A Metric of Visual Comfort for Stereoscopic Motion

Song-Pei Du[1]   Belen Masia[2]   Shi-Min Hu[1]   Diego Gutierrez[2]

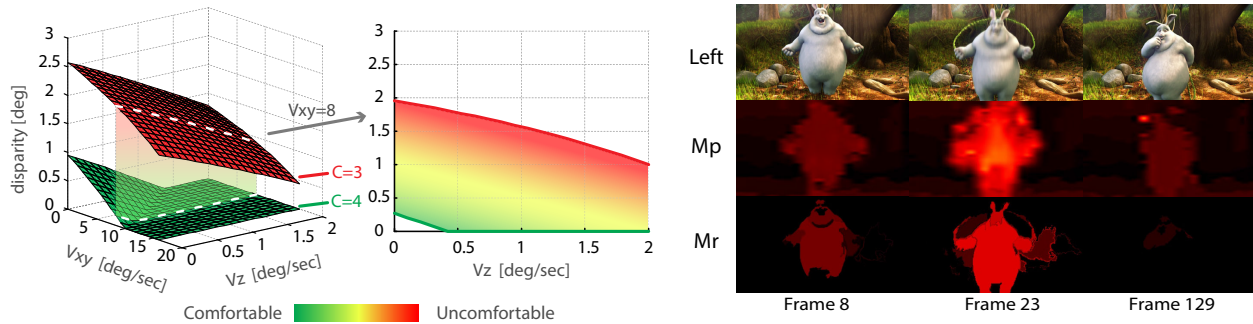[1] TNList, Tsinghua University   [2] Universidad de Zaragoza

Figure 1: *Our novel metric of visual comfort for stereoscopic content takes into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast. Based on our measured comfort function, we derive a metric to predict the degree of comfort for short stereoscopic videos. Left: Example slice of the comfort zone computed by our comfort function for a spatial frequency of* 1*cpd (bounded by comfort values of 3 and 4). Right: comfort maps computed using our metric on three representative frames of the* bunny *movie (©Blender Foundation). From top to bottom, input frames, per-pixel results, and per-region results (brighter red indicates less comfort). Our metric predicts less comfort with faster movement (frame 23), in agreement with the perceptual experiments.*

Abstract

We propose a novel metric of visual comfort for stereoscopic motion, based on a series of systematic perceptual experiments. We take into account disparity, motion in depth, motion on the screen plane, and the spatial frequency of luminance contrast. We further derive a comfort metric to predict the comfort of short stereoscopic videos. We validate it on both controlled scenes and real videos available on the internet, and show how all the factors we take into account, as well as their interactions, affect viewing comfort. Last, we propose various applications that can benefit from our comfort measurements and metric.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image generation—display algorithms, viewing algorithms;

Keywords: stereo, motion, visual comfort

Links: ◆DL 🗎PDF 🌐WEB

1 Introduction

Over the last few years, there has been a renewed interest in stereoscopic displays. Stereoscopic content is generated for movies, games and visualizations for industrial, medical, cultural or educational applications. This has in turn spurred research on aspects of the human visual system that relate to stereo vision [Pol-

lock et al. 2012]. Recent studies analyze the comfort zone for the vergence-accommodation conflict, the influence of luminance on stereo perception, or the depiction of glossy materials, to name just a few [Shibata et al. 2011; Didyk et al. 2012; Templin et al. 2012]. The goal is to understand different aspects of our visual system in order to produce stereo content that guarantees a comfortable viewing experience.

As opposed to natural viewing of the real 3D world, stereoscopic viewing implies conflicting vergence and accommodation cues, which is widely accepted to be a main cause of visual discomfort. However, despite recent advances and the extensive existing literature [Howard and Rogers 2002; Julesz 2006; Didyk et al. 2012; Didyk et al. 2011], some aspects of binocular vision remain largely unexplored. One of the main reasons is the large number of different factors involved, as well as their complex interaction [Cutting and Vishton 1995]. As a consequence, generating stereo content that guarantees a comfortable viewing experience remains a challenging task, often reserved to technicians with a large experience in the field [Lang et al. 2010; Mendiburu 2009].

Thus, one of the goals of stereography is to minimize the discomfort that stereoscopic viewing can cause, and numerous works have been devoted to explaining and characterizing the causes [Kooi and Toet 2004; Lambooij et al. 2009; Shibata et al. 2011]. However, fewer have explored how object *motion* affects this discomfort in stereoscopic viewing. Object motion in stereoscopic movies can in fact be a source of discomfort: Researches and experiments have revealed that visual comfort has a close relationship with some oculomotor functions, including eye movements induced by motion in the scene [Bahill and Stark 1975; Ostberg 1980]. In this work we analyze visual discomfort due to motion in short stereoscopic movies by means of a comprehensive statistical study. Unlike previous work [Yano et al. 2004; Jung et al. 2012], we take into account the interplay of motion velocity both on the screen plane and on the depth axis, as well as *signed* disparity and luminance spatial frequency. Our goal is not only to help understand the phenomena that may lead to visual discomfort; we provide a practical metric to assess existing 3D content as well. This can be used as a guideline for the generation of new stereo content, or to keep navigation pa-

rameters in virtual reality environments within comfortable limits, for instance.

Contributions: Specifically, we make the following contributions:

- We show that all the factors included in our study, as well as their interaction, do affect viewing comfort, and should be considered in the design of stereo content.

- We derive a statistical measurement that models the influence of motion, luminance spatial frequency, and signed disparity in visual discomfort.

- We propose a metric to predict potential comfort in short stereoscopic videos, and validate it by means of a user study.

- We propose several direct applications that could benefit from our measurements and metric, including a novel visual comfort zone for stereoscopic production, visualization techniques and stereoscopic retargeting.

Limitations: Although our measurements and metric are the most complete and exhaustive up to date, we do not aim at providing here the ultimate solution to this problem. Our methodology and results represent a solid step towards fully characterizing discomfort due to motion, and we hope that they will help others build more sophisticated models. However, there are a number of limitations that should be addressed by follow-up work. First, a comfortable stereo viewing experience may be related to many other factors not considered here, such as luminance contrast, disparity spatial frequency, viewing time, flicker, or imperfect content, to name a few [Kooi and Toet 2004; Mantiuk et al. 2011; Didyk et al. 2011; Hoffman et al. 2011; Cho and Kang 2012]. Taking all these factors and their interactions into account would make the problem intractable. Additionally, we limit our study to supra-threshold stimuli. Last, our metric is devised for short video sequences (up to 30 seconds in our results). This is convenient, since the average shot in modern TV and movies is only a few seconds. It would be interesting, however, to analyze how to extend our approach to longer sequences (even entire films) to take into account cumulative discomfort effects.

2 Related Work

Many previous works have investigated various aspects of stereoscopic perception (see for instance [Howard and Rogers 2002; Palmer 1999]). Recently, researchers have begun to explore the problem from the perspective of computer graphics and its applications. For instance, Templin et al. [2012] introduce a novel technique for stereoscopic depiction of glossy materials. Closer to our approach, Didyk et al. [2011] propose a model of disparity based on perceptual experiments, which is later extended to take into account the influence of luminance contrast [Didyk et al. 2012]. We also propose a measurement based on perception-driven studies, although we tackle a different problem, focusing on visual discomfort in the presence of stereoscopic motion.

Existing works have shown that visual discomfort in stereoscopy has a close relationship with oculomotor functions. It is widely accepted that the vergence-accommodation conflict is a key factor of visual discomfort, and that there exists a comfortable zone within which little discomfort occurs [Ostberg 1980; Hoffman et al. 2008; Tam et al. 2011]. In general, eye movement can be a source of discomfort when viewing stereoscopic content [Bahill and Stark 1975], which means that the motion component needs to be explicitly considered when measuring visual discomfort.

Kooi and Toet [2004] investigate various factors that may affect the visual comfort of viewing stereo images, including optical errors, imperfect filters and disparity. Hoffman et al. [2011] investigate the influence on the stereo viewing experience caused by flicker, motion and depth artifacts for various temporal presentation methods. Other works offer a quantitative measurement of visual comfort: Jin et al. [2005] evaluate the stereoscopic fusion disparity range based on the viewing distance and field of view of the display. Vertical misalignment has also been shown to affect visual comfort, and the maximum tolerable vertical misalignment has been measured as a unified metric based on different kinds of geometric misalignment [Jin et al. 2006]. Shibata et al. [2011] design a series of experiments to evaluate the zone of comfort for different vergence-accommodation combinations, while Yang et al. [2012] introduce a binocular viewing comfort predictor. None of them, however, consider motion. Lambooij et al. [2009] present a review of causes of visual discomfort, and conclude that visual discomfort might still occur within the so-called comfortable zone because of fast motion, insufficient depth information and unnatural blur.

Existing experiments have also confirmed the correlation between the velocity of moving objects and visual comfort [Yano et al. 2002; Yano et al. 2004; Ukai and Howarth 2008]. Speranza et al. [2006] investigate the relationship between visual discomfort and object size, motion-in-depth and disparity. Jung et al. [2012] introduce a novel visual comfort metric for stereoscopic video based on salient object motion, by computing three different discomfort functions for motion in horizontal, vertical and depth, respectively. The authors then use the mean or min operations to assess the global visual comfort, which is an ad-hoc solution for the co-occurrence of different motion components. Cho and Kang [2012] measure the visual discomfort as a function of disparity and viewing time for three levels of motion-in-depth (slow, medium and fast). Last, Li et al. employ pair-comparison experiments and propose a visual discomfort model based on the disparity and motion on the screen plane; they use both experts-only subjects using the Thurstone-Mosteller model [Li et al. 2011b], and non-experts subjects using the Bradley-Terry model [Li et al. 2011a].

All previous works on visual discomfort of stereoscopic motion either consider a single component of the motion vector, or simply combine conclusions obtained through separate experiments. In contrast, we offer a comprehensive study and systematically explore a larger parameter space, including the influence of the luminance spatial frequency, which is known to play an important role in depth perception. From our studies, we build a reliable measurement of visual comfort for stereoscopic motion, which we use to derive a predictive metric.

3 Methodology

In this section we describe the subjective experiments performed to measure the subjects' level of comfort when watching stereoscopic motion.

3.1 Parameter Space

As explained in Section 2, two key factors related to visual comfort in stereoscopic images and videos are the disparity value d and the velocity of motion $\mathbf{v} = (v_x, v_y, v_z)$, where subindices x, y refer to the screen plane and z indicates the direction perpendicular to the screen, i.e. depth. Recent studies have found that v_x and v_y have a similar effect on visual comfort [Jung et al. 2012]; we thus reduce the dimensionality of our problem by focusing on planar motion (v_{xy}) plus motion in depth (v_z). We measure d in terms of angular disparity α (in deg) while v_z and v_{xy} are the derivatives of the angular disparity and viewing direction β, respectively (in
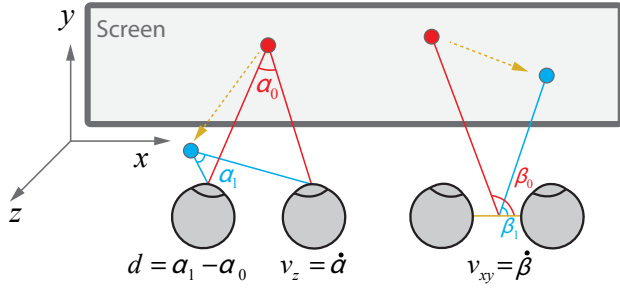
Figure 2: *Disparity and motion defined as functions of angular disparity α and viewing direction β. The viewing direction is defined from the middle interocular point.*

Figure 3: *Sample stimuli for the case of corrugation orientations $\psi = 0°$. Three successive frames are shown as anaglyphs. Under our viewing configuration, a $0.1°$ amplitude corresponds to about 7mm in depth.*

Figure 4: *Different combinations of d and v_z. There are no nonlinear motion components and no sign conversions of d at non-zero disparity. In each case, the upper and lower dashed lines represent $d = -2°$ and $d = 2°$ respectively.*

deg/sec). These parameters are shown in Figure 2. Intuitively, v_{xy} corresponds to a change in the gaze direction, while v_z corresponds to a variation of eye vergence.

The influence of luminance spatial frequency f_l on disparity perception is well known [Lee and Rogers 1997; Hess et al. 1999], and it was recently used to develop a perceptual disparity model [Didyk et al. 2012]. However, its influence on visual comfort for stereo motion remains largely unexplored; to overcome this, we add a parameter for luminance spatial frequency f_l in our experiments. Our parameter space is then four-dimensional: d, v_{xy}, v_z, f_l. Similar to other works, in order to reduce dimensionality, we fix the values of other parameters in our experiments.

3.2 Stimuli

Each stimulus consists of two-second animations, defined by a sinusoidal depth corrugation textured with a luminance image of noise of spatial frequency f_l. Each corrugation moves at a speed (v_{xy}, v_z). The mean disparity of the corrugation is d, and the amplitude of the sinusoid is fixed for all stimuli at $0.1°$ (6 arcmin), defined as the difference between mean and peak. In the case in which $v_z \neq 0$, and thus the mean disparity of the corrugation changes over time, d is defined as the mean disparity of the whole two-second stimulus. Previous work by Shibata et al. [2011] used 4 arcmin, but did not consider motion nor the influence of luminance spatial frequencies. We thus choose a slightly larger value which allows to clearly distinguish the corrugation. The corrugation's disparity spatial frequency is set to 0.3 cpd, which has been reported to be near the peak sensitivity of the human visual system [Didyk et al. 2011]. We sample each dimension of our parameter space as follows:

- $d = \{-2, 0, 2\}$ [°]
- $v_{xy} = \{0, 8, 16\}$ [°/sec]
- $v_z = \{0, 1, 2\}$ [°/sec]
- $f_l = \{1, 4, 16\}$ [cpd]

This makes a total of 81 different stimuli. Additionally, for each stimulus we explore four different corrugation orientations $\psi = \{0, 45, 90, 135\}$ [°], defined as degrees over the horizontal (see Section 3.3). The stimuli are shown on a fixed window at the center of the screen (the viewing angle of the window's diagonal is $22°$), surrounded by a 50% gray background. To obtain the stereo pair, we use image warping for the left and right views [Didyk et al. 2011]. This is done to avoid the "keystone" distortion in the toed-in camera configuration, and the perspective effect which would make the depth corrugations look non-uniform. We pre-compute the stimuli by warping offline, which works well in practice; no artifacts were
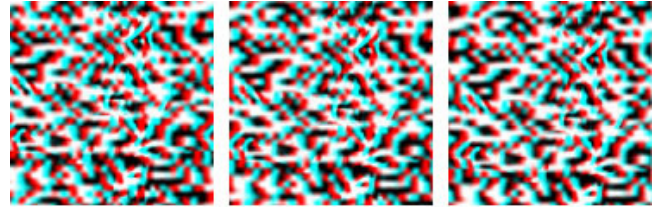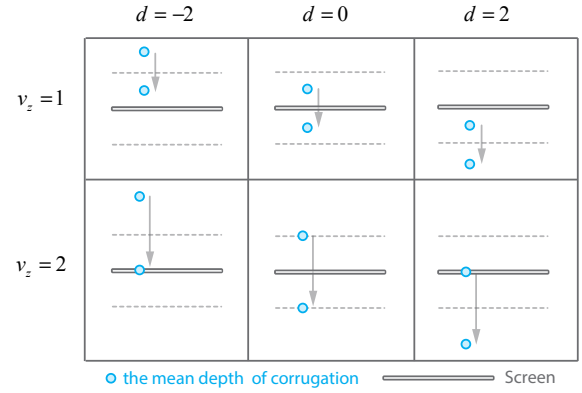
reported by the subjects.

We explore v_{xy} by moving the depth corrugation along the $45°$ diagonal. To avoid potential discomfort from nonlinear motion gradients [Jung et al. 2012], we only consider positive values of v_z, that is, motion towards the subject. Additionally, as noted by Speranza et al. [2006], zero-crossings in the disparity signal will affect the visual comfort: We thus limit depth corrugation motion in the z axis from $d - |v_z|$ to $d + |v_z|$ during the two-second span. Example stimuli are shown in Figure 3 while different combinations of d and v_z are illustrated in Figure 4.

3.3 Procedure

We use a 23-inch interleaved 3D display (1920 × 1080 pixels, 400 cd/m^2 brightness) with passive polarized glasses. The viewing distance is 50 cm and we assume the interpupillary distance to be 65 mm. Twenty subjects participated in our experiments, all with normal or corrected-to-normal vision, and with no difficulty in stereoscopic fusion. Their ages range from 20 to 30 years. Horizontal stripes are visible at this viewing distance and this should be universal for all such polarization displays. Subjects were aware of this and did not report complaints in the experiments.

The experiment is divided into 81 sub-sessions, which correspond to all possible combinations of our stimuli with fixed corrugation orientation ψ. In each sub-session, the subject is asked to rate their comfort level after doing a series of *visual oddity* tasks (three-interval, forced choice) [Shibata et al. 2011]. Specifically, for a given sample (d, v_{xy}, v_z, f_l), three stimuli are presented sequentially, with a 0.5 second break between stimuli during which a 50%

gray image is shown. Two of the three stimuli have the same corrugation orientation ψ, while the other has a $45°$ difference. For example, the three sequentially presented orientations may be $(0°, 0°, 45°)$, $(90°, 135°, 90°)$, $(45°, 0°, 45°)$, etc. The order and choice of orientations are random. After presenting the three stimuli ($2 \times 3 + 0.5 \times 2 = 7$ seconds), the subject is forced to select which stimulus had the odd orientation. Each sub-session contains ten such oddity tasks; after completing each sub-session, the subject is asked to rate their comfort level on a 5-point Likert scale, based on the following two questions:

- How do your eyes feel? (From 1 to 5: severe strain, moderate strain, mild strain, normal, very fresh).

- How comfortable was the viewing experience? (From 1 to 5: very uncomfortable, uncomfortable, mildly comfortable, comfortable, very comfortable).

Each sub-session takes about 80 seconds. We split the experiment into three parts and each part contains 27 sub-sessions with the same f_l value (1, 4, or 16 cpd); the three parts are done on three consecutive days. In each part the order of the 27 sub-sessions is random across subjects. To avoid accumulation effects [Cho and Kang 2012], subjects have to take a two-minute rest between sub-sessions, plus a longer, ten-minute break after 13 sub-sessions. They are nevertheless encouraged to take a longer rest if they want. For one subject, the whole experiment takes between 1.5 and 2 hours each day. Our choices are based on pilot tests performed before the regular experiments, which show that after an 80-second sub-session, subjects do perceive discomfort and that they can recover well after a two-minute break between sub-sessions.

4 Analysis

We first compute each subject's comfort score for every sub-session by averaging the two scores from the Likert scale, thus obtaining a total of $20 \times 81 = 1620$ scores. Then the comfort score for each session is computed as the average across the 20 subjects. To verify that these averages are statistically reliable, we perform a one-way repeated measures analysis of variance (ANOVA) of our data. This yields an F-value $F(80, 1520) = 24.01$, which is much larger that the F-test critical value for $p = 0.01$. This means the inter stimuli (intra subjects) variances are much larger than intra stimuli (inter subjects) variances, and thus our average scores are statistically reliable.

Our data are then used to determine our statistical measurement C of visual comfort for stereoscopic motion. We measure C as $C = C_{\mathbf{v},d} + C_{f_l}$, that is, a function of both the *combination* of velocity and depth, and luminance frequency. Previous works analyzing only velocity and disparity separately suggest that comfort seems to be approximately linear with those parameters [Li et al. 2011b; Jung et al. 2012]. Although Jung et al. [2012] fitted their model using logarithmic functions, the non-linear components are relatively small. Thus, we begin by using the following polynomial to fit $C_{\mathbf{v},d}$:

$$C_{\mathbf{v},d} = p_1 v_{xy} + p_2 v_z + p_3 v_{xy} v_z + p_4 dv_{xy} + p_5 dv_z + p_6 d + p_7 \quad (1)$$

Didyk et al. [2012] model a discrimination-threshold function s as: $s \approx 0.257 \log^2(f_l) - 0.3325 \log(f_l) + s(f_d, m_d)$ where f_d and m_d represent the frequency and magnitude of disparity respectively. To measure the influence of luminance spatial frequency f_l in visual discomfort, we therefore define C_{f_l} as a quadratic component of C:

$$C_{f_l} = p_8 \log^2(f_l) + p_9 \log(f_l) \quad (2)$$

Additionally, we want to explore how the *sign* of d affects the comfort score. We expand Equation 1 and include
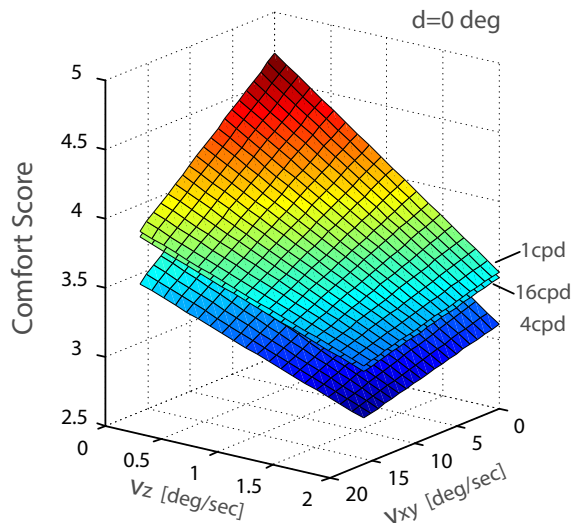

Figure 5: *Slices of our measurement function C. For $d = 0°$, three slices are shown corresponding to three different luminance frequencies: 1, 4 and 16 cpd. Higher comfort score refers to better visual comfort level predicted by our measurement. We provide the slices for $d = -2°$ and $2°$ in the supplementary material.*

(p_4^+, p_5^+, p_6^+) and (p_4^-, p_5^-, p_6^-) for $d \geq 0$ and $d < 0$, respectively. The resulting eleven-dimensional coefficient vector $\mathbf{P} = [p_1, p_2, p_3, p_4^+, p_5^+, p_6^+, p_4^-, p_5^-, p_6^-, p_7, p_8, p_9]$ is computed by solving the following quadratic optimization using linear least squares:

$$\arg \min_{\mathbf{P} \in \mathbb{R}^{10}} \sum_{i=0}^{80} (C(\mathbf{x}_i) - C_i)^2 \quad (3)$$

where $\mathbf{x}_i|_{i=0}^{80}$ are the 81 samples described in Section 3 and $C_i|_{i=0}^{80}$ corresponds to the 81 average scores across subjects. This yields a vector $\mathbf{P} = [$-0.0556,-0.6042,0.0191,0.0022,0.1833,-0.6932,-0.0043,-0.1001,0.2303,4.6567,0.9925,-1.1599$]$. The R^2 measure of goodness of fit is $R^2 = 0.9306$.

4.1 Discussion

Several slices of our visual comfort function C are visualized in Figures 5 and 6. Our measurement agrees with previous observations from existing works: Increasing disparity values (Figure 6(a)), motion on the screen plane (Figure 6(b)) or motion in depth (Figure 6(c)) introduce larger discomfort. Additionally, our measurement allows us to infer other important conclusions:

- The *sign* of the disparity also affects visual comfort (see Figure 6(a)). This effect was previously reported for the case of static stimuli [Shibata et al. 2011]; our experiments show that this behavior applies to stereoscopic motions as well. Additionally, we provide a quantitative measurement of this difference ($|p_6^+| = 0.6932$ and $|p_6^-| = 0.2303$).

- The *combination* of different values of v_{xy} and v_z has a strong influence in comfort ($p_3 = 0.0191$), as shown in Figure 6(b). Comfort decreases differently as v_{xy} and v_z increase ($p_1 = -0.0556$, $p_2 = -0.6042$). In particular, the influence of v_{xy} in viewing discomfort diminishes as v_z increases.

- Last, luminance spatial frequency f_l is a non-linear factor in viewing comfort. For $f_l \in [1 \text{ cpd}, 16 \text{ cpd}]$ in our experiments, the comfort score has a minimum near 4 cpd ($p_8 = 0.9925$,
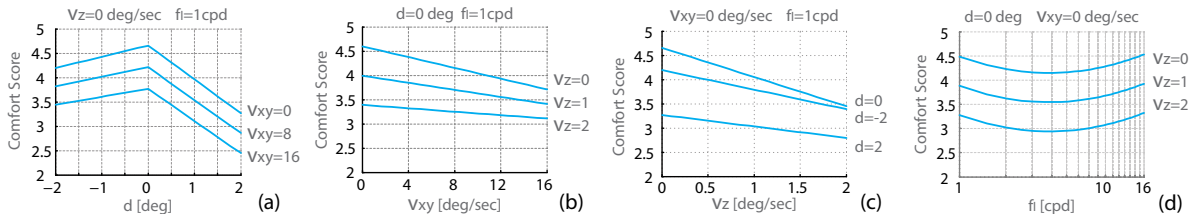
Figure 6: *Different projections of our measurement function C. We refer the reader to the supplementary material for more projections of the function.*

$p_9 = -1.1599$), as shown in Figure 6(d). Didyk et al. [2012] computed the influence of f_l on perceived depth, and found a similar minimum. This could mean that, when taking motion into account, smaller perceived depths may produce a more comfortable 3D viewing experience.

5 Metric of Visual Comfort

Our measurement can be used to predict the level of comfort when viewing short stereoscopic videos. In particular, we derive a metric to compute both a pixel-wise comfort map of each frame in the video $M_p(i, j)$, which allows to identify the particular areas or objects in each frame that are potential sources of discomfort, and a global comfort score M_g for the whole video.

Given an input stereoscopic video consisting of two corresponding left-right image sequences $(I_L(t), I_R(t))|_{t=0,1,...}$, we first compute the motion $\mathbf{v} = (v_x, v_y, v_z)$ and disparity value d at each pixel (i, j) in each frame $I_L(t)$. Since the binocular views have similar content, we assume they share the same comfort map and we will use the left view for computing velocities and f_l, and both views when computing the disparities. For real-world videos where this information is not usually available, we rely on optical flow (for motion), or motion and depth estimation [Jung et al. 2012]. We define the motion on the screen plane as $v_{xy} = \sqrt{v_x^2 + v_y^2}$.

Pixel-wise metric Our pixel-wise metric leverages our measurement of comfort C presented in the previous section. To take into account luminance spatial frequency, we construct a Laplacian pyramid of the luminance of $I_L(t)$, from a starting base frequency f_{l_0}. Multi-scale decompositions are frequently used to model the varying sensitivity of the human visual system (HVS) to different spatial frequencies. In the case of luminance-contrast frequencies, Laplacian pyramids are an efficient approximation that works well in practice [Mantiuk et al. 2006; Didyk et al. 2012]. Similarly, our local contrast term is an efficient and convenient approximation to assess the influence of each luminance spatial frequency channel on comfort. For each frame, we then compute the comfort score at each (i, j) using our measurement function C as:

$$M_p(i, j) = \sum_{k=0}^{n} C(v_{xy}, v_z, d, \frac{f_{l_0}}{2^k}) \times \frac{L_k(i, j)}{\sum_k L_k(i, j)} \quad (4)$$

where $L_k(i, j)$ is defined as the contrast value of the $(2^{k+1} + 1)$-neighborhood at (i, j) at the k-th Laplacian level, and n is the number of Laplacian levels. In practice we select n such that $f_{l_0}/2^n < 1$ cpd. From the resulting $M_p(i, j)$ we obtain a two-dimensional comfort map per frame, which can be used to visualize the spatial location and distribution of the uncomfortable viewing regions. By stacking maps over time, we obtain a three-dimensional $M_p(i, j, t)$ map for the whole video, which allows to visualize the temporal evolution of the discomfort regions. Figure 7 shows M_p for two time instants of two different video sequences.
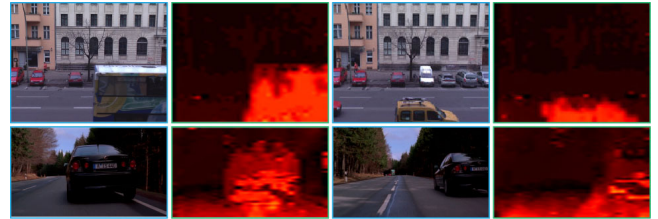

Figure 7: *Representative frames with their computed pixel-wise comfort map M_p for* bus *(top row, ©Fraunhofer HHI) and* car *(bottom row, ©KUK Filmproduktion GmbH) scenes.*

Global metric To compute a global comfort score M_G for the whole video we pool partial metrics both in the spatial and temporal domains. For the spatial, per-frame pooling, existing research suggests that the overall perception of any single frame is dominated by its "worst" area [Keelan 2002]. We thus take a conservative approach and assume that the most uncomfortable region in a frame dictates the discomfort of the whole frame. We further modulate such per-frame discomfort by taking saliency into account. Saliency maps have been employed before in related scenarios, like comfort assessment [Jung et al. 2012], editing of stereo content [Lo et al. 2010], or image and video retargeting [Rubinstein et al. 2010]. We use saliency maps under the reasonable hypothesis that human subjects pay more attention to visually salient regions, and therefore those will have a greater influence in comfort. In our implementation, we obtain a saliency-based segmentation using the method by Cheng et al. [2011], which also yields a per-region saliency value between zero and one. We reduce the discomfort in non-salient regions based on the saliency value, and the comfort M_g for frame t_k is then given by:

$$M_g(t_k) = \min_r (5 - S_r(t_k) \cdot (5 - M_r(t_k))) \quad (5)$$

where S_r represents the saliency values of a given region r and $M_r = \frac{1}{|r|} \sum_{(i,j) \in r} M_p(i, j)$ is the average per-pixel comfort in r.

For temporal pooling, various approaches have been proposed for video quality assessment [Ninassi et al. 2009; Barkowsky et al. 2009]. To obtain our final global metric M_G, we simply pool the results over the whole video by computing the median of the comfort scores over all frames, as validated in previous work [Jung et al. 2012].

Figure 1 (right) shows how the different stages of the metric perform for several sample frames of a stereo movie. For each frame (first row), first a per-pixel map $M_p(i, j)$ is computed (second row), which is then averaged to obtain a per-region comfort measure M_r (third row).
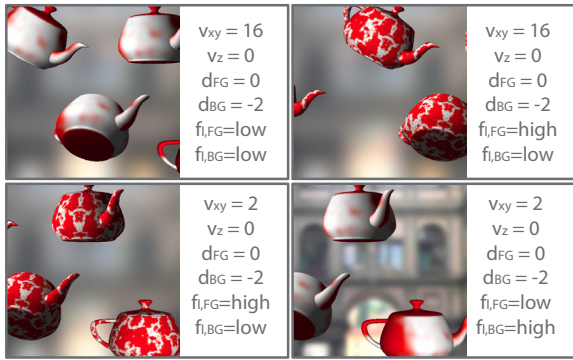
Figure 8: *Representative frames of the stimuli used to test the validity of the approach used to incorporate the influence of luminance contrast spatial frequency. In reading order, stimulus A, B, C and D.*

6 Validation

Our metric is based on the comfort measure as derived in Section 5. In this section we conduct a set of experiments to validate our assumptions regarding the dependence on luminance contrast and the influence of saliency. Please refer to supplementary material for the complete stimuli employed.

Controlled scenes We first test our metric on four simple, controlled scenes, shown in Figure 8. They all consist of a static background ($v_{xy,BG} = 0$) and some moving objects in the foreground. The spatial frequency of luminance contrast f_l varies across stimuli, both for foreground and background. Textures and velocities also vary across stimuli, as detailed in Figure 8. We fix $v_z = 0$ in all cases, as well as the disparity d of both foreground and background to $d_{FG} = 0$ and $d_{BG} = -2$, respectively. We explore four combinations of different frequencies for both foreground and background; additionally, the screen plane velocity of the foreground can take two distinct values, $v_{xy,FG} = \{2, 16\}$.

A user study is run with ten subjects, in which they are asked to rate their comfort when viewing the different clips. Each clip lasts 30 seconds, with a minimum resting period of 60 seconds between videos, and the order of presentation is randomized across subjects. We then compare the comfort values yielded by our metric with assessments given by subjects. Figure 11 (left) summarizes the score predicted by our metric and the average score given by the users. It can be seen how our metric follows very closely such user scores. Since our ranking data does not necessarily follow a normal distribution, we compute Spearman's correlation coefficient, which yields $\rho_S = 1$, with $p = 0.0833$. Pearson's linear correlation coefficient is $\rho_P = 0.9916$ with $p = 0.0084$, indicating that the variables are correlated. This test shows that, even in conflicting scenarios such as stimulus B (high frequency foreground with no disparity but high screen plane velocity) and D (high frequency background with high disparity but no screen plane velocity), our metric is able to capture the relative discomfort elicited by the stimuli.

Additionally, we want to test the influence of saliency on viewing comfort. A second experiment was conducted as a separate session with the same users (resting time between sessions was a minimum of 20 minutes), with a procedure analogous to the one described in the previous paragraph. In this case we use two stimuli, shown in Figure 9, consisting of a series of objects in the foreground, with disparities d_{FG}, on a constant static mid-gray background with dis-
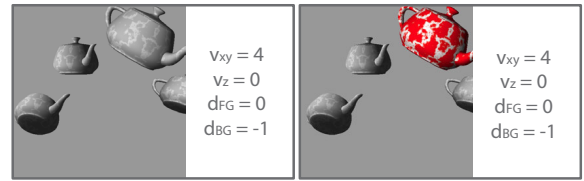

Figure 9: *Representative frames of the stimuli used to test the validity of the saliency scheme integrated in the global metric.*

parity d_{BG}. The objects move erratically with slow screen plane velocity ($v_{xy} = 4$ and $v_z = 0$). In the second clip, we increase the saliency of one of the objects by making it clearly stand out in red. As Figure 11 (center) shows, making the object more salient made the user scores drop, since the discomfort caused by the moving (foreground) teapot now becomes more relevant; this behavior was also predicted by our metric.

We also compare our metric with Jung et al.'s [2012]. Since they do not consider the influence of luminance, their metric will yield the same values for cases A and B (and C and D) in Figure 11 (left), and again for the two cases in Figure 11 (center). For the first experiment, Spearman's correlation coefficient is $\rho_S = 0.8944$ ($p = 0.3333$), while Pearson's coefficient is $\rho_P = 0.9045$ ($p = 0.0955$). Recall that our metric, in contrast, yields higher correlations with the measured data: $\rho_S = 1$ and $\rho_P = 0.9916$.

Real scenes We use four different scenes –*bus*, *bunny*, *horse* and *car*– to further validate our metric. They exhibit a variety of motion combinations as well as different luminance frequencies and disparity ranges. Figure 10 shows representative frames of each one. Ten subjects were asked to rate their comfort level after viewing the clip, and a two-minute rest is forced between two clips. Again, we compute the comfort score according to the presented metric and compare it against the score given by the subjects; results are shown in Figure 11 (right). Again, there is a strong correlation between predicted and user scores, with Spearman's rank correlation coefficient yielding a value of $\rho_S = 1$ ($p = 0.0833$), while Pearson's correlation coefficient is $\rho_P = 0.9514$ ($p = 0.0486$). Although our metric tends to slightly overestimate comfort, the predicted value is generally inside the 95% confidence interval for the mean. The difference for the *car* scene is significantly larger, which is due to its complex v_z velocity field: Objects move in opposite directions, with nonlinear motion in depth and changes in the sign of v_z with respect to the camera. Given the high linear correlation in the data, we can compensate the overestimation by fitting a linear function to the metric, to obtain the final global expected score: $M_{exp} = 3.45 M_G - 10.20$ ($R^2 = 0.91$).

7 Applications

Our work is a contribution towards a comfortable viewing experience. In this section, we describe various applications of our experiments and metric, including: Stereoscopic production, scientific visualization and retargeting.

Stereoscopic Production Various rules and guidelines have been proposed for practical use in stereoscopic content production in order to provide a comfortable viewing experience [Mendiburu 2009; Smolic et al. 2011b; Smolic et al. 2011a]. Often, these guidelines are based on years of experience in the production industry. Our quantitative measurements can complement that know-how. The challenge is the multidimensional nature of the measurements: the function presented in Figure 5 may be too elaborate for practical
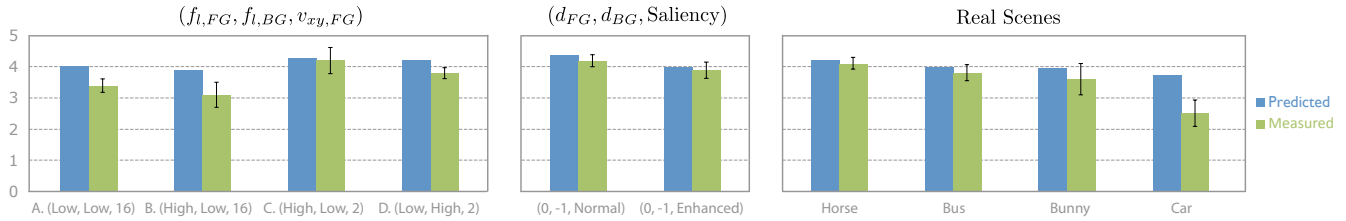
Figure 11: *Results from the metric validation study.* Left: *Comparison between measured user scores for visual comfort and the predicted score using our metric to assess the validity of the approach used to incorporate the influence of luminance contrast spatial frequency.* Center: *Comparison between predicted and measured scores when testing the influence of saliency in the prediction of viewing comfort.* Right: *Scores measured and predicted by our metric for the videos of real scenes. Error bars show standard error of the mean.*

Figure 10: *Representative anaglyph frames of the stimuli used to test the validity of our metric in real scenarios. From left to right, top to bottom: bus, horse (©KUK Filmproduktion GmbH), bunny and car.*

use in a production scenario. However, comfort zones can be derived from the measurements, which can in turn be used as a guide for content design. We can define the comfort zone $Z_{l,h}$ as:

$$Z_{l,h} = \{(d, v_{xy}, v_z, f_l) | l \leq C(d, v_{xy}, v_z, f_l) \leq h\}$$

where l and h are the given lower and upper bounds. Slices for the sample case of $l = 3$ and $h = 4$, $Z_{3,4}$, are shown in Figure 1 (left). Considering, for instance, a moving object with a given v_{xy}, the illustrated comfort zone defines a safe range of disparities on which such object can lie as a function of its velocity in z. Additionally, automatic computation of camera placement [Oskam et al. 2011; Heinzle et al. 2011; Jones et al. 2001], could potentially benefit from incorporating information from our measurements.

Stereoscopic Retargeting   The adaptation of stereoscopic content to a disparity range that provides a comfortable viewing experience has motivated a number of recent works that focus on disparity retargeting [Lang et al. 2010; Kim et al. 2011; Masia et al. 2013; Kellnhofer et al. 2013].

Our work can be applied to this area in two ways. First, as shown in the previous paragraph and in Figure 1 (left), our measurements can be used to define zones of comfort which could be incorporated as constraints when defining a retargeting operator $\phi(d)$. Second, our metric can provide information about the change of predicted comfort caused by the retargeting operation, taking into account the motion in the scene; this can in turn be used to evaluate or select a given operator. Figure 12 shows our predicted distribution of discomfort for the *horse* video, for two different disparity retargeting operators. This provides users with a more insightful view of potential sources of discomfort, both in the temporal and disparity domains.

Visualization   Visualization of complex, three dimensional data is extensively used in various fields, including engineering, geoscience, medicine, biology, architecture or education. In some cases, this visualization can be improved by employing stereoscopic techniques [Ebert et al. 1996; Johnson et al. 2006]. In such systems, uncontrolled user navigation may lead to visual discomfort; our measurement can be used to provide constraints on the navigation (motion), which would guarantee a comfortable viewing experience.

Given a desired lower bound of the visual comfort level, our measurement could be used to dynamically map the user's input from the navigation device into a comfortable range of camera motions and velocities. In a simple implementation, a sampling of the scene would be performed: At sampled positions, the disparity d would be queried for the current frame(s), as well as an estimation of the dominant luminance spatial frequency f_l. This would be used to set boundaries on the maximum v_{xy} and v_z allowed for navigating the scene. A simple, conservative approach would take the minimum values from all the sampled data; more sophisticated approaches can incorporate importance sampling strategies, either task-oriented or based on visual saliency. This can be made more practical by processing batches of frames.

Assessment Metric   A number of tools have recently appeared that focus on the editing of stereoscopic content, such as copy-and-pasting [Lo et al. 2010; Luo et al. 2012], drawing [Kim et al. 2013], converting 2D images to stereo pairs [Konrad et al. 2012; Do et al. 2011], or warping [Niu et al. 2012]. While the initial content may be assumed to have been carefully generated, it is still hard to predict how any of these editing operations would affect the resulting viewing experience. The metric we propose in this paper can be used to evaluate the discomfort that may arise from post-processing operations such as the ones mentioned above.

8   Conclusion and Future Work

We have introduced a novel measurement for the visual discomfort caused by motion in stereoscopic content. A four-dimensional space is explored which includes disparity, planar and depth velocities, as well as the spatial frequency of luminance contrast. Based on these measurements, a metric is proposed to evaluate the level of comfort associated to viewing short stereoscopic videos.

There is ample opportunity for exciting future work. Given the complexity of the HVS, other factors not taken into account here may affect visual comfort, such as the spatial frequency of the disparity, or the temporal frequency of luminance contrast. Similarly, investigating the effect of higher order components of motion (acceleration) can help analyze more complex scenes. Additionally, our saliency estimation does not consider motion or disparity; this
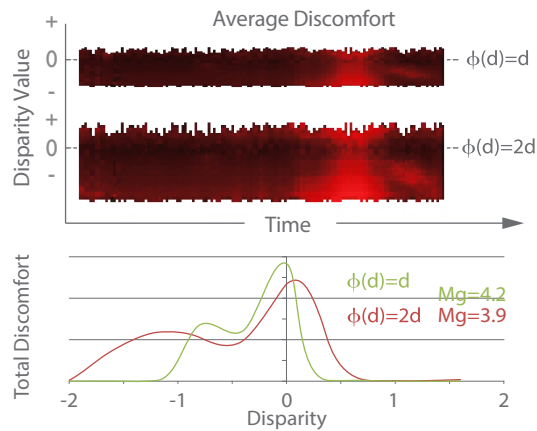
Figure 12: *For the example video horse, we show the distribution of pixel-wise discomfort $(5 - M_p)$ for the original disparity $\phi(d) = d$ and a linear mapping operator $\phi(d) = 2d$. Top: average pixel-wise discomfort. Each vertical line corresponds to one frame. Color indicates discomfort. Bottom: the distribution of total discomfort over d for the whole video. The change of predicted global comfort M_g is also provided.*

is a possible cause of the current overshooting of our metric, which we fix with a fitting function, but deserves further investigation. Our measurements have been tested for a near viewing distance (50 cm in our experiments): Different viewing conditions could be studied using our methodology. Moreover, more sophisticated metrics and models should probably use qualitative information gathered from industry experts. We hope that our work fosters future research in this area, including both stereo applications and a deeper understanding of the mechanisms of our visual system.

9 Acknowledgments

References

BAHILL, A. T., AND STARK, L. 1975. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Exp Neurol 48*, 1, 95–106.

BARKOWSKY, M., BIALKOWSKI, J., ESKOFIER, B., BITTO, R., AND KAUP, A. 2009. Temporal trajectory aware video quality measure. *IEEE Journal of Selected Topics in Signal Processing 3*, 2, 266–279.

CHENG, M.-M., ZHANG, G.-X., MITRA, N. J., HUANG, X., AND HU, S.-M. 2011. Global contrast based salient region detection. In *IEEE CVPR*, 409–416.

CHO, S.-H., AND KANG, H.-B. 2012. Subjective evaluation of visual discomfort caused from stereoscopic 3d video using perceptual importance map. In *TENCON 2012 - 2012 IEEE Region 10 Conference*, 1–6.

CUTTING, J. E., AND VISHTON, P. M. 1995. *Perception of Space and Motion*. Academic Press, ch. Perceiving Layout and Knowing Distances: The integration, relative potency, and contextual use of different information about depth.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Trans. on Graph. 30*, 4.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., SEIDEL, H.-P., AND MATUSIK, W. 2012. A luminance-contrast-aware disparity model and applications. *ACM Trans. Graph. 31*, 6, 184:1–184:10.

DO, L., ZINGER, S., AND DE WITH, P. H. N. 2011. Warping error analysis and reduction for depth-image-based rendering in 3dtv. vol. 7863, 78630B–78630B–9.

EBERT, D. S., SHAW, C. D., ZWA, A., AND STARR, C. 1996. Two-handed interactive stereoscopic visualization. In *IEEE Visualization*.

HEINZLE, S., GREISEN, P., GALLUP, D., CHEN, C., SANER, D., SMOLIC, A., BURG, A., MATUSIK, W., AND GROSS, M. 2011. Computational stereo camera system with programmable control loop. *ACM Transactions on Graphics 30*, 94:1–10.

HESS, R. F., KINGDOM, F. A. A., AND ZIEGLER, L. R. 1999. On the relationship between the spatial channels for luminance and disparity processing. *Vision Research 39*, 559–68.

HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. VergenceCaccommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3.

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*, 271–297.

HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth. Volume 2: Depth Perception*. I Porteous.

JIN, E. W., MILLER, M. E., ENDRIKHOVSKI, S., AND CEROSALETTI, C. D. 2005. Creating a comfortable stereoscopic viewing experience: effects of viewing distance and field of view on fusional range. *Proc. SPIE 5664*, 10–21.

JIN, E. W., MILLER, M. E., AND BOLIN, M. R. 2006. Tolerance of misalignment in stereoscopic systems. *Proc. ICIS*, 370–373.

JOHNSON, A. E., LEIGH, J., MORIN, P., AND KEKEN, P. V. 2006. GeoWall: Stereoscopic Visualization for Geoscience Research and Education. *IEEE Computer Graphics and Applications 26*, 10–14.

JONES, G., LEE, D., HOLLIMAN, N., AND EZRA, D. 2001. Controlling perceived depth in stereoscopic images. *Proceedings of SPIE 4297*, 42–53.

JULESZ, B. 2006. *Foundations of Cyclopean Perception*. MIT Press.

JUNG, Y. J., LEE, S.-I., SOHN, H., WOOK PARK, H., AND MAN RO, Y. 2012. Visual comfort assessment metric based on salient object motion information in stereoscopic video. *Journal of Electronic Imaging 21*, 1, 011008–1–011008–16.

KEELAN, B. W. 2002. Predicting multivariate image quality from individual perceptual attributes. In *PICS 2002: IS&T's PICS Conference, An International Technical Conference on Digital Image Capture and Associated System*, 82–87.

KELLNHOFER, P., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2013. Optimizing disparity for motion in depth. In *Proc. of EGSR*, vol. 32.

KIM, C., HORNUNG, A., HEINZLE, S., MATUSIK, W., AND GROSS, M. 2011. Multi-perspective stereoscopy from light fields. *ACM Trans. Graph. 30* (Dec.), 190:1–190:10.

KIM, Y., LEE, Y., KANG, H., AND LEE, S. 2013. Stereoscopic 3d line drawing. *ACM Trans. Graph. 32*, 4 (July), 57:1–57:13.

KONRAD, J., BROWN, G., WANG, M., ISHWAR, P., WU, C., AND MUKHERJEE, D. 2012. Automatic 2d-to-3d image conversion using 3d examples from the internet. vol. 8288, 82880F–82880F–12.

KOOI, F. L., AND TOET, A. 2004. Visual comfort of binocular and 3D displays. *Displays 25*, 99–108.

LAMBOOIJ, M., IJSSELSTEIJN, W., AND FORTUIN, M. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Technology and Science 53*, 1–14.

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3d. *ACM Trans. Graph. 29*, 3, 10.

LEE, B., AND ROGERS, B. 1997. Disparity modulation sensitivity for narrow-band-filtered stereograms. *Vision Research 37*, 1769–1777.

LI, J., BARKOWSKY, M., AND LE CALLET, P. 2011. The influence of relative disparity and planar motion velocity on visual discomfort of stereoscopic videos. In *Third International Workshop on Quality of Multimedia Experience (QoMEX), 2011*, 155–160.

LI, J., BARKOWSKY, M., WANG, J., AND CALLET, P. L. 2011. Study on visual discomfort induced by stimulus movement at fixed depth on stereoscopic displays using shutter glasses. In *International Conference on Digital Signal Processing*, 1–8.

LO, W.-Y., VAN BAAR, J., KNAUS, C., ZWICKER, M., AND GROSS, M. H. 2010. Stereoscopic 3d copy & paste. *ACM Trans. Graph. 29*, 6, 147.

LUO, S.-J., SHEN, I.-C., CHEN, B.-Y., CHENG, W.-H., AND CHUANG, Y.-Y. 2012. Perspective-aware warping for seamless stereoscopic image cloning. *Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2012) 31*, 6, 182:1–182:8.

MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept. 3*, 3 (July), 286–308.

MANTIUK, R., KIM, K. J., REMPEL, A. G., AND HEIDRICH, W. 2011. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph. 30*, 4, 40:1–40:14.

MASIA, B., WETZSTEIN, G., ALIAGA, C., RASKAR, R., AND GUTIERREZ, D. 2013. Display Adaptive 3D Content Remapping. *Computers & Graphics 37*, 8.

MENDIBURU, B. 2009. *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press.

NINASSI, A., LE MEUR, O., LE CALLET, P., AND BARBA, D. 2009. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing 3*, 2, 253–265.

NIU, Y., FENG, W.-C., AND LIU, F. 2012. Enabling warping on stereoscopic images. *Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia 2012) 31*, 6.

OSKAM, T., HORNUNG, A., BOWLES, H., MITCHELL, K., AND GROSS, M. 2011. OSCAM - Optimized stereoscopic camera control for interactive 3D. *ACM Trans. on Graph. 30*, 6, 189:1–189:8.

OSTBERG, O. 1980. Accommodation and visual fatigue in display work. *Displays 2*, 2, 81 – 85.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. The MIT Press.

POLLOCK, B. T., BURTON, M., KELLY, J. W., GILBERT, S., AND WINER, E. 2012. The Right View from the Wrong Location: Depth Perception in Stereoscopic Multi-User Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics 18*, 581–588.

RUBINSTEIN, M., GUTIERREZ, D., SORKINE, O., AND SHAMIR, A. 2010. A comparative study of image retargeting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia) 29*, 5, 160:1–160:10.

SHIBATA, T., KIM, J., HOFFMAN, D. M., AND BANKS, M. S. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision 11*, 8.

SMOLIC, A., KAUFF, P., KNORR, S., HORNUNG, A., KUNTER, M., MULLER, M., AND LANG, M. 2011. Three-dimensional video postproduction and processing. *Proceedings of the IEEE 99*, 4, 607–625.

SMOLIC, A., POULAKOS, S., HEINZLE, S., GREISEN, P., LANG, M., HORNUNG, A., FARRE, M., STEFANOSKI, N., WANG, O., SCHNYDER, L., MONROY, R., AND GROSS, M. 2011. Disparity-aware stereo 3d production tools. In *Visual Media Production (CVMP), 2011 Conference for*, 165–173.

SPERANZA, F., TAM, W. J., RENAUD, R., AND HUR, N. 2006. Effect of disparity and motion on visual comfort of stereoscopic images. In *Proceedings of the SPIE*, vol. 6055, 94–103.

TAM, W. J., SPERANZA, F., YANO, S., SHIMONO, K., AND ONO, H. 2011. Stereoscopic 3d-tv: Visual comfort. *IEEE Transactions on Broadcasting 57*, 2, 335–346.

TEMPLIN, K., DIDYK, P., RITSCHEL, T., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. Highlight Microdisparity for Improved Gloss Depiction. *ACM Trans. Graph. 31*, 4.

UKAI, K., AND HOWARTH, P. A. 2008. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays 29*, 106–116.

YANG, X., ZHANG, L., WONG, T.-T., AND HENG, P.-A. 2012. Binocular tone mapping. *ACM Trans. Graph. 31*, 4 (July), 93:1–93:10.

YANO, S., IDE, S., MITSUHASHI, T., AND THWAITES, H. 2002. A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays 23*, 191–201.

YANO, S., EMOTO, M., AND MITSUHASHI, T. 2004. Two factors in visual fatigue caused by stereoscopic HDTV images. *Displays 25*, 141–150.