https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-501

Perceptual Display: Apparent Enhancement of Scene Detail and Depth

Karol Myszkowski, MPI Informatik
Okan Tarhan Tursun, MPI Informatik
Petr Kellnhofer, CSAIL MIT
Krzysztof Templin, MPI Informatik
Elena Arabadzhiyska, Saarland University, MMCI
Piotr Didyk, Saarland University, MMCI and MPI Informatik
Hans-Peter Seidel, MPI Informatik

Abstract

In this paper, we show how specific properties of human visual perception can be used for improving the perceived image quality, often without a need for enhancing physical display parameters. The focus of our work is on depth perception and improvement of viewing comfort on stereoscopic and multiscopic displays by exploiting complex interactions between monocular depth cues (such as motion parallax) and binocular vision. We also consider stereoscopic cinematographic applications, where proper handling of scene cuts, reflective/refractive objects, and film grain requires special attention to provide the required level of perceptual quality. Finally, we discuss gaze-driven depth manipulations to enhance perceived scene depth, and we present our predictor for saccade landing position which significantly reduces undesired effects of inherent system latency in foveated rendering applications.

Introduction

Over the last decade, stereoscopic 3D has become a mainstream feature in digital imaging, gaming, and film production. This, in turn, has triggered significant research efforts to improve the overall viewing experience and depth perception. A deeper understanding of 3D display limitations and the human perception has lead to smarter ways of content processing in this field. Traditional stereoscopic 3D as displayed on recent display devices may cause serious problems which negatively affect the viewing experience. For instance, the discomfort caused by rivalry or excessive disparity in junction with the so-called vergence-accommodation conflict can easily diminish potential advantages of binocular perception, causing fatigue and consequently shifting preference of viewers towards legacy 2D viewing. Perceptual modeling of the human visual system (HVS) is a way to address this problem and to optimize the content for the best viewing experience of both the general audience and individual observers by taking specific properties of display devices into account for personalization. In this work, we present a number of solutions for disparity manipulation that are relevant for modern multimedia systems. Such systems include virtual and augmented reality (VR/AR) technologies where stereoscopic displays are commonly used with head and eye trackers. The proposed solutions are suitable for real-time systems as well as high-quality off-line techniques targeted for cinematographic applications.

The HVS interprets 3D scenes by utilizing different depth cues [9, 27]. While a number of pictorial monoscopic cues contribute to this process, a complete stereoscopic impression requires the presence of more effective binocular cues such as the stereoscopic disparity which triggers vergence reflex of eyes. The well-known vergence-accommodation conflict puts serious constraints on the range of disparity cues that can be reproduced on stereo displays without causing visual discomfort [18, 24]. The main reason behind this limitation is the inconsistency between the image depth level reproduced by stereo disparity and the optical focus state of the eye which is fixed on the screen plane. Reproducing the accommodation cue can solve this problem [14], but currently available commercial solutions do not offer such a capability with sufficient level of quality. For example, light-field displays [29] support the eye accommodation; however, the objects which are positioned distant from the screen plane must be strongly blurred due to a limited number of views and this results in angular sampling deficit which otherwise would manifest as aliasing artifacts [51, 48]. The limit of reproduced depth ranges may be even more strict than for traditional stereo 3D displays. Because of the aforementioned reasons, it is crucial to take any possible opportunity that allows us to improve depth reproduction within the comfortable disparity range.

In this work, we present disparity manipulation solutions that optimize depth perception, depending on the scene content and gaze position. First, we discuss interaction of motion parallax and binocular disparity as important depth perception cues, and we present a perceptual model that predicts the perceived depth that results from such interaction. The model can be used to improve the depth reproduction on stereo 3D and automultiscopic displays without increasing disparity range. Second, we address selected problems of disparity manipulation in cinematographic applications with a common goal of reducing viewing discomfort while retaining realistic and artistically meaningful scene depiction. We start from the problem of disparity adjustment at scene cuts and then we focus on optimizing the eye vergence response, so that the time of clear vision and depth perception is maximized. Also, we briefly discuss challenges in reproducing film grain and rendering glossy materials. Third, we present a gaze-contingent method for disparity manipulation. The key idea of this method is to compress disparity at peripheral vision where the HVS sensitivity is reduced. At the same time, the disparity in foveal region is enhanced, and the spatial distance between fixated objects and the screen is reduced. For a natural viewing experience, these

manipulations are performed gradually and remain mostly unnoticed by the user. This requires the knowledge of the viewer's fixation point, and the scope of disparity manipulation depending on the duration of fixation time. Finally, we present our solution for saccade landing position prediction in gaze-contingent rendering, which allows more profound scene manipulations by taking the advantage of saccadic suppression to hide such manipulations from viewers.

Motion Parallax in Stereo 3D

According to a standard definition [33, 32], motion parallax is present when observer motion results in different retinal velocities of scene features at different depth. Motion parallax is a strong monocular depth cue, and its impact on depth perception is only slightly weaker than binocular disparity [9]. This fact has been exploited in several applications, such as *wiggle stereoscopy* [50] where motion parallax is used as a metaphor for stereoscopic images, or *parallax scrolling* [49] used in games where by moving foreground and background with different speeds depth sensation is evoked. A striking example of motion parallax efficiency are species that introduce subtle head movements to enable motion parallax [23]. This mechanism has been incorporated into cameras where apparent depth is enhanced by subtle motion of the sensor [37, 45].

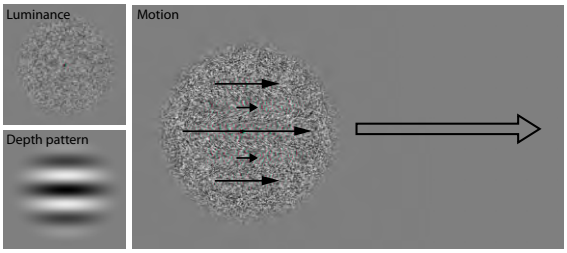

Figure 1: Stimulus used in our experiment in which we investigate depth perception induced by motion parallax. *Left*: Luminance and depth of our stimuli. *Right:* Anaglyph version of the same. For the dynamic stimulus variant, translating horizontally as indicated by the large arrow, motion parallax results in the flow depicted by the small arrows.

Perceptual Model Since motion parallax is such a strong depth cue that can be reproduced on a 2D screen without any limits, we consider to manipulate disparity signal according to the strength of motion parallax to improve the overall depth reproduction. We propose a computational model for detecting motion parallax and quantifying the amount of apparent depth it induces together with binocular disparity. To this end, we conducted a perceptual experiment where subjects were asked to match two stimuli according to the depth they provided (refer to Figure 1 for the stimulus example). One of the stimuli contained both motion parallax and binocular disparity, while the other had only a disparity signal. The matches provided by users allowed us to find the disparity signal that is equivalent to a given parallax-disparity combination.

The data from the experiment was used to derive an analytical model of perceived depth due to a combination of motion parallax and disparity (refer to Figure 2). The model maps a joint stimulus involving both the motion and disparity cues to

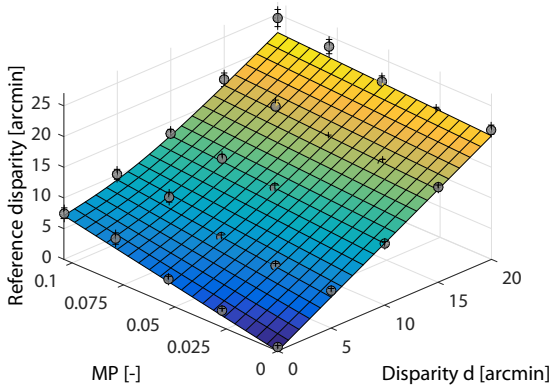

Figure 2: Our perceptual model showing the matched static disparity *(vertical axis)* as a function of binocular disparity *d* and relative depth from motion parallax *MP (x and y axis)*. *Gray dots* depict data acquired in our experiment.

the matching depth that is perceived using only disparity. As can be seen from Figure 2 motion parallax is a weaker depth cue compared to binocular disparity. Nawrot et al. [31] hypothesize that this is because in the latter case recovering depth information relies on the knowledge of fixed interocular distance, while an analogous processing for motion parallax involves the magnitude of eye movements possibly associated with head motion that can be estimated by the HVS with a much lower accuracy. However, since in stereoscopic displays disparity is usually strongly compressed to maintain viewing comfort [25, 11, 12], the apparent depth from motion parallax becomes comparable as it always refers to the original, uncompressed depth.

Stereo 3D Displays We exploit this observation and we propose a new disparity mapping operator which takes an advantage of our motion parallax model. The goal of the technique is to compress disparity range for regions that are supported by motion parallax, and use the additional disparity budget in static regions. As our goal is only to reallocate disparities, the technique can easily be combined with other existing disparity mapping approaches by applying them to our input disparity map beforehand. Figure 3 illustrates such disparity reallocation to the train interior, as strong motion parallax is present for the countryside view in the window and disparity that was originally allocated for this scene region can be strongly compressed without significant loss of depth perception.

Autostereoscopic Displays One of the main drawbacks of current autostereoscopic displays is a significant angular aliasing when a large depth range is shown on such a screen [51]. The aliasing originates from the limited angular resolution of the display, and it reveals itself as an unfavorable ghosting artifact which breaks the stereoscopic effect (Figure 4a). We can utilize our disparity manipulation technique to improve the image quality on such screens. To this end, instead of reallocating disparity signal as proposed before, we simply remove it from locations where motion parallax sufficiently supports depth perception. This reduces the required disparity range, and thus, the amount of visible aliasing.

We have tested our approach on a Full HD display Tridelity

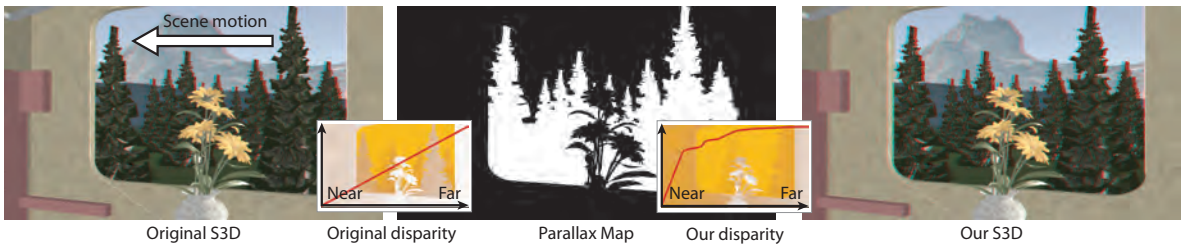

Figure 3: Starting from a stereoscopic video content with a static observer in a moving train *(Left)*, our method detects regions where motion parallax acts as an additional depth cue *(Center, white color)* and uses our model to redistribute the disparity depth budget from such regions *(the countryside)* to regions where it is more needed *(the train interior) (Right).*

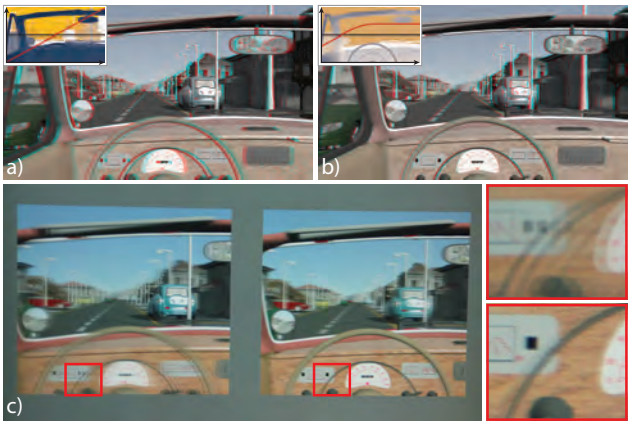

Figure 4: Example of disparity compression for autostereoscopic displays. a) The original linear compression with the corresponding disparity map and mapping curve shown in the inset. b) Our manipulation applied on the top of (a) to compress stronger regions that benefit from motion parallax(the street view). c) A photo of our display. Aliasing artifacts are strongly present in the content with only linear mapping (insets).

MV2600va that utilizes parallax barrier to deliver 5 views. We compared a linear mapping without and with our manipulation on the top (Figure 4). Disparities of both results are centered around the screen to minimize the aliasing of the display [51]. While the reference footage suffers from aliasing in the front and far parts of the scene, our output was able to convey a very similar depth impression without violating the usable depth range and introducing visual artifacts due to severe aliasing.

More details on analysis and quantification of structure perception from motion parallax for complex image sequences can be found in [22]. We believe that our model of depth induced by motion parallax in the presence of binocular disparity with different compression/enhancement levels (Figure 2) is a significant step towards better understanding of perception for new output devices such as head-mounted or lightfield displays where motion parallax is the inherent cue obtained from observers' head movements.

Disparity Manipulation at Film Cuts

The Hollywood style of combining shots developed into a set of formal conventions that obey dynamics of visual attention and control the continuity of space, time, and action. In modern movies cuts play the most important role (99% of all edits), while dissolves and wipes have vanished almost completely. An extensive analysis in [8] shows that average shot duration over past 75 years has declined from ca. 15 s to ca. 3.5 s! Clearly, short shots increase the viewer engagement by forcing eyes to quickly follow newly appearing content. However, this might be challenging in stereoscopic 3D movies, where unpredictable and large change in disparity means that binocular fusion is lost, and a confusing double image is seen (diplopia). Moreover, vergence system needs to quickly adapt to new conditions, in spite of the conflicting goal of the interconnected accommodation system [18, 24]. Clearly, such rapid disparity changes may lead to confusion, reduced understanding of the scene, and overall attractiveness of the content. In most cases the problem cannot be solved by simply matching the depth around the cut, as this would typically lead to excessive flattening the scene [42].

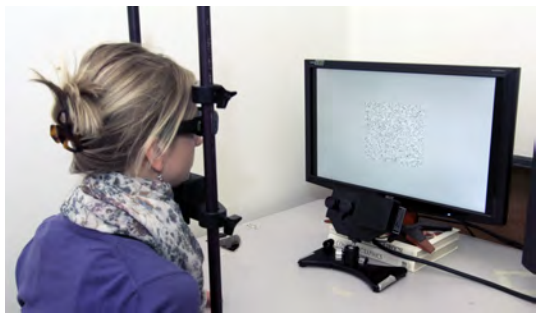

Figure 5: Setup used to measure eye vergence responses to step-like changes in disparity.

Perceptual Model To address the problem of rapid depth changes, we propose to relate the transition quality at the cut to vergence adaptation time, as it more directly measures the duration of imperfect vision state (diplopia, poor depth perception) than just the magnitude of disparity difference. This decision is justified by an observation that for the same absolute disparity differences the vergence adaptation time may be different as a function the initial and target disparity. To model this effect we measure the vergence response time using a high-end eye tracker (Figure 5) for a wide range of the initial disparities $d_i = 0, \pm 30, \pm 60, \pm 90$ px., as well as the target disparities ranging from ± 30 to ± 180 px. The "+" and "-" signs indicate here fixations behind the screen (uncrossed disparity) and in front of the screen (crossed disparity), respectively. The assumed disparity range corresponds approximately to the comfort zone in desktop viewing conditions [41, Fig. 23]. The stimulus in our experiment was a low-pass filtered white-noise patch (Figure 5) changing its

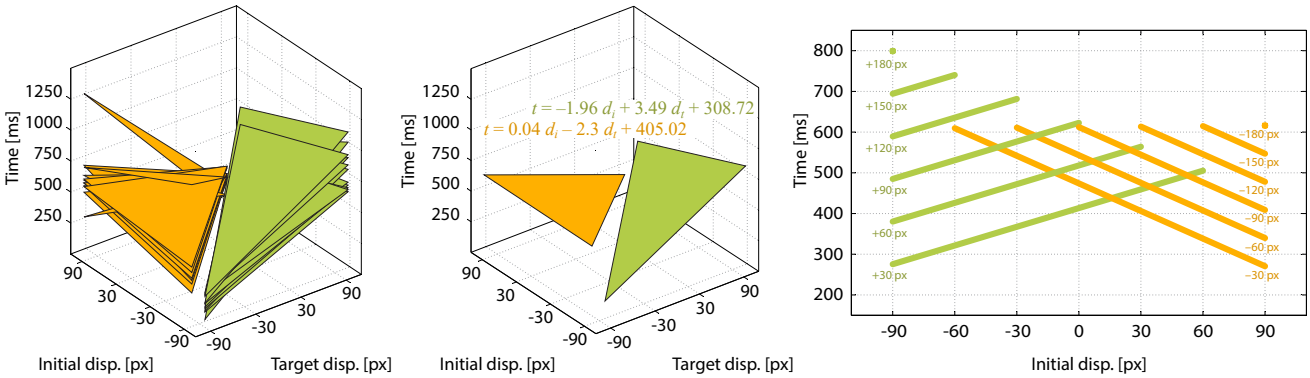

Figure 6: The vergence adaptation times as a function of the initial and target disparity as measured in the eye tracking experiment (Figure 5). The observer-specific *(left)* and average-observer *(middle)* fits to a two-plane model are derived with low standard deviation error. The right panel depicts a number of diagonal sections of the average planes *(middle)*. Each line represents disparity steps of the same magnitude and direction, but with different initial disparities.

disparity in the specified discrete steps over time. Figure 6-left shows the measured vergence adaptation times for a number of participants. In each case the vergence adaptation times can be almost perfectly modeled by two planes that represent convergence (*yellow*) and divergence (*green*). Figure 6-middle shows the average of all fits from the left panel, along with equations of the planes. We leave a gap between the planes, where times begin to increase due to Panum's fusional area and tolerance of the visual system to vergence errors. The diagonal is a singularity, where no vergence adaptation is needed, because the initial and target disparity are equal.

As expected, our measurements show that given the initial disparity and direction, steps with larger magnitude lead to longer vergence adaptation times (Figure 6-middle). An interesting finding is that the adaptation time depends also on the step direction and initial disparity. Given the initial disparity (Figure 6-right, abscissae) and step magnitude (one yellow and one green line per magnitude), steps *towards* the screen are generally faster: To the right of the graph, yellow lines (convergent steps) have lower times than the corresponding green lines (divergent steps). To the left, this is reversed. Note, that corresponding yellow and green lines intersect near the point of zero initial disparity (screen plane). We hypothesize that it is related to vergence-accommodation coupling, which attracts vergence towards the screen plane, where the vergence-accommodation conflict disappears.

Applications By using an off-the-shelf S3D display in our experiments, and validating our vergence adaptation model with real-world images [42], we ensure that the conditions are possibly similar to the ones in expected applications, where accommodation and pictorial cues may affect the vergence. In a subject's performance study aiming at 3D object recognitions tasks we demonstrated that by optimizing disparities before and after a cut we can reduce the eye vergence times and thereby improve the scene understanding. Figure 7 shows an example of interactive application that guides disparity retargeting, so that the vergence adaptation time is optimized for selected fixation points before and after the cut. The most likely fixation points can be measured, as eye scanpaths form highly repetitive patterns between different spectators for the same video sequences [47]. Moreover, skilled directors are capable of precise guiding and predicting viewers' attention. More details on this work can be found in [42].

Apart from optimizing the vergence adaptation time at the cut, another important problem is temporal frame-to-frame transition of required disparity retargeting, so that it smoothly blends with the original disparity. We address this problem in the following section.

Figure 7: Optimization of film editing operations to best match the human depth adaptation abilities based on the vergence response time model. The colors of the lines connecting points of interest before and after the cut visualize the corresponding vergence adaptation times. Pictures from *Big Buck Bunny* by Blender Foundation.

Film Grain and Specular Objects Other important problems in cinematographic stereo 3D applications that require disparity manipulation for comfortable depiction are *film grain management* as well as as handling *specular objects*.

Due to technical or artistic reasons, films often contain considerable amounts of film grain, and techniques for matching, adding, or removing it, play significant role in the post-production process. Intuitively, grain should be treated independently in each view of a stereoscopic image, however, the grain fusion by the HVS becomes difficult. A state-of-the-art solution projects the grain onto the geometry of the scene, but this approach has certain drawbacks of perceptual and aesthetic nature, e.g., the grain

confounds with the surface texture. To address this problem, a technique for volumetric scattering of grain in the proximity of scene surfaces has been proposed [43], which facilitates the grain fusion by the HVS and at the same time separates the scene information from the media.

The nature of problems with specular objects is essentially quite similar. Since highlights are view-dependent reflections of the light-sources in the scene, they have their own parallax and, consequently, their own position in depth, which differs from the position of the object they appear on. Moreover, the highlights often have inconsistent shape or topology across the views, which makes them difficult to fuse. Projecting consistent highlights onto the surfaces, is a simple approach that improves visual comfort, however, it decreases the realism of glossy object appearance. An intermediate solution, highlight microdisparity, which detaches highlights from the surface and places them in its proximity, has been proposed [44]. This way highlight fusion by the HVS becomes comfortable, while the glossy object appearance is not affected. The technique can be generalized to multiple reflections or other view-dependent effects, such as those observed in refractive media [10].

Gaze-driven Disparity Manipulation

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 [36, 38]. Additionally, objects around the fixation location are moved towards the screen [35] to reduce artifacts such as visual discomfort (stereoscopic displays or virtual reality systems) or reduced spatial resolution (multi-view/lightfield displays). 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. Since the HVS sensitivity to temporal disparity changes is relatively low [20] practically meaningful depth manipulations are feasible during the eye fixation period after the saccade is completed.

- 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.

Perceptual Model To address these requirements, we first study the sensitivity of the HVS to the temporal disparity changes. As most disparity manipulations can be approximated by local scaling and shifting of depth (Figure 8), we limit our study to these two manipulations.

In the first experiment we determine the minimum speed at which a continuous shift of disparity becomes visible to an observer (Figure 8b). The task was to decide which of the two stimuli contained motion or other temporal distortions (Figure 9a). The velocity of the moving stimuli was adjusted using the QUEST procedure. The results of the experiments are presented in Figure 10a. We observed a large variance of stereo sensitivity between subjects as expected for a general population [6]. While

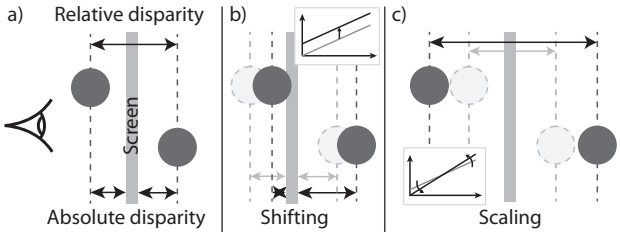

Figure 8: Basic disparity manipulation scenarios. (a) Two objects as displayed in depth on a stereoscopic screen with their absolute disparity from the screen and the relative disparity between them. (b) Shifting of disparity moves both objects jointly, and thus changes absolute but preserves relative disparities. (c) Scaling of disparity changes mutual distances between objects and therefore both absolute and relative disparities.

our initial hypothesis was that the speed threshold depends on the initial disparity, an additional analysis of variance did not show any effect ($F(6,72) = 0.42$, $p = 0.42$). Consequently, we model the disparity change thresholds as a constant:

$$v_b = c_0, \tag{1}$$

where $c_0 = 17.64$ arcmin/s.

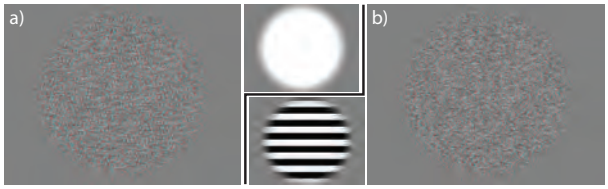

Figure 9: The random dot stereograms used in seamless disparity manipulation experiments: (a) Flat stimuli for Experiment 1. (b) Spatial corrugation for Experiment 2. For the reader convenience the actually used disparity patterns are depicted in the luminance domain in the middle.

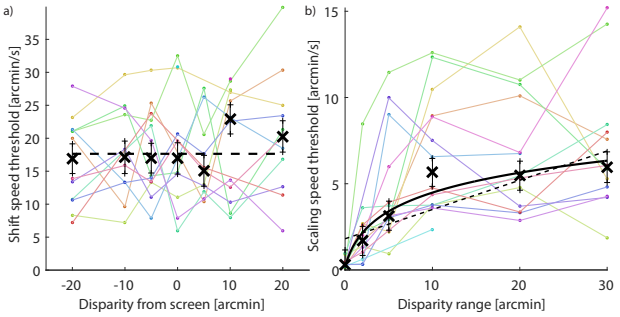

Figure 10: Results of Experiments 1 and 2 and our fitted model (a,b). Colors encode individual subjects; black crosses are median values across all subjects. (a) Thresholds as a function of the disparity from the screen (Experiment 1) with global median values shown as dashed line. (b) Thresholds as a function of the disparity range (Experiment 2) with both linear and logarithmic fit to median values.

In the second experiment we measured how quickly the scene disparity range can be scaled before the temporal changes become visible (Figure 8c). 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 dis-

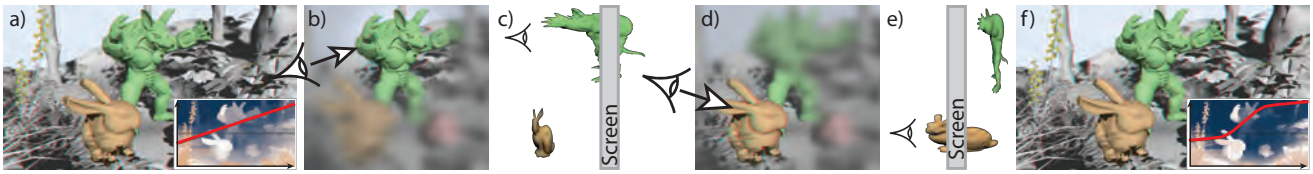

Figure 11: Beyond common disparity mapping (a) our approach adapts to attended regions such as the Armadillo (b) or Bunny (d) and modifies disparity in that region to enhance depth and reduce discomfort from the accommodation-vergence conflict (c,e). To that end we build non-linear remapping curves and use our novel perceptual model to ensure a seamless transition between them (f).

parity of the patch as a change of peak-to-trough amplitude over time (Figure 9b). The results of the experiments are presented in Figure 10b. 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 [4]. We model the disparity range change thresholds as a function of the disparity magnitude:

$$v_g(s) = c_1 + c_2 \cdot \log(s+1), \tag{2}$$

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 $R^2 = 0.89$.

Applications 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. We propose a real-time disparity manipulation technique that adjusts disparity information in the stereoscopic content taking into account gaze information (Figure 11). 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 [38] (compare the disparity mapping curves in the insets in Figure 11, a and f). An additional improvement can be achieved by bringing the attended part of the image close to the screen [35, 17] (compare the fixated object positions in Figure 11, c and e). 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. More details on our technique can be found in [21].

In the previous section we discussed disparity manipulation at the stereo 3D film cuts, but we ignored the problem of temporal frame-to-frame transition in required disparity retargeting. Our model for seamless disparity changes can be employed to address this issue by adjusting the disparity before and after the cut, so that frame-to-frame disparity changes always remain below the visibility threshold for all pixels (Figure 12). This way smooth blending of disparity in the proximity of the cut with the original disparity, which is likely to follow the artistic goals, can be achieved.

Latency issues and the range of depth manipulation 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 present we initiate our disparity manipulations after the new eye fixation is established. Such manipulations can be performed only during the eye fixation period and because of limits imposed on the speed of such changes that make them seamless, their range is somehow limited. In the following section we present our approach for the prediction of saccade landing position, which enables to perform more profound depth changes during the saccade by exploiting saccadic suppression of perceived visual information. In such case our seamless disparity manipulation could be used to compensate for possible prediction inaccuracies after the fixation point is already established.

Saccade Landing Position Prediction for Gaze-Contingent Rendering

Gaze-contingent rendering methods use the gaze position obtained from an eye tracker to adapt the rendered scene according to the perceptual capabilities of the HVS. Most prominent examples are the foveated rendering techniques [30, 16, 46, 34] which take advantage of the decay in visual acuity towards the periphery [15, 7, 3], and provide high image quality only for the fovea. A similar principle can be employed to improve realism and viewer experience by simulating depth-of-field effects [28] and local luminance adaptation [19], reducing the vergence-accommodation conflict during a stereoscopic presentation [13], or enhancing depth impression by gaze-driven disparity manipulations [21].

Although gaze-contingent rendering can lead to significant gains in computational cost and improvements in perceived visual quality, it is very sensitive to system latency [40]. When gaze location changes rapidly during fast eye motions called saccades [26], even short delays may result in visible artifacts which make the gaze-contingent rendering unfavorable. One solution to this issue may be to use an eye tracker which provides very high sampling rates and render the scene with high frame rates to minimize the system latency. We introduce another solution which exploits a natural phenomenon related to the HVS called the saccadic suppression [26]. Our solution is predicting the saccade landing position in real-time during the saccade, and rendering according to the predicted position on the display. The prediction is refined each time the eye tracker provides a new sample during the saccade. Rendering based on the landing position instead of the most recently sampled gaze position results in a mismatch between the actual gaze position and the rendering position during the saccade, but the saccadic suppression masks this mismatch. When the saccade ends, and the saccadic suppression loses its effect, the desired rendering result is already made available in the target gaze position as a result of our prediction. This approach success-

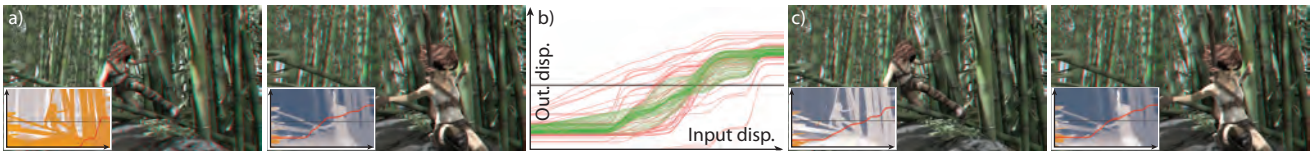

Figure 12: Our seamless disparity manipulation applied to a per-frame saliency-based remapping [25]. (a) Two frames from the per-frame remapped sequence. (b) All per-frame (red) and our seamless transition (green) curves. (c) Our results. The Sintel video and disparity information are courtesy of the Blender Foundation and the authors of [5], respectively.

fully alleviates the problem of system latency in gaze-contingent rendering tasks.

Our method consists of two steps. The first step is an offline data collection phase and a learning phase for the prediction model. The second step is the online detection of saccades and predicting the landing position for rendering. We use a velocity-threshold based (I-VT) saccade detection algorithm in both steps [1, 39]. To provide a robust detection, we choose a relatively large detection threshold $V_d = 130°/s$ and then search for the first gaze sample where the gaze velocity drops below $V_a = 60°/s$ within a temporal window of $30ms$. This point represents the beginning of the saccade and it is called the anchor point. The end of the saccade is detected with another velocity threshold, $V_f = 60°/s$. In order to include possible corrective motions of the eye called glissades [26], we consider samples which occur up to $15ms$ after this threshold is reached. Figure 13a shows these points on velocity and displacement profiles of a typical saccade. This figure also shows the noisy nature of saccade velocity profiles and multiple peaks during fixations that would lead to false detections if the detection of saccade beginning was based on only one velocity threshold.

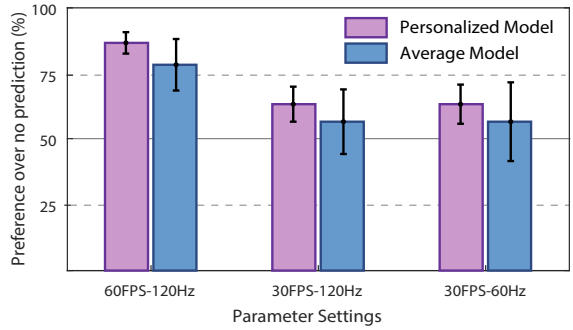

Figure 14: Results from the subjective free-viewing experiment where the participants compare standard gaze-contingent rendering with our method.

Saccades have specific acceleration, deceleration, peak velocity and duration characteristics depending on their amplitude. Due to the fast nature of saccades, the HVS does not have time to change the trajectory during a saccade because it takes approximately $70ms$ for visual information to travel from the retina to the oculomotor mechanisms of the brain [26, Ch. 3]. We show how these properties result in different displacement-time curves in Figure 13b for different saccade amplitudes. To model the saccade amplitude as a function of displacement and time, we collect gaze samples from 22 participants where each participant performs at least 300 saccades with different amplitudes during a 5-minute session. Among all gaze samples recorded in this experiment, the ones which correspond to saccades are detected and classified using the thresholds given above. Then we represent each saccade, S_k, performed by a participant as a set of such subsequent gaze samples:

$$S_k = \{s_{k0}, s_{k1}, s_{k2}, ..., s_{kN}\}, \tag{3}$$

where s_{k0} is the anchor point and s_{kN} is the end point of the saccade. Each one of the gaze samples is expressed in terms of triplets:

$$s_{kl} = \langle t_{kl}, d_{kl}, \theta_{kl} \rangle, \tag{4}$$

where t_{kl} is the timestamp, d_{kl} is the displacement and θ_{kl} is the direction of the sample, measured with respect to s_{k0}. According to this notation, the amplitude of the saccade S_k is defined as:

$$|S_k| = d_{kN}, \tag{5}$$

and the direction as:

$$\angle S_k = \theta_{kN}. \tag{6}$$

As most saccades follow a linear or approximately linear trajectory [26, Ch. 3], our estimate for the direction of the saccade, $\widehat{\angle S_k}$, at a point with timestamp t_{kl} is equal to the direction of the last observed gaze sample:

$$\widehat{\angle S_k}(t_{kl}) = \theta_{kl}. \tag{7}$$

In order to estimate the landing position, the task reduces to predicting the displacement in this gaze direction. We observe that the displacement profiles obtained from our experiment tend to form a consistent surface in the 3D space where the x- and y-axis are the displacement and time axes, while the z-axis corresponds to the amplitude values (see Figure 13c). Motivated by this observation, we treat the amplitude prediction as a regression problem, which is formally defined as:

$$\widehat{|S_k|}(t_{kl}) = f(t_{kl}, d_{kl}), \tag{8}$$

where we seek the optimal function f, minimizing the following squared error function:

$$\sum_{k,l} (|S_k| - f(t_{kl}, d_{kl}))^2. \tag{9}$$

We compare two different parameterizations for function f. The first one is interpolation of amplitude between collected gaze samples. For this approach, we preprocess the data by applying a 2D median filter to reduce the effect of noise, outliers and duplicates among the collected gaze samples. The second one is finding a polynomial fit whose degree is chosen according to cross-validation errors. According to our analysis, the best polynomial

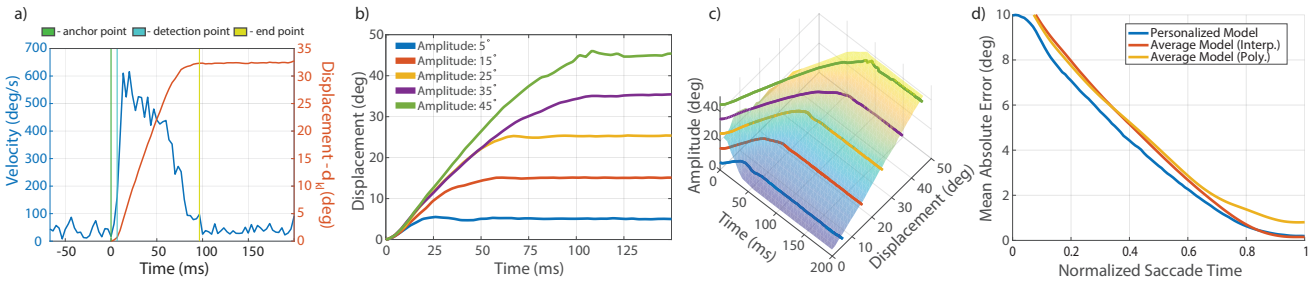

Figure 13: Anchor, detection and end points are shown on velocity and displacement profiles of a typical saccade with an amplitude of approximately 32° (a). Saccade motion has ballistic properties which exhibit a close relationship between displacement profiles and saccade amplitudes (b). A sample prediction surface obtained from the collected gaze samples is displayed in (c) (only 5 displacement curves are shown on the plot for simplicity). Improvement in the prediction accuracy over the course of saccades is shown in (d).

for the prediction of saccade amplitude is:

$$\begin{aligned} f_{poly}(t,d) = &-10.19t + 19.11t^2 - 17.15t^3 + 6.251t^4 \\ &- 0.7552t^5 - 23dt + 26.89dt^2 - 14.74dt^3 \\ &+ 3.882dt^4 - 0.4005dt^5 - 11.84d^2t \\ &+ 9.326d^2t^2 - 2.583d^2t^3 - 0.01058d^2t^4 \\ &+ 0.06587d^2t^5 + 20.18d + 5.071d^2 + 18.15, \end{aligned} \tag{10}$$

where $t = (t_{kl} - 47.39)/33.4$ and $d = (d_{kl} - 14.67)/11.72$ are the time and displacement measurements normalized with their respective means and standard deviations from the training data. Using this polynomial is a convenient way to predict the landing position when a simple implementation is desired. Otherwise, it is possible to obtain slightly more accurate predictions in terms of Mean Absolute Error by interpolating on collected gaze data (see Figure 13d). This figure also shows the improvement in the prediction accuracy when each participant is treated separately to find a personalized model. Although the personalized models provide the best accuracy, the average prediction model is a good alternative when it is not possible to record gaze data for personalization. The gaze samples collected during our experiments and interpolation grids which can be used for prediction are publicly available for the use of researchers on the project website of our work [2].

We validate the effectiveness of our saccade prediction method with a subjective free-viewing experiment, in which the observers compare the standard gaze-contingent rendering with our approach. The average preference of the observers significantly favors our method when the frame rate and sampling frequency settings are as low as 60FPS and 120Hz (see Figure 14). These settings correspond to capabilities of widely available eye trackers and graphics hardware on the market. In case of more significant system latencies, such as those which are longer than an average saccade duration, the benefit of making predictions start to diminish because the average length of saccades is not long enough to compensate for all of the system latency. But even in those extreme cases, our method is still favorable. Although our experiments are applied to foveated rendering, our method is easily applicable to other gaze-contingent rendering tasks as well. For details please refer to our original work [2].

Conclusion

In this work, we demonstrated that specialized models based on the perceptual capabilities of the HVS can be used as a guide for disparity optimization techniques for both the apparent enhancement of perceived depth ranges and overall improvement of 3D scene depiction quality.

We developed a model of perceived depth which can be used for the evaluation of different motion parallax and arbitrarily compressed disparity settings. To enhance overall depth perception, we showed an application of this model for smart allocation of disparity budget to different regions depending on the presence of supporting monocular depth cues such as the motion parallax. Also, we proposed a perceptual model that predicts the adaptation time of the ocular vergence reflex to rapid disparity changes in scene cuts. This model can guide disparity manipulation both before and after the scene cuts, which effectively minimizes the duration of reduced quality for scene and depth vision. In order to make disparity manipulations as seamless as possible, we measured the HVS sensitivity to basic disparity manipulations in temporal domain. The findings from these measurements allow keeping gaze-contingent disparity manipulations below the visibility threshold and prevent the display system latency from being a strict constraint on the rate of these manipulations. This approach assumes that the eye fixation is established and the disparity changes are seamlessly performed during fixation periods. Even more profound scene changes can be seamlessly performed by employing saccadic suppression, when the HVS sensitivity is very low. This requires a reliable estimation of saccade landing position, which also increases the limits on tolerable display system latency.

We believe that the perceptual models that we presented in this work have a more fundamental nature, although we focused mostly on their applied aspects.

References

[1] Richard Andersson, Linnéa Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, 2016.

[2] Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 36(4), 2017. http://dx.doi.org/10.1145/3072959.3073642.

[3] Martin S Banks, Allison B Sekuler, and Stephen J Anderson. Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *J. Opt. Soc. Am.*, 8(11):1775–1787, 1991.

[4] Allen Brookes and Kent A Stevens. The analogy between stereo depth and brightness. *Perception*, 18(5):601–614, 1989.

[5] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), editor, *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012.

[6] Ben E Coutant and Gerald Westheimer. Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics*, 13(1):3–7, 1993.

[7] Christine A Curcio, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson. Human photoreceptor topography. *Journal of Comparative Neurology*, 292(4):497–523, 1990.

[8] J.E. Cutting, K.L. Brunick, J.E. Delong, C. Iricinschi, and A. Candan. Quicker, faster, darker: Changes in hollywood film over 75 years. *i-PERCEPTION*, 2(6):569–76, 2011.

[9] J.E Cutting and P.M. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In William Epstein and Sheena Rogers, editors, *Perception of Space and Motion (Handbook Of Perception And Cognition)*, pages 69–117. Academic Press, 1995.

[10] Lukasz Dabala, Petr Kellnhofer, Tobias Ritschel, Piotr Didyk, Krzysztof Templin, Przemyslaw Rokita, Karol Myszkowski, and Hans-Peter Seidel. Manipulating refractive and reflective binocular disparity. *Computer Graphics Forum (Proc. EUROGRAPHICS)*, 33(2):53–62, 2014. http://dx.doi.org/10.1111/cgf.12290.

[11] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. A perceptual model for disparity. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 30(4):96, 2011.

[12] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, Hans-Peter Seidel, and Wojciech Matusik. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(6):184, 2012.

[13] Andrew T. Duchowski, Donald H. House, Jordan Gestring, Rui I. Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. Reducing visual discomfort of 3D stereoscopic sisplays with gaze-contingent depth-of-field. In *Proc. ACM Symp. on Appl. Perc. (SAP)*, pages 39–46, 2014.

[14] David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE Transactions on Visualization and Computer Graphics (Proc. VR)*, 23(4):1322–1331, 2017.

[15] O-J Grüsser and U Grüsser-Cornehls. Physiology of vision. In *Fundamentals of Sensory Physiology*, pages 144–198. 1986.

[16] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3D graphics. *ACM Trans Graph (Proc SIGGRAPH Asia)*, 31(6):164, 2012.

[17] Philippe Hanhart and Touradj Ebrahimi. Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience. In *Proc. SPIE vol. 9011*, pages 0D–1–11, 2014.

[18] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33–33, 2008.

[19] David Jacobs, Orazio Gallo, Emily Cooper, Kari Pulli, and Marc Levoy. Simulating the visual experience of very bright and very dark scenes. *ACM Trans Graph (TOG)*, 34(3):25, 2015.

[20] David Kane, Phillip Guan, and Martin S Banks. The limits of human stereopsis in space and time. *The Journal of Neuroscience*, 34(4):1397–1408, 2014.

[21] Petr Kellnhofer, Piotr Didyk, Karol Myszkowski, Mohamed M. Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. GazeStereo3D: Seamless disparity manipulations. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 35(4), 2016. http://dx.doi.org/10.1145/2897824.2925866.

[22] Petr Kellnhofer, Piotr Didyk, Tobias Ritschel, Belen Masia, Karol Myszkowski, and Hans-Peter Seidel. Motion parallax in stereo 3D: Model and applications. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH Asia)*, 35(6), 2016. http://dx.doi.org/10.1145/2980179.2980230.

[23] Karl Kral. Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behavioural Processes*, 64(1):1–12, 2003.

[24] Marc Lambooij, Marten Fortuin, Ingrid Heynderickx, and Wijnand IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009.

[25] Manuel Lang, Alexander Hornung, Oliver Wang, Steven Poulakos, Aljoscha Smolic, and Markus Gross. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 29(4):75, 2010.

[26] R John Leigh and David S Zee. *The neurology of eye movements*, volume 90. Oxford University Press, USA, 2015.

[27] Thomas Leimkühler, Petr Kellnhofer, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Perceptual real-time 2D-to-3D conversion using cue fusion. *IEEE Trans. Visualization and Computer Graphics*, 2017.

[28] Radosław Mantiuk, Bartosz Bazyluk, and Anna Tomaszewska. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Int Conf on Serious Games Dev & Appl*, pages 1–12, 2011.

[29] Belen Masia, Gordon Wetzstein, Piotr Didyk, and Diego Gutierrez. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, 37(8):1012 – 1038, 2013.

[30] Hunter Murphy and Andrew T. Duchowski. Gaze-contingent level of detail rendering. *Eurographics Short Presentations*, 2001.

[31] Mark Nawrot, Michael Ratzlaff, Zachary Leonard, and Keith Stroyan. Modeling depth from motion parallax with the motion/pursuit ratio. *Frontiers in Psychology*, 5, 2014.

[32] Mark Nawrot and Keith Stroyan. The motion/pursuit law for visual depth perception from motion parallax. *Vision Research*, 49(15):1969–1978, 2009.

[33] Mika E Ono, Josée Rivest, and Hiroshi Ono. Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3):331, 1986.

[34] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph (Proc SIGGRAPH Asia)*, 35(6):179, 2016.

[35] Eli Peli, T Reed Hedges, Jinshan Tang, and Dan Landmann. A binocular stereoscopic display system with coupled convergence and accommodation demands. In *SID Symposium Digest of Technical Papers*, volume 32, pages 1296–1299, 2001.

[36] Simon J.D Prince and Brian J Rogers. Sensitivity to disparity corrugations in peripheral vision. *Vision Research*, 38(17):2533–2537, 1998.

[37] Dennis Proffitt and Tom Banton. Perceived depth is enhanced with parallax scanning. Technical report, University of Virginia-Cognitive Science Department, 1999.

[38] Samuel C. Rawlings and T. Shipley. Stereoscopic acuity and horizontal angular distance from fixation. *J. Opt. Soc. Am.*, 59(8):991–993, 1969.

[39] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proc. Symp. on Eye Tracking Res. and Appl. (ETRA)*, pages 71–78, 2000.

[40] DR Saunders and RL Woods. Direct measurement of the system latency of gaze-contingent displays. *Behavior Research Methods*, 46(2):439–447, 2014.

[41] Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *J. Vision*, 11(8):11, 2011.

[42] Krzysztof Templin, Piotr Didyk, Karol Myszkowski, Mohamed M. Hefeeda, Hans-Peter Seidel, and Wojciech Matusik. Modeling and optimizing eye vergence response to stereoscopic cuts. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 33(4):1–8, 2014. http://doi.acm.org/10.1145/2601097.2601148.

[43] Krzysztof Templin, Piotr Didyk, Karol Myszkowski, and Hans-Peter Seidel. Perceptually-motivated stereoscopic film grain. *Computer Graphics Forum (Proc. Pacific Graphics)*, 33(7):349–358, 2014.

[44] Krzysztof Templin, Piotr Didyk, Tobias Ritschel, Karol Myszkowski, and Hans-Peter Seidel. Highlight microdisparity for improved gloss depiction. *ACM Transactions on Graphics (Proc. ACM SIGGRAPH)*, 31(4):1–5, 2012. http://doi.acm.org/10.1145/2185520.2185588.

[45] v3© Imaging. www.inv3.com, 2015.

[46] Karthik Vaidyanathan, Marco Salvi, Robert Toth, Tim Foley, Tomas Akenine-Möller, Jim Nilsson, Jacob Munkberg, Jon Hasselgren, Masamichi Sugihara, Petrik Clarberg, et al. Coarse pixel shading. In *High Performance Graphics*, 2014.

[47] Helena X. Wang, Jeremy Freeman, Elisha P. Merriam, Uri Hasson, and David J. Heeger. Temporal eye movement strategies during naturalistic viewing. *J. Vision*, 12(1):16, 2012.

[48] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):1–11, 2012.

[49] Wikipedia. Parallax scrolling — Wikipedia, the free encyclopedia, 2015. http://en.wikipedia.org/wiki/Parallax_scrolling [Online; accessed 2-June-2015].

[50] Wikipedia. Wiggle stereoscopy — Wikipedia, the free encyclopedia, 2015. http://en.wikipedia.org/wiki/Wiggle_stereoscopy [Online; accessed 2-June-2015].

[51] Matthias Zwicker, Wojciech Matusik, Fredo Durand, and Hanspeter Pfister. Antialiasing for automultiscopic 3D displays. In *Proc. of EGSR*. The Eurographics Association, 2006.