

A Luminance-Contrast-Aware Disparity Model and Applications

Piotr Didyk^{1,2} Tobias Ritschel¹ Elmar Eisemann³ Karol Myszkowski¹ Hans-Peter Seidel¹ Wojciech Matusik²

¹MPI Informatik ²CSAIL MIT ³Delft University of Technology / Télécom ParisTech

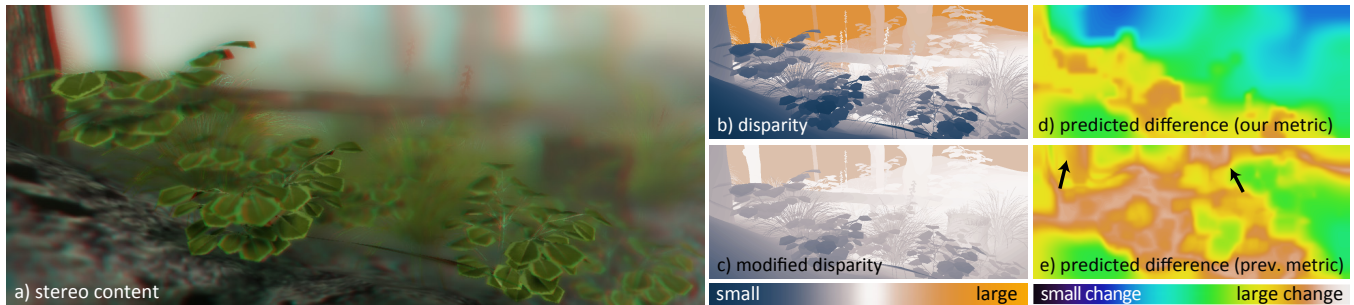


Figure 1: When stereo content (a; b) is manipulated (c), we quantify the perceived change considering luminance, and disparity (d), whereas previous work leads to wrong predictions (e) e. g., for low-texture areas, fog, or depth-of-field (arrows). Please note that all images in the paper, except for disparity and response maps are presented in anaglyph colors.

Abstract

Binocular disparity is one of the most important depth cues used by the human visual system. Recently developed stereo-perception models allow us to successfully manipulate disparity in order to improve viewing comfort, depth discrimination as well as stereo content compression and display. Nonetheless, all existing models neglect the substantial influence of luminance on stereo perception. Our work is the first to account for the interplay of luminance contrast (magnitude/frequency) and disparity and our model predicts the human response to complex stereo-luminance images. Besides improving existing disparity-model applications (e. g., difference metrics or compression), our approach offers new possibilities, such as joint luminance contrast and disparity manipulation or the optimization of auto-stereoscopic content. We validate our results in a user study, which also reveals the advantage of considering luminance contrast and its significant impact on disparity manipulation techniques.

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

Keywords: perception; stereoscopy

Links: [DL](#) [PDF](#) [WEB](#) [VIDEO](#) [DATA](#)

1 Introduction

The human visual system (HVS) combines information coming from many different cues [Howard and Rogers 2002] to determine spatial

layout. Binocular disparity, due to differences of the projected retinal positions of the same object in both eyes, is one of the strongest cues, in particular for short ranges (up to 30 meters) [Cutting and Vishton 1995]. Current 3D display technology allows us to make use of binocular disparity, but, in order to ensure viewing comfort, disparity should be limited to a so-called comfort zone [Rushton et al. 1994; Lambooi et al. 2009; Shibata et al. 2011]. Smaller screens often imply a smaller disparity range, auto-stereoscopic displays only have a reduced depth of field, and artistic manipulations can enhance certain features [Ware et al. 1998; Jones et al. 2001; Lang et al. 2010; Didyk et al. 2011]. Whenever such modifications are applied, it is important to analyze the impact. Furthermore, such a prediction also leads to a better control of the changes. However, so far, no existing perception model considers the influence of RGB image content on depth perception. Intuitively, a certain magnitude of luminance contrast is required to make disparity visible, while stereopsis is likely to be weaker for low-contrast and blurry patterns. In this work, we show that luminance contrast (magnitude/frequency) does have a significant impact on depth perception and should be taken into account for a more faithful computational model. One key challenge of a combined luminance contrast and disparity model is the growing dimensionality, which we limit to 4D by considering: spatial frequency and magnitude of disparity, as well as spatial frequency and magnitude of luminance contrast. We ignore image brightness, pixel color and saturation, which seem to have a lower impact on depth perception (Sec. 7). Our model improves the performance of existing applications [Lang et al. 2010; Didyk et al. 2011] such as disparity retargeting, difference metrics, compression, and even enables previously-impossible applications. Precisely, we make the following contributions:

- A disparity-perception model accounting for image content;
- Measurements of perceived disparity changes for stimuli with different luminance and disparity patterns;
- New methods to automatically retarget disparity and to manipulate luminance contrast to improve depth perception;
- A user study to validate and illustrate the advantages of our method.

First, we will discuss previous work (Sec. 2) and survey the psychophysical evidence for the link between depth and contrast perception (Sec. 3). Then, we describe our computational model and explain how to jointly process disparity and luminance contrast (Sec. 4). Further, we describe the necessary measurements for our model. Next, we show several of its applications and compare to previous work (Sec. 5). Our approach is validated (Sec. 6) and its limitations discussed (Sec. 7), before we conclude.

2 Previous Work

Since our main goal is perception-informed manipulations of stereo content, we will give an overview of existing disparity adjustment methods, which are usually designed to fit the scene's entire disparity range into a limited depth range (called comfort zone) where the conflict between accommodation and vergence is reduced [Lambooj et al. 2009; Shibata et al. 2011]. Jones et al. [2001] presented a mathematical framework for stereo-camera parameters, such as interaxial distance and convergence. Recently, Oskam et al. [2011] proposed a similar approach for real-time applications to optimize camera parameters according to control points that assign scene depth to a desirable depth on a display device. Heinzle et al. [2011] built a computational stereo camera, which can alter its interaxial distance and convergence-plane during stereo acquisition. Other techniques work directly on pixel disparity to map the scene depth into the comfort zone [Lang et al. 2010]. Such operations can also be performed in a perceptually linearized disparity space [Didyk et al. 2011], where the impact of spatial-disparity variations at different frequency scales can be considered [Tyler 1975; Filippini and Banks 2009].

No current solution considers the influence of arbitrary luminance patterns on perceived disparity. Stereoacuity thresholds are found by applying different depth corrugations to carefully-textured images. However, such conditions hardly reflect real images; band-limited, or low-magnitude contrast patterns strongly affect stereo vision. Our goal is to account for such influences in complex images and make disparity manipulations more effective. We build upon findings in spatial vision that analyze this link between luminance patterns and stereoacuity and give a brief overview in the following section.

3 Background

Spatial band-pass channels Although it is often assumed that correspondence matching in stereo is achieved at the level of luminance edges, there is direct evidence that band-pass limited channels in the luminance domain play an important role in disparity processing [Heckmann and Schor 1989]. The observation is not surprising since contrast processing in the HVS follows such principles and contrast is required for stereo matching. Hence, one can expect a strong correlation between stereoacuity and contrast characteristics such as the compressive contrast nonlinearity at suprathreshold levels [Wilson 1980] and the contrast sensitivity function (CSF) [Barten 1989], which we discuss next.

Luminance contrast magnitude Legge and Gu [1989] and Heckmann and Schor [1989] investigated stereoacuity for luminance sine-wave gratings and found that perceivable disparity thresholds decrease with increasing luminance contrast magnitude, which can be modeled using a compressive power function with exponents falling into the range from -0.5 to -0.7 . Similar results have been obtained for narrow-band-filtered random-dot stereograms by Cormack et al. [1991]. We rely on their data when modeling the dependence of stereoacuity on luminance contrast magnitudes. For low values, they observed a significant reduction of stereoacuity

(below a tenth multiple of the detection threshold), which relates to the lower reliability of edge localization in stereo matching due to a poorer signal-to-background-noise ratio in band-pass luminance channels [Legge and Gu 1989]. For contrast at suprathreshold levels, stereoacuity is little affected. Luminance contrast also does not alter the upper disparity limits for comfortable binocular fusion.

Spatial luminance-contrast frequencies Legge and Gu [1989] measured the necessary luminance-contrast thresholds to detect a fixed disparity for sine-wave gratings of various spatial corrugation frequencies. They neglect disparity magnitude, but derive a CSF for stereopsis, whose shape is similar to the luminance-CSF shape. Monocular-luminance thresholds are usually assumed to be 0.3–0.4 log units smaller than the luminance contrast needed for stereovision. We consider suprathreshold luminance contrast, more complex disparity patterns, and explore, how band-tuned luminance contrast interacts with corrugated depth stimuli at various spatial frequencies.

Lee et al. [1997] measured the impact of luminance frequency on disparity perception for band-pass-filtered random-dot stereograms. They showed that the relationship between disparity sensitivity and luminance frequency exhibits a band-pass characteristic with the maximum located at a luminance frequency of 4 cpd, which is shifted for lower-frequency disparity modulation below 0.25 cpd to around 3 cpd. Their conclusion was that the observed differences in sensitivity result from the stimulation of different visual channels, which are tuned to different spatial modulations of luminance and disparity. They also observed that there is a mostly weak influence of luminance frequency on disparity sensitivity at suprathreshold disparities, except for high luminance frequencies as well as low disparity frequencies. Lee et al. considered relatively narrow ranges of luminance frequency 1–8 cpd, disparity corrugation frequency 0.125–1.0 cpd, and disparity magnitudes up to 4 arcmin. In this work, we significantly expand them to 0.3–20 cpd for luminance frequency, 0.05–2 cpd for disparity corrugation frequency and up to 20 arcmin for disparity magnitudes. The results by Lee et al. have been challenged by Hess et al. [1999], who experimented with randomly-positioned Gabor patches with modulated disparity. Hess et al. found that low-frequency disparity modulations were detected equally well for low and high-luminance frequencies. However, for high-frequency disparity corrugations, perception of depth was enhanced when a high-frequency luminance pattern was used, which improves its localization and thus facilitates stereo matching. Our goal is to elaborate a computational model which accounts for such effects in the context of complex luminance and depth configurations with channel-based luminance and disparity processing.

Asymmetry effect An interesting observation is that asymmetry effects in depth perception can occur (Fig. 2). We consider two patches, one in front of the other, each with a luminance texture that we refer to as the *support*, or say that it *supports* the disparity. Limiting the deeper patch to a lower-frequency support, makes the step between the patches less visible and, finally, disappear. When swapping the luminance patterns, the depth difference becomes visible again. This observation has not been reported so far and needs more investigation. Here, we will give an overview of a few insights and show how we believe that this aspect can be integrated into our model.

Our interpretation of the asymmetry effect (see Fig. 3) assumes that, in the considered depth range, we are more sensitive to binocular disparity than pictorial cues such as texture density or relative size [Cutting and Vishton 1995, Fig. 1]. The effect could then relate to lower frequency luminance patterns being difficult to localize accurately in space (e. g., a completely white wall does not allow us to deduce its distance, as our visual system cannot rely on corre-

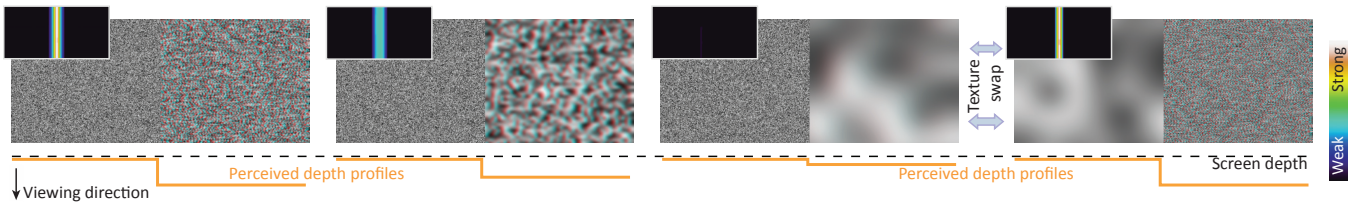


Figure 2: Influence of spatial luminance patterns on depth perception. The physical depth of all stimuli is equal, yet the perceived depth (orange profiles) varies depending on the applied texture patterns. High-frequency removal from the texture on the right/deeper patch leads to a perceived depth reduction (second and third stimuli). While the third stimulus barely exhibits any perceivable depth, just swapping textures leads to a strong depth impression for the fourth stimulus. The insets present the strength of perceived disparity, as predicted by our model. In the additional materials, we provide full-resolution stereo images of the stimuli.

spondence points). Occlusions (such as the step edge) introduce a sharp boundary which results in a well visible discontinuity. Hereby, a relative depth localization (with respect to the background) becomes easier when occluding a high-frequency luminance pattern. We will concentrate on this occlusion/correspondence aspect when integrating the effect into our model. This choice might exclude other factors that could play an important role (e. g., pictorial depth cues and other higher level cues) and further investigations will be needed to fully explain this phenomenon.

One could think that pictorial depth cues overrule the influence of binocular disparity and might explain the effect. Mather et al. [2002] showed that when only pictorial cues are considered, the blurred textures in Fig. 2 should appear as being behind the others. For the first three stimuli, this finding holds when the figure is viewed stereoscopically. However, in the fourth, the blurred texture appears in front. The only relevant difference between our stimuli and the ones presented by Mather and Smith is that ours contain a binocular disparity cue. This shows that disparity has a significant impact on depth perception and can overrule pictorial cues in some cases. In practice, our model seems to work acceptably and the previous observation could indicate that our simplifying assumptions are suitable for our purposes.

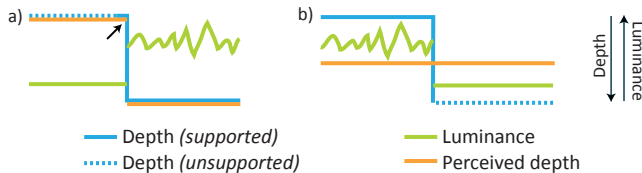


Figure 3: Luminance patterns (green) influence the depth perception (orange) of the same depth profile (blue). Some allow us to well discriminate depth (solid blue), while others do not (dotted blue). The frontal-patch edge can benefit from the luminance contrast between patches (a, arrow). If the luminance pattern of the deeper patch renders localization difficult the depth step disappears (b).

4 Disparity and Luminance Processing

Here, we explain our perceptual model to predict the HVS response to a disparity signal in presence of a supporting luminance pattern. We then illustrate how to use our model to express physical values in perceptually linear units, which is achieved by constructing so-called *transducer functions* and derive the computational model used in our applications. Finally, we present how to determine the few remaining parameters in a psychophysical experiment.

4.1 Threshold Function

The first step in deriving a model is to acquire a threshold function $th(f_d, m_d, f_l, c_l)$, which for each combination of its parameter values (disparity frequency f_d and disparity magnitude m_d , luminance frequency f_l , and luminance-contrast magnitude c_l) defines the smallest perceivable change (i.e., equivalent to 1 JND) in disparity magnitude (expressed in units of arcmin).

As indicated by Legge and Gu [1989], only low-level luminance-contrast magnitude affects stereoacuity, while otherwise having little to no influence. Further, Cormack [1991], presented a corresponding disparity-threshold function for luminance-contrast magnitude. Consequently, we decided to factor out the luminance-contrast magnitude dimension, leading to the following model:

$$th(f_d, m_d, f_l, c_l) = s(f_d, m_d, f_l) / Q(f_l, c_l), \quad (1)$$

where s is a discrimination-threshold function assuming maximal luminance-contrast magnitude and Q is a function that compensates for the increase of the threshold due to a smaller luminance-contrast magnitude c_l .

We model s via a general quadratic polynomial function:

$$\begin{aligned} s(f_d, m_d, f_l) = & p_1 \log_{10}^2(f_d) + p_2 m_d^2 + p_3 \log_{10}^2(f_l) \\ & + p_4 \log_{10}(f_d) m_d + p_5 \log_{10}(f_d) \log_{10}(f_l) + p_6 m_d \log_{10}(f_l) \\ & + p_7 \log_{10}(f_d) + p_8 m_d + p_9 \log_{10}(f_l) + p_{10}, \end{aligned} \quad (2)$$

where $\mathbf{p} := [p_1, \dots, p_{10}]$ is a parameter vector obtained by minimizing the following error: $\arg \min_{\mathbf{p} \in \mathbb{R}^{10}} \sum_{i=1}^n ((s(o_i) - \Delta m_i) / (\Delta m_i))^2$, where o_i are stimuli with their corresponding thresholds Δm_i , as determined in our psychophysical experiment (Sec. 4.5). Hereby, we obtain $\mathbf{p} = [0.3655, 0.0024, 0.2571, 0.0416, -0.0694, -0.0126, 0.0764, 0.0669, -0.3325, 0.2826]$, which results in the disparity discrimination function th visualized in Fig. 4. The use of the log domain is motivated by previous work [Didyk et al. 2011] and leads to better results. The range of disparity detection thresholds specified by our model is in good agreement with the data in [Lee and Rogers 1997] for measured mid-range disparity and luminance frequencies. For more extreme ranges, similar to [Hess et al. 1999], we observe that, for low-frequency disparity corrugations, a wide range of luminance frequencies lead to good stereoacuity, while for higher-frequency disparity corrugations stereoacuity is weak for low luminance frequencies.

To determine the scaling function Q , we use the data by Cormack [1991], expressed in units of threshold multiples c_m , to which we fit a cubic polynomial in the logarithmic domain:

$$T(c_m) = \exp(r_1 \log_{10}^3(c_m) + r_2 \log_{10}^2(c_m) + r_3 \log_{10}(c_m) + r_4), \quad (3)$$

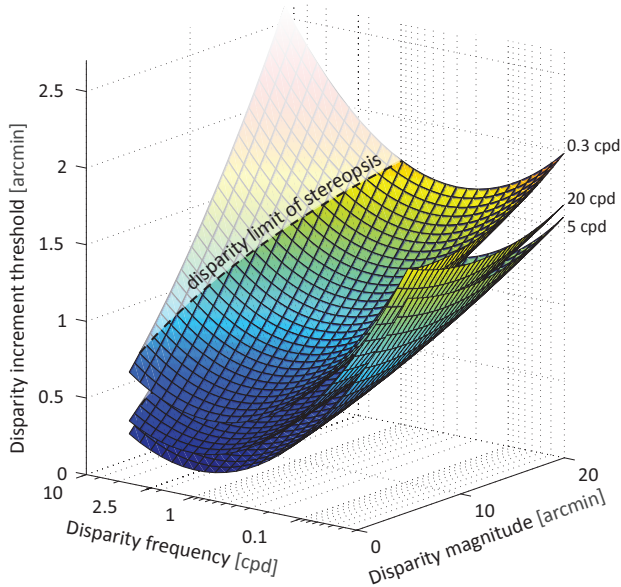


Figure 4: Plot visualizing slices of our model of the disparity discrimination function for sinusoidal corrugations. We illustrate three surfaces corresponding to different luminance frequencies (0.3 cpd, 5 cpd and 20 cpd) and a well visible contrast (above 10 JNDs). The model is limited by the disparity limit of stereopsis measured by Tyler et al. [1975]. In the additional materials we provide a plot which shows our model in a bigger range of disparity magnitude values such that the disparity limit is visible across whole range of disparity frequencies.

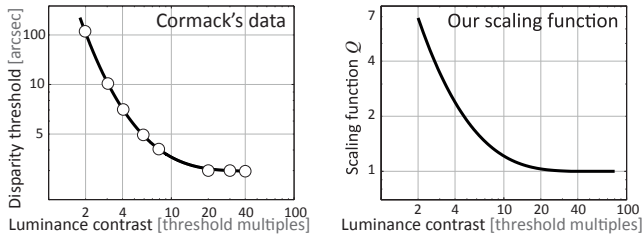


Figure 5: Our function fitting to Cormack's data (marked by empty circles), as well as our scaling function Q .

where $\mathbf{r} := [-0.9468, 4.4094, -6.9054, 4.7294]$ is a parameter vector obtained from fitting the above model to the experimental data of Cormack [1991]. Q is then expressed as:

$$Q(f_1, c_1) = \begin{cases} T(c_1 \cdot csf(f_1))/T(u) & \text{if } c_1 \cdot csf(f_1) \leq u \\ 1 & \text{if } c_1 \cdot csf(f_1) > u \end{cases}, \quad (4)$$

where csf is the luminance contrast sensitivity function [Barten 1989] and $u := 35.6769$ specifies when the luminance contrast has no further influence on the disparity threshold [Legge and Gu 1989], meaning $T'(u) = 0$. Our fit is illustrated in Fig. 5.

4.2 Transducer

A transducer function relates physically measurable quantities to the HVS response (in JND units). Typically transducers are specified for luminance-contrast magnitude [Wilson 1980; Mantiuk et al. 2006], but, recently, disparity magnitude has also been considered [Didyk et al. 2011]. Didyk et al. assumed a perfectly visible luminance

pattern and proposed a two-dimensional transducer of disparity frequency and magnitude, which leads to a conservative prediction. Consequently, perceived disparity is generally overestimated. We extend their solution to a four-dimensional transducer $t(f_d, m_d, f_1, c_1)$, which we build directly from the threshold function th :

$$t(f_d, m_d, f_1, c_1) = \int_0^{m_d} th(f_d, x, f_1, c_1)^{-1} dx \quad (5)$$

The function $t(f_d, \cdot, f_1, c_1) : \mathbb{R} \rightarrow \mathbb{R}$ (a partial application of t to f_d, f_1, c_1) is monotonic, hence, there usually¹ exists an inverse transducer $(t(f_d, \cdot, f_1, c_1))^{-1}$. t maps disparity-luminance stimuli to a perceptually linear space of disparity and t^{-1} can be used to reconstruct the stimuli. E.g., for disparity compression, mapping via t makes removing imperceptible disparities easy and t^{-1} can be used to reconstruct the modified disparity map. Similarly, we can build a transducer to convert luminance contrast to a uniform space. For more details on constructing transducer functions please refer to work by Wilson [1980] and Mantiuk et al. [2006].

In practice, a transducer function t can be evaluated by numerical integration and stored in a table. t^{-1} can be implicitly defined via a binary search. Nonetheless, in four dimensions, the memory and performance costs can be significant. A better solution makes use of the factorization: $t(f_d, m_d, f_1, c_1) = t'(f_d, m_d, f_1)/Q(f_1, c_1)$, where $t'(f_d, m_d, f_1) = \int_0^{m_d} s(f_d, x, f_1)^{-1} dx$. Functions t' (and t'^{-1} if wanted) can be discretized, precomputed, and conveniently stored as 3D arrays. The inverse transducer for a given f_d, f_1, c_1 is then: $m_d = t'^{-1}(f_d, Q(f_1, c_1) \cdot R, f_1)$, where R is a JND-unit response to disparity.

In order to account for the HVS limits of perceivable stereopsis, we use our threshold function only within the limits measured by Tyler et al. [1975] (Fig. 4). Beyond this range, transducer functions should remain flat but stay invertible. We achieve this by enforcing the functions to be strictly increasing beyond the stereopsis limit, but keeping their total increase below 1 JND.

4.3 Computational Model

The above transducer is valid for abstract stimuli. For real content, we decompose the input's luminance and disparity into corresponding Laplacian pyramids, such as it has been done independently for luminance [Mantiuk et al. 2006] and disparity [Didyk et al. 2011] before.

For luminance, we compute a Laplacian pyramid C of the luminance pattern, which contains Michelson contrast values c_l , which are already required for Q in Eq. 4. Pixel-disparity values are transformed into vergence (world-space angles) [Didyk et al. 2011], and we build a Laplacian pyramid D [Burt and Adelson 1983]. The value $D_i(\mathbf{x})$ corresponds to the disparity value at location $\mathbf{x} \in \mathbb{R}^2$ in the i -th level frequency of the pyramid i. e., $\alpha/2^i$ cpd (where $\alpha \approx 20$ for our setup). To convert disparities into JND units, we apply the transducer function to the values of the Laplacian pyramid. Disparity frequency as well as disparity magnitude are defined directly in the pyramid D : $f_d = \alpha/2^i$ and $m_d = D_i(\mathbf{x})$. To evaluate the transducer, we also need to know the frequency f_1 and contrast c_1 of the supporting luminance pattern.

To combine luminance and disparity, we follow the independent-channels hypothesis for disparity processing as proposed by Marr and Poggio [1979]; stereoacuity is determined by the most sensitive channel and remains uninfluenced by other channels. Consequently, given a disparity frequency f_d , we assume that the response is the **maximum** of all responses for all higher luminance frequencies f_1

¹only for constant luminance patterns, the function cannot be inverted

in the image region corresponding to half a cycle of f_d . This choice is justified in more detail in Sec. 7.

Formally, the response is then:

$$D'_i(\mathbf{x}) = \max_{j \in (0, \dots, i-1)} t(\alpha/2^i, D_i(\mathbf{x}), \alpha/2^j, S_j(\mathbf{x})), \quad (6)$$

where $S_j(\mathbf{x})$ evaluates the luminance support, defined as the average of all contrast values C_j of the j -th level of the luminance decomposition that fall into a rectangular region $\sigma_i(\mathbf{x}) = (\mathbf{x} - (w, w)^T, \mathbf{x} + (w, w)^T)$ of size $w = 2^j$ around \mathbf{x} (Fig. 6). The values of S can be pre-computed from C and later accessed in constant time. The resulting structure is a Laplacian pyramid with a MIP map defined on each of its levels, as visualized in Fig. 6, right. Note, that computing the maximum of S_j over all levels and, then, applying a transducer independently is *not* equivalent.

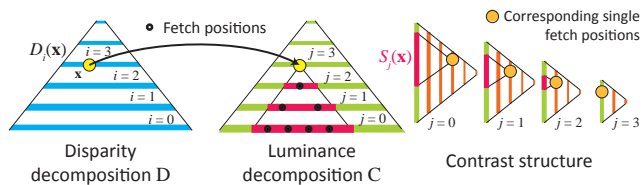


Figure 6: For a disparity $D_i(\mathbf{x})$ at location \mathbf{x} , the model needs to involve levels $j < i$ in the luminance Laplacian pyramid C . In each level j , an average contrast $S_j(\mathbf{x})$ of a region $\sigma_i(\mathbf{x})$ (marked in red) around \mathbf{x} is computed and its impact evaluated. For acceleration, averages can be pre-computed in MIP maps for each level (right).

4.4 Asymmetries

So far, our computational model does not account for the asymmetries described in Sec. 3, as it would be necessary to study an even higher-dimensional space including neighborhood configurations. Nonetheless, we can exploit a few observations to derive a perceptually motivated model that we verify practically (Sec. 6).

In fact, in order to perceive a sinusoidal depth corrugation, peaks as well as valleys of the sinusoid need to be well supported by luminance contrast. The HVS relies on clear correspondences, which might not always be easy to discern, as illustrated in Fig. 7. To account for the full wave, a 3×3 neighborhood at the given level of the Laplacian decomposition is evaluated and the minimum response chosen. Hereby, we ensure that a full cycle is well supported and visible.

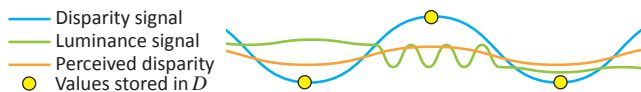


Figure 7: A weak luminance contrast can attenuate the disparity response; in the valleys of the sinusoidal depth function the low spatial frequency of the luminance signal weakens the overall perceived corrugation.

While this extension already explains several cases in Fig. 2, it is insufficient to explain the entire asymmetry. The texture swap would not yet be detected to influence depth perception. In order to better model the response, we need to take disocclusion into account. In fact, the occluding patch’s edge introduces a high-contrast luminance edge in superposition with the patch beneath. If they are present in both views (left and right eye), these high frequencies allow us to localize the edge in space - we disregard pathological cases where

disparity and luminance frequency perfectly agree. Consequently, we propose to evaluate the luminance contrast for both views and use the maximum response. Hereby, a point on the deeper patch will be disoccluded in one view and reveal its high-frequency luminance neighborhood, while points on the edge will maintain a high-contrast edge in both views. Also note, that this effect affects not only the points directly on the edge, but also in a small neighborhood near the edge. This relates to findings on backward-compatible stereo [Didyk et al. 2011]. Similarly to the Cornsweet effect for luminance, the HVS extrapolates depth information to neighboring locations. Although heuristic, this solution performs well in practice (Sec. 6).

4.5 Psychophysical Experiment

To derive the parameters of s (Eq. 2), our experiment explores: disparity frequency f_d (measured in cpd), disparity magnitude m_d (measured in arcmins), and luminance frequency f_l (measured in cpd).

Stimuli All stimuli are horizontal sinusoidal disparity corrugations with luminance noise of a certain frequency. First, we create a luminance pattern by producing a noise of frequency f_l and scale it to match the maximal reproducible contrast on our display. Using such a texture excludes any external depth cues, such as shading. Next, we create a disparity pattern – a sinusoidal grating with frequency f_d and magnitude m_d . Such disparity gratings do not produce occlusions. Finally, the luminance pattern is warped according to the disparity pattern to produce an image pair for the observer’s left and right eye [Didyk et al. 2010]. All steps are adjusted to the viewing conditions, i. e., the screen size and viewing distance. We assume a standard intra-ocular distance of 65 mm.

Equipment We use a Samsung SyncMaster 2233RZ display (1680 \times 1050 pixels, 1000 : 1 contrast), along with NVIDIA 3D Vision active shutter glasses, observed from a distance of 60 cm, ensured by a chin-rest. Measurements were performed in controlled, office-lighting conditions.

Subjects All subjects in our experiment were naïve, paid and had normal or corrected-to-normal vision. Before conducting the experiment, we checked that subjects are not stereo-blind [Richards 1971]. In total there were 24 participants who took part in the experiment (12 F, 12 M). They were all between 22 and 30 years old. One participant was discarded due to very high thresholds (on average 3 times higher than the thresholds of others).

Task In this experiment, we seek measuring a disparity-discrimination threshold for stimuli defined in three-dimensional parameter space. For a given stimulus $o = (f_d, m_d, f_l)$, we run a threshold estimation procedure. In each step, we show two stimuli o and $o + [0, \Delta m_d, 0]$. One located on the left-hand side of the screen and the other on the right. The position is randomized. The task of the participant is to judge which stimulus exhibits larger depth magnitude and choose using the “left” and “right” arrow keys. Depending on the answer, Δa is adjusted in the next step using the QUEST procedure [Watson and Pelli 1983]. When the standard deviation of the estimated value is lower than 0.05, the process stops.

Each participant performed 35 adjustment procedures. One session took from 30 to 100 min. Subjects were allowed to take a break whenever they felt tired. In total, we obtained 805 measured thresholds to which we fit our model.

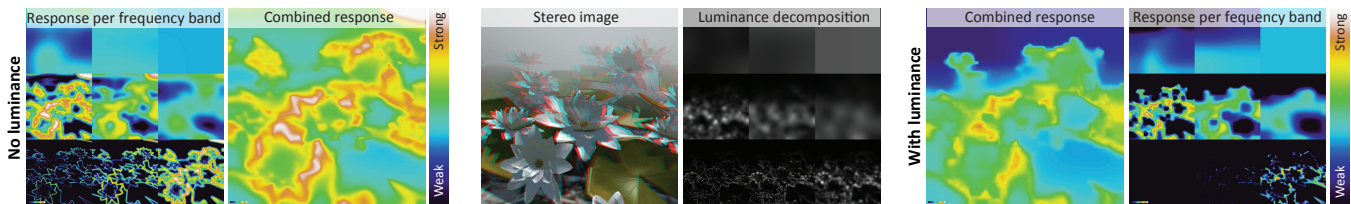


Figure 8: Comparison of perceived disparity as predicted by the previously proposed model by Didyk et al. that ignores image content (left) and our model (right). Responses per frequency band and the combined response are shown for both, as well as the original stereo image with the multi-band decomposition of the luminance pattern (middle).

5 Applications

Previous work [Didyk et al. 2011] demonstrated a number of applications for disparity models, including a stereo-image metric and compression. In this section, we show how our new model improves these results. We also present new applications such as disparity optimization, including the case for a multi-view auto-stereoscopic display as well as joint (luminance and disparity) manipulations. Most of those techniques were not possible using luminance-ignoring models.

5.1 Stereo Image Metric

Our model can be used to predict the perceived difference between two stereo images: a reference image and a second image which underwent a distortion, such as compression. Perceptual image metrics have previously been proposed independently for luminance contrast [Mantiuk et al. 2006] and disparity [Didyk et al. 2011].

To overcome this limitation, we first use our model to map both input images into our perceptually linear space. The transducer function is applied after the *phase uncertainty operation*, similarly to previous work [Lubin 1995]. Per-band differences (a simple subtraction) then indicate the detectability of disparity changes. All bands can be combined using a Minkowski summation to produce a spatially varying difference map. We use the same parameters as those reported in [Didyk et al. 2011] for both—phase uncertainty and Minkowski summation.

A comparison of our metric and the method of Didyk et al. [2011] is shown in Fig. 1 and 8. Our approach successfully detects the human inability to perceive changes of disparity when the luminance support is not adequate, i. e., low luminance contrast because of missing texture, fog, or depth-of-field (Fig. 1). Previous metrics are too conservative and report invisible differences (false positives).

5.2 Stereo Compression

Key to many perceptual compression approaches is to map the signal into a perceptually linear space, such that the perception of artifacts can be reliably controlled. This is the idea behind classic image compression such as JPG [Taubman and Marcellin 2001], but also disparity compression [Didyk et al. 2011]. All values below the detection threshold, as predicted by our model, are removed. Including luminance leads to more compact compression (Fig. 9).

5.3 Disparity Optimization

One of our new applications is perceptual disparity optimization, which automatically fits the disparity of stereo content into a limited range by analyzing disparity and luminance contrast via our model. The objective is to achieve a small difference between the original and the re-mapped content according to our disparity metric. Due

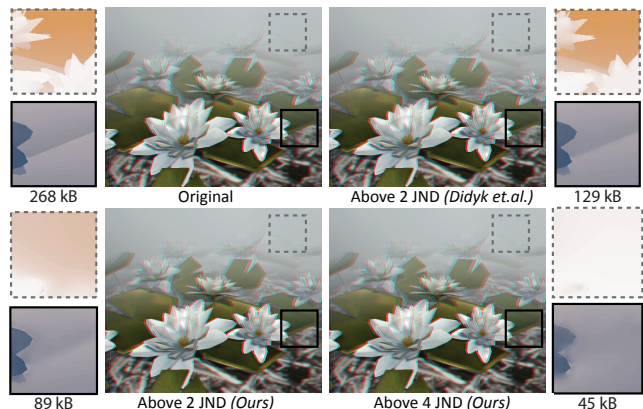


Figure 9: Comparison of disparity compression using our method and the more conservative technique by Didyk et al. Our method can account for regions where a poor luminance pattern reduces sensitivity to depth changes. Therefore, it can remove information more aggressively than previous techniques. The insets show zoomed-in parts of pixel disparity maps. The size corresponds to the size of our disparity representation compressed using LZW.

to many non-linearities of human disparity-luminance perception the optimization is challenging and the search space of all possible disparity re-mappings is difficult to tackle.

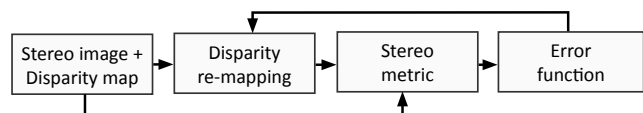


Figure 10: Our disparity optimization. From left to right: Input is a stereo image and a disparity map. A disparity mapping P is applied to the input. Our metric computes the difference between input and remapped content. The difference is converted into a single error value, and a new mapping P is chosen. The process is repeated until the error is low enough or a fixed iteration number is reached.

To make the problem tractable, we restrict the search space to the subset of all global and piecewise-defined mapping curves, as done for automatized tone mapping [Mantiuk et al. 2008] (Fig. 10). Such curves can be defined using a small number of n (we use $n = 7$) control points with values at fixed locations $P := \{(0, y_0), \dots, (1.0, y_n)\}$ combined with a simple (e.g., piecewise-cubic) reconstruction. Given the original stereo content A and a remapping $r(A, P)$ of A using the control points P , simulated annealing is used to minimize the integrated perceived difference over the image domain Ω

$$\min_{P \in \mathbb{R}^n} \int_{\Omega} A \ominus r(A, P) dx,$$

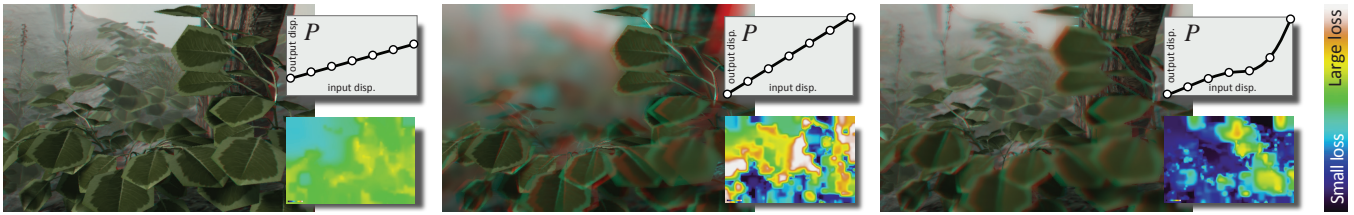


Figure 12: Trade-off between the depth range and sharpness on a multi-view auto-stereoscopic display. The insets show disparity mapping functions and the loss of depth perception due to blur. Left to right: simple mapping that fits entire scene in the depth-of-field region (marked in white on curve plots), disparity mapping using the entire pixel disparity range, our mapping. Our mapping leads to a good balance between depth perception and depth-of-field constraints.

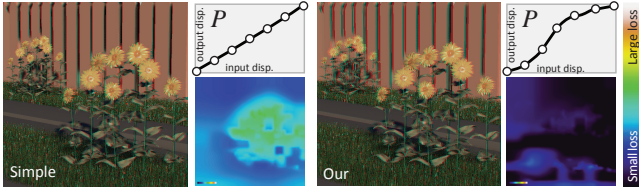


Figure 11: Our optimization compared to linear disparity mapping. Insets visualize mapping curves and disparity perception loss compared to the original stereo image, as reported by our metric.

where the \ominus operator denotes our perceptual metric of disparity difference. By implementing our method on a GPU, the disparity optimization can be performed at interactive speed e. g., while a user navigates inside the scene (Fig. 11). In order to maintain temporal coherence, we use the last frame’s solution as the initial guess for P in the next frame. We can further smoothly interpolate previous solutions over a couple of frames to improve the smoothness of the animation. A similar approach was recently used in [Oskam et al. 2011].

5.4 Multi-view Autostereoscopic Display

Disparity optimization is particularly important for multi-view auto-stereoscopic displays, where the affordable disparity range is very shallow. Beyond this range, depth-of-field blur is usually applied in order to avoid intersperspective aliasing [Zwicker et al. 2006]. Therefore, two extreme strategies (Fig. 12, left) are possible. Either, the whole scene needs to fit into the small range where everything can be sharp or a bigger range can be used, but then prefiltering (blur) is necessary. The trade-off between these two solutions is not obvious. Our metric can predict the strength of perceived depth in the presence of blur due to depth-of-field. Therefore, using our optimization scheme along with the metric, leads to an optimal trade-off between sharpness and depth range. Two modifications are required: First, based on the display specification, the focal range (ϕ_0, ϕ_1) has to be computed. Second, a depth of field operator $d(A, \phi_0, \phi_1)$ has to be applied to the luminance content A [Zwicker et al. 2006]. The solution is given by:

$$\arg \min_{P \in \mathbb{R}^n} \int_{\Omega} A \ominus d(r(A, P), \phi_0, \phi_1) dx$$

An example of this optimization is presented in Fig. 12, right.

5.5 Joint Disparity and Luminance Manipulations

We can predict the perceived change of distorted disparity, just like the effect of luminance distortions on perceived depth. Hence, we can identify image regions, where the stereo impression is weak

due to poor luminance support. We can quantify this effect by comparing two stereo images with the same disparity pattern but an assumed-perfect luminance pattern in one of them.

By improving the luminance contrast in areas where the original support proves insufficient, we re-introduce the impression of depth as shown in Fig. 13. We also use this technique to illustrate the successful detection of asymmetries (Fig. 14).

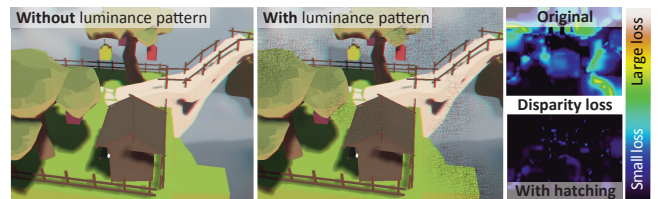


Figure 13: An insufficient luminance support in the original stereo image (left), lowers its depth perception (top right). By adding a hatching pattern, guided by our metric, the resulting stereo image (middle) shows significantly less stereo loss (bottom right).

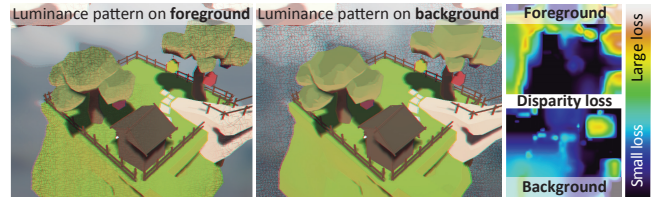


Figure 14: To illustrate the prediction of asymmetries, we show two cases: hatching on the foreground (left) and the background (middle). Compared to foreground hatching (right top), background hatching creates more pronounced differences due to disocclusions, leading to better depth perception (right bottom), as correctly predicted by our metric.

6 Evaluation

In order to evaluate model and applications, we conducted an additional user study with 17 new participants. The first part verified how well our metric predicts actual JND values. We used a stereo image from Fig. 9 and applied a scaling to the disparity in order to create images that differ in depth perception. One image was modified to match an average error of 0.5 JND (with minimum 0.4 JND and maximum 0.8 JND). For a second image the average difference was 3 JND (with minimum 2.5 JND and maximum 3.5 JND). We showed the modified images side by side (randomized) with the original image and asked about perceived differences. Each pair was shown ten times in randomized order. The 0.5 JND difference image

was detected in 58 % cases, which is close to a random answer, as expected. For the 3 JND case the probability of the detection was 91 %.

To evaluate our compression, we used the examples in Fig. 9. We compared the original to images where all disparities below 2 JND were removed using our model, as well as the conservative model presented by Didyk et al. [2011]. Again, we employed ten randomized repetitions. We asked participants which compression technique produces images that are closer to the original in terms of depth. In 51 % our new compression method was chosen as the one closer to the original, suggesting that our technique improves the compression ratio without introducing artifacts.

To evaluate our disparity optimization, we compared it to existing techniques in a pair-wise comparison with three different scenes (Fig. 15) and four different techniques: camera-parameter adjustment [Jones et al. 2001; Oskam et al. 2011] (CAM), perceptual disparity scaling [Didyk et al. 2011] (PCT), the proposed optimization scheme of this paper without (OUR-D), as well as with accounting for the luminance support (OUR-CD). For each method we ensured that the resulting disparities spanned the same range. In total, 18 pairs of stereo images were shown in a randomized order to the 17 participants who were asked to indicate which stereo image exhibits a better depth impression. In order to analyze the obtained data we computed scores (the average number of times each method was preferred) and computed a two-way ANOVA test. To reveal the differences between the methods, we performed a multiple comparison test. Detailed results of the study are presented in Fig. 16.



Figure 15: Scenes used in our study. For all images used in our study please refer to the accompanying additional materials.

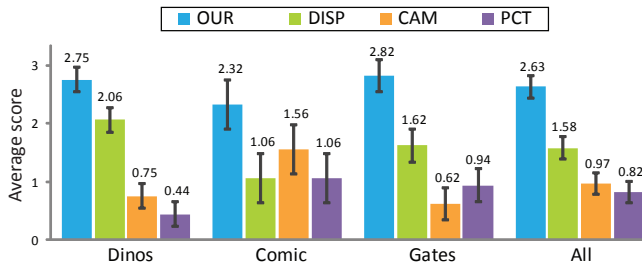


Figure 16: Statistical data obtained in our study. The error bars show 95 % confidence intervals.

The study showed, that for the scenes *Dinos* (courtesy of [Lee et al. 2009]) and *Gates*, our optimization was preferred over all other methods and the effect was significant. The lower performance of CAM, as well as PCT is due to the inability to effectively compress disparities in regions that are less crucial for depth perception. In the *Comic* scene, the difference between OUR-CD and CAM is not statistically significant for $p = 0.05$, but it is when assuming $p = 0.1$. This observation indicates that in some cases our solution may perform similarly to others. In the case of the *Comic* scene, this effect can be well explained; the biggest depth-range compression can be obtained in the back, due to the low luminance frequency in the sky,

which is correctly detected by our model. The CAM solution mostly affects the background, actually even a bit too much. Our solution more evenly distributes the depth impression (refer to the images in the additional materials) and while the foreground looks similar, the background has more depth information. Nonetheless, this difference is very localized in the scene. Generally, the results show that including luminance in the model improves the performance of the disparity optimization significantly.

We also illustrate the usefulness of our metric for autostereoscopic displays, where depth-of-field and disparity perception are linked and, hence, no luminance-insensitive metric would work. We used the examples from Fig. 12. We compared our method separately to the mapping that linearly fits everything into the depth-of-field region and the one that uses the full display-disparity range. 13/14 out of 17 participants preferred the depth impression delivered by our method to using the entire depth-of-field/disparity range. For completeness, we also tested whether our luminance pattern in Fig. 13 improved depth perception and 16 out of 17 participants chose our solution. A two-sided binomial statistical test revealed that in both studies results were statistically significant with $p < 0.05$.

7 Discussion

The independent-channels hypothesis for disparity processing [Marr and Poggio 1979], was applied when computing the perceived disparity $D'_i(\mathbf{x})$ in Eq. 6. It implies that stereoacuity is determined by the most sensitive channel and remains uninfluenced by others. This hypothesis has been confirmed in psychophysical studies where stereoacuity has been investigated for independent, as well as summed up sine-wave stimuli of different luminance-contrast frequencies and magnitude [Heckmann and Schor 1989]. It turns out that the phase relationship of sine-wave components, which affects also the local shape of the resulting luminance gradients, is not utilized in stereoacuity. What matters are mostly peak-to-through luminance gradients. Even more convincing is that the thresholds obtained for sinusoidal luminance gratings, for which stereoacuity is best (in the range of 3–10 cpd), are the same as those obtained for multi-frequency square-wave luminance stimuli [Legge and Gu 1989, Fig. 3]. In all cases, the same Michelson luminance-contrast magnitude has been considered.

To reduce dimensionality, we decided to exclude the influence of luminance-contrast magnitude from our measurements; stereo increment thresholds per luminance spatial frequency channel actually increase for low contrast as a power-law function [Rohaly and Wilson 1999, Fig. 6]. We considered this influence in a simplified form by expressing the signal in each luminance channel in JND units including its normalization via the CSF function. We then compute stereoacuity per channel using a compressive function (Eq. 4), which we derived based on the data from [Cormack et al. 1991].

In our perceptual model, we do not consider temporal aspects [Lee et al. 2007]. It would require adding additional dimensionality to our experimental data, and we relegate such an extension as future work. Also, we ignore chromatic stereopsis, which is less contrast sensitive, leads to weaker stereoacuity, and features a more limited disparity range with respect to its luminance counterpart [Kingdom and Simmons 2000].

Finally, we do not consider image brightness because stereoacuity weakly depends on luminance in mesopic and photopic levels (over 0.1 cd/m^2), which are typical for standard stereo 3D displays [Howard and Rogers 2002, Chapters 19.5.1].

Our disparity space is linear in the same way as CIELUV or CIELAB color spaces. It is constructed via integration of the threshold function as it has been done before for luminance [Wilson 1980; Mantiuk

et al. 2006], and similarly, the linearity cannot be global. It is also important to underline that we do not make *absolute depth* perceptually linearized, but disparity, which is defined as in the perceptual literature [Howard and Rogers 2002, Fig. 19.1].

In order to account for the disparity limit of stereopsis we used the data provided by Tyler et al. [1975] (Sec. 4.2). Alternatively, the finding of Burt et al. [1980] could be used. We chose Tyler's data because he considered disparity limits for sinusoidal patterns, which better fit our model, while Burt et al. used points.

Concerning the generality of our model, we did not repeat the experiment for different display technologies (e. g., anaglyph, polarization), which may result in a slightly different stereoacuity. However, measurements with different equipment are not a problem and our model and techniques remain valid. For displays with different parameters (e. g., size, resolution, contrast ratio), our model is directly applicable; it uses physical values which can be computed from the display specification and viewing conditions. Our evaluation was conducted on a different group of people than the threshold measurements. The positive results of the study suggest that, although stereoacuity varies among people, our model is general enough to be successfully used in practice.

Comparing our disparity optimization to other enhancement techniques, such as Cornsweet profiles [Didyk et al. 2012] could be considered. However, these can be used as an additional step atop any disparity adjustment.

8 Conclusion

We presented a model to capture the interaction of disparity and luminance contrast, while previous work focused on these aspects separately. To our knowledge, our model is the first of its kind and enables effective stereo-content modification. A user study allowed us to derive a new disparity-sensitivity function and we explained how we believe that certain neighborhood-related effects, such as asymmetry, could be integrated as well. While modern rendering effects (depth of field, lens flare, motion blur, veiling glare, participating media, as well as poor visibility conditions – rain, night,...) increase realism or are added for artistic/aesthetic reasons, they also affect luminance contrast, which in turn influence the disparity perception. With our technique, adequate disparity handling becomes possible in all these situations. The same holds for non-photorealistic rendering such as toon shading, or hatching techniques. By using our model, we were able to improve existing, but also develop new compelling applications, such as an image optimization for multi-view autostereoscopic displays or joint (luminance/disparity) processing. Our novel disparity optimization method is a good alternative to previous methods for disparity-range control. We showed that considering luminance significantly improves the results of the proposed mapping technique and showed the validity of our results in an additional study.

In the future, models such as ours will be crucial for stereo images and video processing. Many other applications are possible; combined tone and disparity remapping for HDR stereo content, or luminance hatching could be combined with other styles of non-photorealistic rendering. We also believe that our model could be integrated in a 3D video conference system, as, especially in architectural environments, regions with weak luminance variations are common. Further, our way of optimizing 3D content could be used to consider different viewing conditions or even viewers.

Acknowledgments

This work was partially supported by NSF IIS-1116296 and Quanta Computer.

References

- BARTEN, P. G. J. 1989. The square root integral (SQRI): A new metric to describe the effect of various display parameters on perceived image quality. In *Proc. SPIE*, vol. 1077, 73–82.
- BURT, P., AND ADELSON, E. 1983. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* 31, 4, 532–540.
- BURT, P., AND JULESZ, B. 1980. A disparity gradient limit for binocular fusion. *Science* 208, 4444, 615–617.
- CORMACK, L., STEVENSON, S., AND SCHOR, C. 1991. Interocular correlation, luminance contrast and cyclopean processing. *Vis Res* 31, 12, 2195–2207.
- CUTTING, J., AND VISHTON, P. 1995. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of Space and Motion*, Academic Press, W. Epstein and S. Rogers, Eds., 69–117.
- DIDYK, P., RITSCHER, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Adaptive image-space stereo view synthesis. In *Proc. VMV*, 299–306.
- DIDYK, P., RITSCHER, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Trans. Graph.* 30, 96:1–96:10.
- DIDYK, P., RITSCHER, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. Apparent stereo: the cornsweet illusion can enhance perceived depth. In *Proc. SPIE*, vol. 8291, 8291ON.
- FILIPPINI, H., AND BANKS, M. 2009. Limits of stereopsis explained by local cross-correlation. *J Vision* 9, 1, 8:1–8:18.
- HECKMANN, T., AND SCHOR, C. M. 1989. Is edge information for stereoacuity spatially channeled? *Vis Res* 29, 5, 593–607.
- 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 Trans. Graph.* 30, 94:1–94:10.
- HESS, R., KINGDOM, F., AND ZIEGLER, L. 1999. On the relationship between the spatial channels for luminance and disparity processing. *Vis Res* 39, 3, 559–68.
- HOWARD, I. P., AND ROGERS, B. J. 2002. *Seeing in Depth*, vol. 2: Depth Perception. I. Porteous, Toronto.
- JONES, G., LEE, D., HOLLIMAN, N., AND EZRA, D. 2001. Controlling perceived depth in stereoscopic images. In *Proc. SPIE*, vol. 4297, 42–53.
- KINGDOM, F., AND SIMMONS, D. 2000. The relationship between colour vision and stereoscopic depth perception. *J Society for 3-D Broadcasting and Imaging* 1, 10–19.
- LAMBOOIJ, M., IJSSSELSTEIJN, W., FORTUIN, M., AND HEYN- DERICKX, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J Imaging Science and Technology* 53, 030201–14.
- LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph.* 29, 4, 75:1–75:10.
- LEE, B., AND ROGERS, B. 1997. Disparity modulation sensitivity for narrow-band-filtered stereograms. *Vis Res* 37, 13, 1769–77.

- LEE, S., SHIOIRI, S., AND YAGUCHI, H. 2007. Stereo channels with different temporal frequency tunings. *Vis Res* 47, 3, 289–97.
- LEE, S., EISEMANN, E., AND SEIDEL, H.-P. 2009. Depth-of-field rendering with multiview synthesis. *ACM Trans. Graph. (Proc. of SIGGRAPH Asia)* 28, 5.
- LEGG, G., AND GU, Y. 1989. Stereopsis and contrast. *Vis Res* 29, 8, 989–1004.
- LUBIN, J. 1995. A visual discrimination model for imaging system design and development. In *Vision models for target detection and recognition*, World Scientific, E. Peli, Ed., 245–283.
- MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Applied Perception* 3, 3, 286–308.
- MANTIUK, R., DALY, S., AND KEROFISKY, L. 2008. Display adaptive tone mapping. *ACM Trans. Graph.* 27, 3, 68:1–68:10.
- MARR, D., AND POGGIO, T. 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond. Ser. B* 204, 301–28.
- MATHER, G., AND SMITH, D. 2002. Blur discrimination and its relation to blur-mediated depth perception. *Perception* 31, 10, 1211–1220.
- OSKAM, T., HORNUNG, A., BOWLES, H., MITCHELL, K., AND GROSS, M. 2011. Oscan - optimized stereoscopic camera control for interactive 3D. *ACM Trans. Graph.* 30, 189:1–189:8.
- RICHARDS, W. 1971. Anomalous stereoscopic depth perception. *JOSA* 61, 3, 410–14.
- ROHALY, A. M., AND WILSON, H. R. 1999. The effects of contrast on perceived depth and depth discrimination. *Vis Res* 39, 1, 9 – 18.
- RUSHTON, S., MON-WILLIAMS, M., AND WANN, J. P. 1994. Binocular vision in a bi-ocular world: new-generation head-mounted displays avoid causing visual deficit. *Displays* 15, 4, 255 – 260.
- SHIBATA, T., KIM, J., HOFFMAN, D., AND BANKS, M. 2011. The zone of comfort: Predicting visual discomfort with stereo displays. *J Vision* 11, 8, 11:1–11:29.
- TAUBMAN, D. S., AND MARCELLIN, M. W. 2001. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, Norwell, MA, USA.
- TYLER, C. W. 1975. Spatial organization of binocular disparity sensitivity. *Vis Res* 15, 5, 583 – 590.
- WARE, C., GOBRECHT, C., AND PATON, M. 1998. Dynamic adjustment of stereo display parameters. *IEEE*, vol. 28, 56–65.
- WATSON, A. B., AND PELLI, D. G. 1983. QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics* 33, 2, 113–120.
- WILSON, H. 1980. A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics* 38, 171–8.
- ZWICKER, M., MATUSIK, W., DURAND, F., PFISTER, H., AND FORLINES, C. 2006. Antialiasing for automultiscopic 3D displays. In *Proc. EGSR*, 73–82.