automatic convergence adjustment for stereoscopy using eye tracking. In *Eurographics 2013-Posters*, 23–24.
- GEISLER, W. S., AND PERRY, J. S. 1998. A real-time foveated multiresolution system for low-bandwidth video communication. In *Proc. SPIE vol. 3299*, 294–305.
- GUENTER, B., FINCH, M., DRUCKER, S., TAN, D., AND SNYDER, J. 2012. Foveated 3D graphics. *ACM Transactions on Graphics (Proc SIGGRAPH Asia)* 31, 6, 164.
- HANHART, P., AND EBRAHIMI, T. 2014. Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience. In *Proc. SPIE vol. 9011*, 0D–1–11.
- HARRIS, J. M., AND WATAMANIUK, S. N. 1995. Speed discrimination of motion-in-depth using binocular cues. *Vision Research* 35, 7, 885–896.
- HARRIS, J. M., MCKEE, S. P., AND WATAMANIUK, S. N. 1998. Visual search for motion-in-depth: Stereomotion does not ‘pop out’ from disparity noise. *Nature Neuroscience* 1, 2, 165–168.
- HOFFMAN, D., GIRSHICK, A., AKELEY, K., AND BANKS, M. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vision* 8, 3, 1–30.
- HOFFMAN, D. M., KARASEV, V. I., AND BANKS, M. S. 2011. Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth. *Journal of the Society for Information Display* 19, 3, 271–297.
- JACOBS, D., GALLO, O., A. COOPER, E., PULLI, K., AND LEVOY, M. 2015. Simulating the visual experience of very bright and very dark scenes. *ACM Trans. Graph.* 34, 3, 25:1–25:15.
- JONES, G. R., LEE, D., HOLLIMAN, N. S., AND EZRA, D. 2001. Controlling perceived depth in stereoscopic images. In *SPIE vol. 4297*, 42–53.
- KANE, D., GUAN, P., AND BANKS, M. S. 2014. The limits of human stereopsis in space and time. *The Journal of Neuroscience* 34, 4, 1397–1408.
- KIM, T., PARK, J., LEE, S., AND BOVIK, A. C. 2014. 3D visual discomfort prediction based on physiological optics of binocular

- vision and foveation. In *Asia-Pacific Signal and Information Proc. Assoc. (APSIPA)*, 1–4.
- KOMOGORTSEV, O. V., AND KHAN, J. I. 2008. Eye movement prediction by Kalman filter with integrated linear horizontal oculomotor plant mechanical model. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*, 229–236.
- KRISHNAN, V., FARAZIAN, F., AND STARK, L. 1973. An analysis of latencies and prediction in the fusional vergence system. *Am. J. Optometry and Arch. Am. Academy of Optometry* 50, 933–9.
- LANG, M., HORNING, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. (Proc. SIGGRAPH)* 29, 4, 75.
- LIU, S., AND HUA, H. 2008. Spatialchromatic foveation for gaze contingent displays. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*, 139–142.
- LOSCHKY, L. C., AND WOLVERTON, G. S. 2007. How late can you update gaze-contingent multiresolutional displays without detection? *ACM Trans. Multimedia Comput. Commun. Appl.* 3, 4, 7:1–7:10.
- MANTIUK, R., BAZYLUK, B., AND TOMASZEWSKA, A. 2011. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Serious Games Development and Appl.* 1–12.
- MASIA, B., WETZSTEIN, G., ALIAGA, C., RASKAR, R., AND GUTIERREZ, D. 2013. Display adaptive 3D content remapping. *Computers & Graphics* 37, 8, 983–996.
- MCCONKIE, G. W., AND LOSCHKY, L. C. 2002. Perception onset time during fixations in free viewing. *Behavior Research Methods, Instruments, & Computers* 34, 4, 481–490.
- MURPHY, H., AND DUCHOWSKI, A. T. 2001. Gaze-contingent level of detail rendering. *Eurographics Short Presentations*.
- OSKAM, T., HORNING, A., BOWLES, H., MITCHELL, K., AND GROSS, M. H. 2011. OSCAM-optimized stereoscopic camera control for interactive 3D. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 30, 6, 189.
- PELI, E., HEDGES, T. R., TANG, J., AND LANDMANN, D. 2001. A binocular stereoscopic display system with coupled convergence and accommodation demands. In *SID Symposium Digest of Technical Papers*, vol. 32, 1296–1299.
- PORTFORS-YEOMANS, C., AND REGAN, D. 1996. Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vision Research* 36, 20, 3265–3279.
- RAWLINGS, S. C., AND SHIPLEY, T. 1969. Stereoscopic acuity and horizontal angular distance from fixation. *J. Opt. Soc. Am.* 59, 8, 991–993.
- REINHARD, E., WARD, G., DEBEVEC, P., PATTANAIK, S., HEIDRICH, W., AND MYSZKOWSKI, K. 2010. *High Dynamic Range Imaging*. Morgan Kaufmann Publishers, 2nd edition.
- SCHARSTEIN, D., HIRSCHMÜLLER, H., KITAJIMA, Y., KRATHWOHL, G., NEŠIĆ, N., WANG, X., AND WESTLING, P. 2014. High-resolution stereo datasets with subpixel-accurate ground truth. In *Pattern Recognition*. Springer, 31–42.
- SEMMLOW, J., AND WETZEL, P. 1979. Dynamic contributions of the components of binocular vergence. *J Opt Soc Am A* 69, 639–45.
- SEMMLOW, J., HUNG, G., AND CIUFFREDA, K. 1986. Quantitative assessment of disparity vergence components. *Invest. Ophthalmol. Vis. Sci.* 27, 558–64.
- SHERSTYUK, A., DEY, A., SANDOR, C., AND STATE, A. 2012. Dynamic eye convergence for head-mounted displays improves user performance in virtual environments. In *Proc I3D*, 23–30.
- SHIBATA, T., KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *J. Vision* 11, 8, 11.
- TEMPLIN, K., DIDYK, P., MYSZKOWSKI, K., HEFEEDA, M. M., SEIDEL, H.-P., AND MATUSIK, W. 2014. Modeling and optimizing eye vergence response to stereoscopic cuts. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 33, 4.
- TYLER, C. W. 1975. Spatial organization of binocular disparity sensitivity. *Vision Research* 15, 5, 583–590.
- VINNIKOV, M., AND ALLISON, R. S. 2014. Gaze-contingent depth of field in realistic scenes: The user experience. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*, 119–126.
- WATSON, A. B., AND PELLI, D. G. 1983. QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics* 33, 2, 113–120.
- ZHANG, J., AND SCLAROFF, S. 2013. Saliency detection: a Boolean map approach. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*.
- ZILLY, F., KLUGER, J., AND KAUFF, P. 2011. Production rules for stereo acquisition. *Proc. IEEE* 99, 4, 590–606.
- ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3D displays. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques*, Eurographics Association, 73–82.

Appendix

Here we show how an optimal transition between two disparity mappings $d, d' : \mathbb{R} \rightarrow \mathbb{R}$ can be computed for an original disparity map D_o . For this purpose, as mentioned in the main text, we assume that the transition is defined as a sequence of intermediate disparity mappings $d_i : \mathbb{R} \rightarrow \mathbb{R}$, one for each frame. We specify each d_i as an interpolation between d and d' . Consequently, the transition is defined as:

$$d_0 \equiv d, \quad d_n \equiv d',$$

$$d_i(x) = (1 - w_i) \cdot d(x) + w_i \cdot d'(x) \quad 0 \leq i \leq n$$

$$w_0 = 0, \quad w_{i-1} \leq w_i, \quad w_n = 1,$$

where w_i is a sequence of the interpolation weights. This definition is equivalent to a formulation where disparity mappings are replaced with depth values from each stage of the transition:

$$D_0 \equiv D, \quad D_n \equiv D',$$

$$D_i(\mathbf{x}) = (1 - w_i) \cdot D(\mathbf{x}) + w_i \cdot D'(\mathbf{x}), \quad 0 \leq i \leq n,$$

$$w_0 = 0, \quad w_{i-1} \leq w_i, \quad w_n = 1,$$

for $D_i(\mathbf{x}) = d_i(D_o(\mathbf{x}))$. In order to make the transition seamless, we follow our visibility prediction described in Section 6 and obtain the following constraints that restrict absolute and relative disparity

changes:

$$\forall_i \forall_{\mathbf{x} \in \mathbf{R}^2} \frac{|D_i(\mathbf{x}) - D_{i-1}(\mathbf{x})|}{t} \leq v_b \quad (10)$$

$$\forall_i \forall_{\mathbf{x}, \mathbf{y} \in \mathbf{R}^2} \frac{|\Delta D_i(\mathbf{x}, \mathbf{y}) - \Delta D_{i-1}(\mathbf{x}, \mathbf{y})|}{t} \leq v_g(\Delta D_{i-1}(\mathbf{x}, \mathbf{y})), \quad (11)$$

where t is the period of one frame and $\Delta D_i(\mathbf{x}, \mathbf{y}) = D_i(\mathbf{x}) - D_i(\mathbf{y})$. Now let us consider the term $D_i(\mathbf{x}) - D_{i-1}(\mathbf{x})$. It can be shown that:

$$\begin{aligned} D_i(\mathbf{x}) - D_{i-1}(\mathbf{x}) &= \\ &= (1 - w_i) \cdot D(\mathbf{x}) + w_i \cdot D'(\mathbf{x}) - (1 - w_{i-1}) \cdot D(\mathbf{x}) - w_{i-1} \cdot D'(\mathbf{x}) \\ &= (w_i - w_{i-1}) \cdot (D'(\mathbf{x}) - D(\mathbf{x})) \end{aligned} \quad (12)$$

By substituting this into Eq. 10, we can show that the constraint on the absolute disparity changes is equivalent to:

$$\forall_i \forall_{\mathbf{x} \in \mathbf{R}^2} w_i - w_{i-1} \leq \frac{v_b \cdot t}{|D'(\mathbf{x}) - D(\mathbf{x})|} \quad (13)$$

Furthermore, using Eq. 12 we can also obtain:

$$\begin{aligned} \Delta D_i(\mathbf{x}, \mathbf{y}) - \Delta D_{i-1}(\mathbf{x}, \mathbf{y}) &= \\ &= (D_i(\mathbf{x}) - D_{i-1}(\mathbf{x})) - (D_i(\mathbf{y}) - D_{i-1}(\mathbf{y})) \\ &= (w_i - w_{i-1}) \cdot (D'(\mathbf{x}) - D(\mathbf{x})) - (w_i - w_{i-1}) \cdot (D'(\mathbf{y}) - D(\mathbf{y})) \\ &= (w_i - w_{i-1}) \cdot ((D'(\mathbf{x}) - D'(\mathbf{y})) - (D(\mathbf{x}) - D(\mathbf{y}))) \\ &= (w_i - w_{i-1}) \cdot (\Delta D'(\mathbf{x}, \mathbf{y}) - \Delta D(\mathbf{x}, \mathbf{y})) \end{aligned}$$

By substituting this into Eq. 11, we obtain a new form for the constraint on relative disparity changes:

$$\forall_i \forall_{\mathbf{x}, \mathbf{y} \in \mathbf{R}^2} w_i - w_{i-1} \leq \frac{v_g(\Delta D_{i-1}(\mathbf{x}, \mathbf{y})) \cdot t}{|(\Delta D'(\mathbf{x}, \mathbf{y}) - \Delta D(\mathbf{x}, \mathbf{y}))|} \quad (14)$$

By combining Eq. 13 and Eq. 14, we can obtain the weights w_i that define the shortest transition between d and d' , such that it does not violate the constraints in Eq. 1 and Eq. 2:

$$\begin{aligned} w_0 &= 0, \quad w_n = 1, \quad w_i = w_{i-1} + \Delta w_i, \\ \Delta w_i &= \min_{\mathbf{x}, \mathbf{y} \in \mathbf{R}^2} \left(\frac{v_b \cdot t}{|D'(\mathbf{x}) - D(\mathbf{x})|}, \frac{v_g(\Delta D_{i-1}(\mathbf{x}, \mathbf{y})) \cdot t}{|\Delta D'(\mathbf{x}, \mathbf{y}) - \Delta D(\mathbf{x}, \mathbf{y})|} \right) \end{aligned}$$