Motion Parallax in Stereo 3D: Model and Applications

Petr Kellnhofer[1] Piotr Didyk[1,2] Tobias Ritschel[1,2,3] Belen Masia[1,4] Karol Myszkowski[1] Hans-Peter Seidel[1]

[1]MPI Informatik [2]Saarland University, MMCI [3]University College London [4]Universidad de Zaragoza

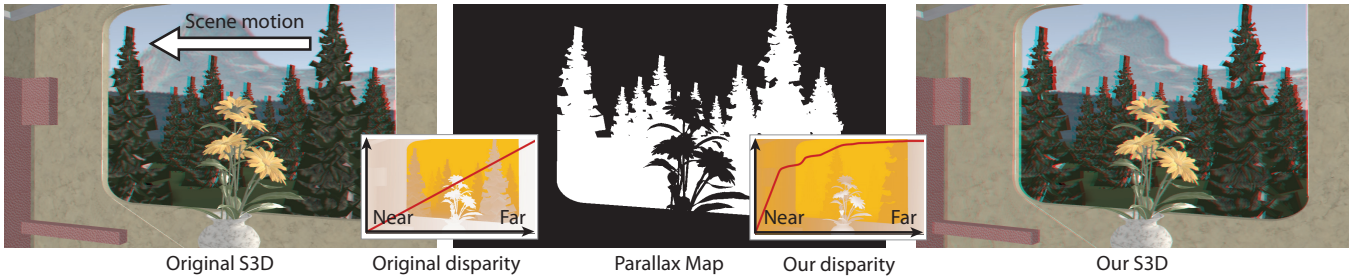

Figure 1: *Starting from stereoscopic video content with a static observer in a moving train* (Left)*, our method detects regions where motion parallax acts as an additional depth cue* (Center, white color) *and uses our model to redistribute the disparity depth budget from such regions* (the countryside) *to regions where it is more needed* (the train interior) (Right).

Abstract

Binocular disparity is the main depth cue that makes stereoscopic images appear 3D. However, in many scenarios, the range of depth that can be reproduced by this cue is greatly limited and typically fixed due to constraints imposed by displays. For example, due to the low angular resolution of current automultiscopic screens, they can only reproduce a shallow depth range. In this work, we study the motion parallax cue, which is a relatively strong depth cue, and can be freely reproduced even on a 2D screen without any limits. We exploit the fact that in many practical scenarios, motion parallax provides sufficiently strong depth information that the presence of binocular depth cues can be reduced through aggressive disparity compression. To assess the strength of the effect we conduct psychovisual experiments that measure the influence of motion parallax on depth perception and relate it to the depth resulting from binocular disparity. Based on the measurements, we propose a joint disparity-parallax computational model that predicts apparent depth resulting from both cues. We demonstrate how this model can be applied in the context of stereo and multiscopic image processing, and propose new disparity manipulation techniques, which first quantify depth obtained from motion parallax, and then adjust binocular disparity information accordingly. This allows us to manipulate the disparity signal according to the strength of motion parallax to improve the overall depth reproduction. This technique is validated in additional experiments.

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

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

SA '16 Technical Papers, December 05 - 08, 2016, Macao

1 Introduction

Perceiving the layout of a scene is one of the main tasks performed by the human visual system. To this end, different depth cues [Cutting and Vishton 1995] are analyzed and combined into a common understanding of the scene. Not all cues, however, provide reliable information. Pictorial cues, such as shadows, aerial perspective or defocus, can often be misleading. Other cues, such as occlusion, provide only a depth ordering. There are also very strong cues, such as ocular convergence and binocular disparity, which provide a true stereoscopic impression. These, however, can only be reproduced in a limited fashion due to significant limitations of the current display technology. The first problem is the lack of a correct accommodation cue in most of the current stereoscopic displays, which leads to visual discomfort [Hoffman et al. 2008; Lambooij et al. 2009] when a large disparity range is shown to observers. The limited range of displayable disparity that does not cause such issues defines the depth budget of the display suitable for practical applications. While current research tries to address this problem with new light-field displays [Masia et al. 2013b], these solutions have even stronger requirements regarding the depth range. This is due to the apparent blur for objects that are located at a significant distance from the screen plane [Zwicker et al. 2006; Wetzstein et al. 2012]. Because of the aforementioned reasons, it is crucial to take any opportunity that allows us to improve depth reproduction without using additional disparity range.

One of the strongest monocular depth cues is motion parallax. It arises when the location of features at different depths results in different retinal velocities. The strength of this cue is relatively high when compared to other monocular cues and also when compared to binocular disparity [Cutting and Vishton 1995]. This fact has been exploited in several applications, such as *wiggle stereoscopy* [Wikipedia 2015b] where motion parallax is used as a metaphor for stereoscopic images, or *parallax scrolling* [Wikipedia 2015a] used in games where, by moving foreground and background at different speeds, a depth sensation is evoked. Striking examples of motion parallax efficiency are species that introduce subtle head movements to enable motion parallax [Kral 2003]. This mechanism has been incorporated into cameras where apparent depth is enhanced by

ISBN: 978-1-4503-4514-9/16/12
DOI: http://dx.doi.org/10.1145/2980179.2980230

subtle motion of the sensor [Proffitt and Banton 1999; v3© Imaging 2015]. Interestingly, motion parallax is not limited to observer motion, but also provides depth information whenever local motion in the scene follows a predictable transformation [Ullman 1983; Luca et al. 2007] (Fig. 2). These facts suggest that motion parallax is a very strong source of depth information for the human visual system (HVS), but it has never been explored in the context of stereoscopic image manipulations.

In this work, we address this opportunity and propose a computational model for detecting motion parallax and quantifying the amount of apparent depth it induces together with binocular disparity. To this end, we conduct a series of psychovisual experiments that measure apparent depth in stereoscopic stimuli in the presence of motion parallax. This is done for simple, sinusoidal corrugation stimuli. Based on the measurements, we propose a computational model that predicts apparent depth induced by these cues in complex images. Furthermore, we demonstrate how the model can be used to improve depth reproduction on current display devices. To this end, we develop a new, motion-aware disparity manipulation technique. The key idea is to reallocate the disparity range from regions that exhibit motion parallax to static parts of the scene so that the overall depth perceived by observers is maximized. To evaluate the effectiveness of our manipulations, we perform additional validation experiments which confirm that by taking the motion parallax depth cue into account, the overall depth impression can be enhanced without extending the disparity budget (Fig. 1). More precisely, we make the following contributions:

- we design and perform psychovisual experiments quantifying the joint contribution of binocular disparity and motion parallax to depth perception,
- propose a computational model that predicts apparent depth induced by these cues for complex scenes, and
- develop new motion-driven disparity manipulations for stereoscopic and multiscopic content.

2 Related work

Motion parallax is a depth cue that results from observer movement. As we move, objects closer to us move in the visual field faster than objects that are farther away. This relative velocity is used by the HVS to recover the distance between different points in the scene. Motion parallax can also be triggered when two points in the scene undergo a rigid transformation, e.g., translation (Fig. 2). In this section, we provide background information about this mechanism, and discuss previous work on modeling it. In particular, we concentrate on the relation of motion parallax to binocular disparity, as well as the interaction between these two cues.

Motion parallax modeling Binocular disparity is usually parameterized as a difference in vergence angles of the point of interest and the fixation point (refer to Fig. 2). Motion parallax can be parameterized in *equivalent disparity* units, which enables a direct comparison to binocular disparity [Rogers and Graham 1982; Ono et al. 1986; Bradshaw and Rogers 1996]. Equivalent disparity is defined as the binocular disparity in orthostereoscopic viewing of the shape that generated such motion parallax [Rogers and Graham 1982].

Modeling motion parallax by means of the equivalent disparity makes an assumption about the absolute scale of depth in the scene. As motion parallax is a purely relative depth cue, we can choose an arbitrary scaling factor and apply it to both the scene geometry and motion. This will still lead to the same 2D image sequence, and hence, the same motion parallax. However, assuming that human inter-ocular distance is fixed, each of these scenes would have different disparities. Similarly, there will be multiple possible

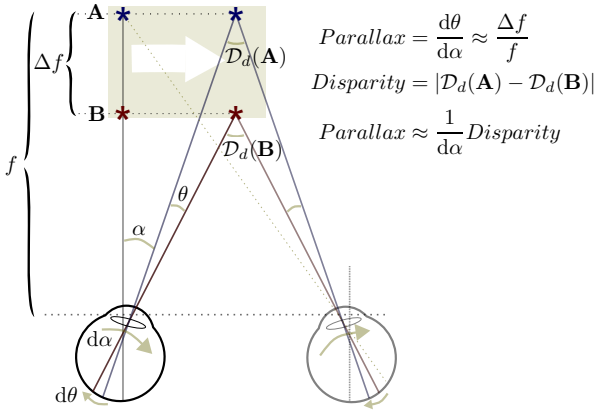

Figure 2: *Motion parallax mechanics [Nawrot and Stroyan 2009]. Translation of a pair of points A and B to the right (or equivalently observer translation to the left) produces motion parallax. Assuming that A is the fixation point, its retinal position does not change, due to the pursuit eye rotation α at angular velocity $\mathrm{d}\alpha/\mathrm{d}t$, while at the same time the retinal position of B changes with respect to A by the angular distance θ producing retinal motion with velocity $\mathrm{d}\theta/\mathrm{d}t$. The relation $\mathrm{d}\theta/\mathrm{d}\alpha \approx \Delta f/f = m$ approximately holds, which, given the fixation distance f, enables recovery of depth Δf from motion parallax. The right side of the figure shows the relation of motion parallax to the angular measure of binocular disparity $d = |\mathcal{D}_d(\mathbf{A}) - \mathcal{D}_d(\mathbf{B})|$ [Howard and Rogers 2012, Fig. 14.5] for the asymptotic case of $f \to \infty$ as derived by Stroyan [2010].*

parallax values for each disparity distribution. This ambiguity can be resolved by directly measuring the relative depth as $m = \Delta f/f$ [Nawrot and Stroyan 2009] (Fig. 2). This particular form prevents division by zero for flat surfaces and restricts possible values to the range from 0 to 1.

Note that as a static scene description the relative depth is not a definition of motion parallax itself. A necessary condition for existence of motion parallax is the presence of rigid motion, which requires a dynamic definition. Nawrot et al. [2009] show that the perception of motion parallax follows the ratio of the retinal motion velocity of distractor $\mathbf{B}$ ($\mathrm{d}\theta/\mathrm{d}t$) and the eye motion velocity during a pursuit of the fixate $\mathbf{A}$ ($\mathrm{d}\alpha/\mathrm{d}t$) (Fig. 2). These two signals allow the HVS to derive the relative depth, as it can be shown that $\mathrm{d}\theta/\mathrm{d}\alpha \approx \Delta f/f$ [Nawrot and Stroyan 2009].

Perceived depth from motion parallax While processing of depth from disparity and motion parallax seem to engage common neural mechanisms [Rogers and Graham 1982; Bradshaw et al. 2006], there are also notable differences between them. The HVS sensitivity to motion parallax as a depth cue has been measured in the equivalent disparity experiment where static sinusoidal depth corrugations have been used as the reference for perceived depth [Rogers and Graham 1982; Bradshaw and Rogers 1996; Bradshaw et al. 2006]. The respective sensitivity function has a shape similar to the disparity sensitivity function (DSF) [Bradshaw and Rogers 1999] with a maximum around 0.4 cpd; however, the threshold magnitudes are 2 to 10 times higher than for disparity.

Unlike for binocular disparity, depth constancy was found to be rather poor for motion parallax, as the estimated depth increased with the square of a feature's distance [Ono et al. 1986]. When geometrically equivalent depth is presented through motion parallax or binocular disparity as the only depth cue, foreshortening of 25%-125% has been observed for the motion parallax presentation

[Durgin et al. 1995]. Nawrot et al. [2014] performed a matching experiment, where the perceived depth as induced by motion parallax has been compared to equivalent depth resulting from binocular disparity, and observed a near-tenfold depth foreshortening. Such depth foreshortening should be understood as a perceived distance reduction for a given object with respect to the fixated point. Nawrot et al. found that to model perceived depth due to motion parallax, compressive nonlinearities (transducers) need to be applied in the ratio θ/α (Fig. 2).We extend the model of Nawrot et al. [2014], which was derived for monoscopic 2D moving stimuli, to stereoscopic stimuli with binocular disparity. This enables modeling perceived depth for different combinations of motion parallax and disparity, which is relevant for stereo 3D applications. Moreover, the model by Nawrot et al. does not account for the relative velocity on the display which directly translates into retinal velocity. Such relative velocity is characterized by its own detection thresholds [Campbell and Maffei 1981], which affect the ability to detect motion parallax as well. Therefore, our model accounts for relative velocity in the presence of binocular disparity.

Disparity vs. motion parallax on 3D displays Both the reduced threshold sensitivity and depth foreshortening for supra-threshold depth ranges indicate that motion parallax is a weaker depth cue compared to binocular disparity. However, since in stereoscopic displays disparity is usually strongly compressed to maintain viewing comfort [Lang et al. 2010; Masia et al. 2013a; Chapiro et al. 2014], the apparent depth from motion parallax becomes comparable, as it always refers to the original, uncompressed depth. In Sec. 5 we seek the opportunities of disparity range compression in regions where depth perception is well supported by motion parallax. Note that disparity is constrained by the fixed interocular distance, while motion parallax can still be detected at far distances, given that the observer velocity is sufficiently high. An additional advantage of motion parallax is that in contrast to disparities which may cause diplopia on certain 3D displays, motion parallax always leads to physically plausible scene interpretations without additional effort.

Disparity and motion parallax cue fusion Different classes of tasks combining motion parallax and disparity depth cues lead to different models of cue fusion. In a surface detection task from random dots, Turner et al. [1997] did not observe the performance to increase when the two cues were present, and noticed that disparity was dominant when contradicting cues were shown. Young et al. [1993] have looked into perceived depth from a combination of (inconsistent) motion and texture cues through perturbation analysis and argued for a weak fusion model, i.e., a weighted linear combination with weights decreasing if a cue is contaminated by noise. Such a strictly modular approach to both cues is not supported by experimental evidence that arises, for example, in surface recognition tasks involving motion parallax and disparity, where cooperative and facilitative interactions have been identified [Howard and Rogers 2012, Ch. 30.2.6]. Landy et al. [1995] presented a *modified weak fusion* model for cue combination, which can handle such interactions through cue promotion when they are in agreement with other cues. Domini et al. [2006] proposed an *intrinsic constraint* model for describing cue fusion. It assumes that families of possible depth layouts are derived for both the disparity and motion parallax cues up to an arbitrary affine transformation, and then their unique best intersection is chosen using a maximum likelihood decision rule. In the following section, we propose a model, which is derived in a perceived-depth matching task, where cue interactions are explicitly measured for various combinations of geometrically correct motion parallax and compressed or expanded disparity values. Our model is fully empirical and does not contradict either of the two aforementioned models. The range of motion parallax and disparity magnitudes modeled fully covers the requirements of stereoscopic image manipulation (Sec. 5).

Kinetic depth effect (KDE) An object structure can be recovered from its rotating motion; this is referred to as the kinetic depth effect (KDE) [Wallach and O'Connell 1953]. Durgin et al. [1995] demonstrated that, as for motion parallax, the recovery of quantitative depth information from object rotation in monoscopic images is weaker than from binocular disparity. Recently, Bista et al. [2016] proposed an approach for KDE triggering from a pair of photographs, where first a rough scene geometry (mesh) is reconstructed, then an optimal location for the scene rotation axis that is parallel to the screen is derived, and finally a swinging rotation of the camera around this axis is introduced. In this work we consider motion parallax, which is inherent for an animation, rather than artificially generated rotations as in [Bista et al. 2016]. Also, we focus on the binocular vision and disparity manipulation, rather than monocular images and mesh deformation, to bring the relative velocity between rotating scene elements into desirable ranges as proposed by Bista et al. [2016].

3 A joint motion parallax and disparity model

In contrast to previous work, we assess the joint influence of motion parallax and binocular disparity on depth perception. To this end, we propose a model which describes a relation between the strength of these cues and the depth impression they induce. A direct advantage of such an approach is that the cue fusion is an integral part of the model, and it does not need to be modeled separately. The model will allow us to identify combination of motion parallax and disparity values that result in a similar depth impression. We will use this equivalence in our applications to reduce binocular disparity in regions where depth perception is well supported by motion parallax.

3.1 Methods

To acquire data for building our model, we conducted a perceptual experiment where subjects were asked to match two stimuli according to the depth they provided. One of the stimuli contained both motion parallax and binocular disparity, while the other had only a disparity signal. The matches provided by users allowed us to find the disparity signal that is equivalent to a given parallax-disparity combination.

Stimuli The stimuli were random-dot stereograms with sinusoidal corrugations in depth. Both disparities and luminance-contrast values were windowed by a Gaussian function (Fig. 3, left). The spatial depth corrugation frequency was 0.4 cpd — the peak sensitivity of the HVS to both binocular disparity and motion parallax [Rogers and Graham 1982]. The luminance pattern was a random pattern with a uniform intensity distribution, which was additionally filtered using a low-pass filter with the cutoff frequency of 10 cpd. To avoid texture depth cues that could provide additional information about the presented shape, the pattern was flatly projected on the surface. Stimuli also included a central black spot of 0.01 degrees to aid fixation. In our experiment, we used two types of stimuli: *dynamic* stimuli with both motion parallax and disparity, and *static* stimuli with disparity but without motion parallax. The dynamic stimuli were translated horizontally across the screen (Fig. 3, right) at different speeds.

We parameterize the stimuli using their relative depth from motion parallax $m = \Delta f/f$, binocular disparity $d = |\mathcal{D}_d(\mathbf{A}) - \mathcal{D}_d(\mathbf{B})|$, and retinal velocity $V = \mathrm{d}\theta/\mathrm{d}t$ (Fig. 2), i.e., (d_i, m_i, V_i). A static stimulus is a special case of the dynamic one, i.e., $(d_i, 0, 0)$. Two possible parameterizations can be considered for the velocity: the

absolute ocular velocity of the stimuli pursuit on the screen $\nu = \mathrm{d}\alpha/\mathrm{d}t$, or the relative retinal velocity of peaks and troughs in the stimuli $V = \mathrm{d}\theta/\mathrm{d}t$ (both expressed in arcmin/s). We use V in the stimuli parameterization as it more directly relates to the visibility of motion parallax.

Note that stimuli can express different combinations of motion parallax and binocular disparity through the experiments, which do not need to be consistent. We specifically seek to measure the effect of non-consistent combinations, namely compressed and expanded disparity, which is important for applications such as those shown in Sec. 5.

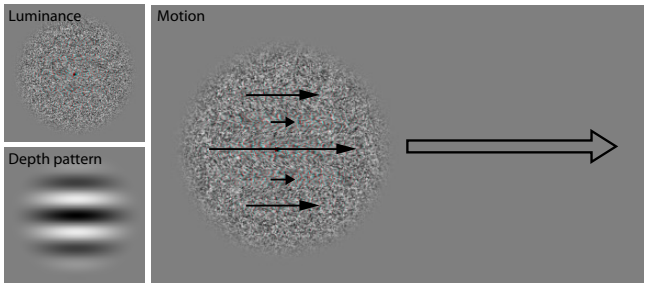

Figure 3: *Stimulus used in our experiment.* Left*: Luminance and depth of our stimuli.* Right: *Anaglyph version of the same. For the dynamic stimulus variant, translating horizontally as indicated by the large arrow, motion parallax results in the flow depicted by the small arrows.*

Equipment Stimuli were presented on a Zalman ZM-M240W polarized-glasses 3D display, with a screen size of $24''$ and a spatial resolution of 1920×1080 pixels, at a viewing distance of 60 cm.

Participants Twenty-seven subjects (5 F, 22 M, 24 to 27 years of age), who had normal or corrected-to-normal vision and passed a stereo-blindness test, took part in the experiment.

Procedure During each trial, subjects were given a pair of static and dynamic stimuli. They were instructed to fixate onto the marked central spot and adjust the disparity of the static stimulus until the perceived depth matched the one of the dynamic stimulus. This was done by pressing left and right arrow keyboard keys. The stimuli were displayed sequentially and subjects could switch between them at will by pressing a key. A 500 ms-blank screen followed each switch. When switching to a dynamic stimulus, visibility was enforced for at least 1000 ms before switching back to a static stimulus to guarantee that the subject had time to observe the motion and did not judge depth based on binocular disparity alone. For each dynamic stimulus each subject performed two trials. Performing the complete set of measurements took around 45 minutes per subject.

The three-dimensional stimulus space was covered by a set of samples combined from two subsets: In the first one, five regular steps of disparity and motion parallax were chosen from a typical range of values: $d \in [0..20]$ arcmin and $m \in [0..0.1]$, resulting in 25 samples. Stimulus translation velocity of 500 arcmin/s was chosen as it was observed to be safely supra-threshold in a pilot experiment.

Since the existence of detectable motion is a necessary condition for the perception of depth from motion parallax, five additional samples with stimulus translation velocity that varied from 0 to 300 arcmin/s were added for three disparity and motion parallax combinations $(0, 0.025, .)$, $(10, 0.075, .)$, and $(5, 0.100, .)$. These points were chosen to lie in the part of the sampling space where the disparity is smaller than it would be in the real world for an object defined by the given motion parallax. This way the contribution of parallax to the estimated depth can potentially be significant. This helps us to derive the threshold velocity required for motion parallax detection. As a result, 15 additional samples are measured, and for two repetitions, the experiment totals 80 samples per subject.

3.2 Data analysis and model fitting

The data from the experiment was used to derive an analytical model of perceived depth due to a combination of motion parallax and disparity. The model maps a joint stimulus involving both the motion and disparity cues to the matching depth that is perceived using only disparity. We call this the matching static disparity. Please refer to the supplemental materials for detailed results of our measurements.

We seek a function $\Phi : \mathbb{R}^4 \to \mathbb{R}$ of an equivalent static disparity for a combination of motion parallax m, binocular disparity d, velocity V, and angular distance s as a separation between neighboring peaks and troughs of our 0.4 cpd depth corrugation stimulus. We obtain the function by fitting an analytic expression to our experimental data. This is done in two steps: First, we fit a 2D supra-threshold velocity model $\Phi(m, d, .)$ for m and d as two free parameters, while the angular distance s stays fixed and the relative velocity V is maintained supra-threshold for all stimuli (Fig. 4a). Second, we model the effect of near-threshold velocities V using an additional, separable, multiplicative 2D function $\Psi(V, s)$ which depends only on V and s (Fig. 4c). Effectively, we assume that the two steps are separable, i.e., the velocity influence is independent of the rest of the parameters. We express the full 4D model by introducing Ψ into Φ as a scaling factor which can be efficiently pre-computed.

Supra-threshold velocity For supra-threshold velocity $\hat{V}$, a specific stimulus angular distance $x = 1.25$ arcmin and each constant d, we express perceived depth as a function of m by fitting a linear function to our data:

$$\Phi_{d,\hat{V}}(m, .) = am + b. \tag{1}$$

To obtain a model for varying d we interpolate the scale a and bias b. For b we use linear interpolation, as it follows the identity prediction for depth perception as a function of disparity. A linear interpolation of the slope a would result in unreliable predictions in extrapolated regions. Instead, we use a normalized sigmoid function λ. Additionally, we enforce that in the absence of both motion parallax and binocular disparity, the perceived depth remains zero, i.e., $\Phi_{\hat{V}}(0, 0) = 0$. The resulting function describing the perceived depth at supra-threshold levels is given by:

$$\Phi_{\hat{V}}(m, d) = (c_1 \cdot \lambda(d) + c_2) \cdot m + c_3 \cdot d, \tag{2}$$

$$\lambda(d) = \frac{2}{1 + e^{-d\beta}} - 1, \tag{3}$$

where $c_1 = -50.88$, $c_2 = 68.56$, and $c_3 = 1.006$ (DOF-adjusted $R^2 = 0.98$) are the fitting parameters to our experimental data and an integral part of the model used in our application (Fig. 4a). We set $\beta = 0.25$ to guarantee that the sigmoid saturates for the maximum disparity considered in the experiment ($d = 20$ arcmin) so that no extrapolation beyond our measurements occurs. The slope in the m-axis direction for $d \geq 20$ arcmin is then constant and equal to the sum $c_1 + c_2$. This determines the contribution of parallax to the total perceived depth for high disparity magnitudes. The value c_2 determines the slope of $\Phi_{\hat{V}}(m, d)$ as a function of m for $d = 0$, that is, the effect of motion parallax as the only available depth cue in the absence of binocular disparity. The value of c_3 determines the slope in the d-axis direction for $d \geq 20$ arcmin. As its value is close

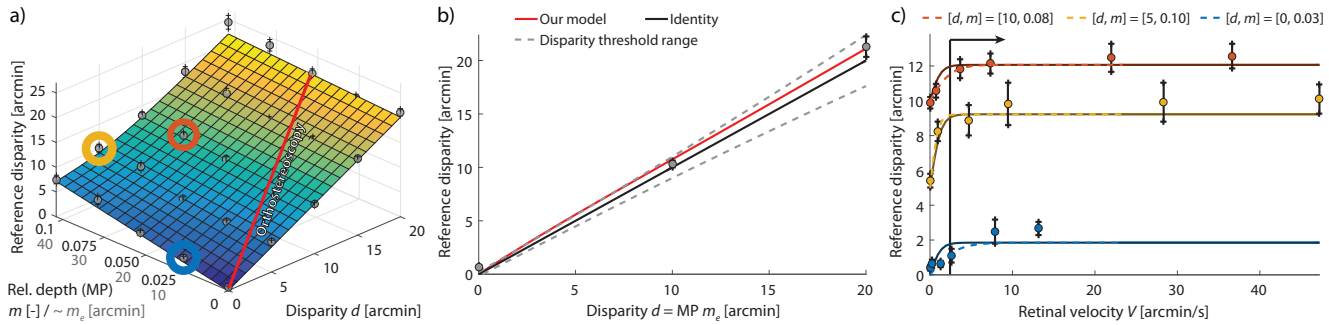

Figure 4: a) *Data acquired from our experiment* (gray dots) *and our model showing the matched static disparity* (vertical axis) *as a function of binocular disparity and relative depth from motion parallax* (x and y axis). *The red line shows the points corresponding to real-world observation conditions ($d = m$), plotted in b).* b) *Comparison of measured points (gray dots), our model (red line) and the theoretical identity (blue line) for real-world observation conditions ($d = m_e$). Dashed lines mark deviation from identity within the disparity discrimination thresholds for a given disparity [Didyk et al. 2011]. Difference under the threshold range confirms that our model does not predict more depth for a moving stimulus than for a static stimulus if conflict of cues is avoided.* c) *Perceived depth as a function of retinal velocity of motion V for three selected points from the two-dimensional space in a). Dashed lines are fitted separately for each sequence while full lines were fitted jointly to build our model in Eq. 4. Corresponding points for the maximum velocity marked in a). Vertical line is the 95th percentile threshold value.*

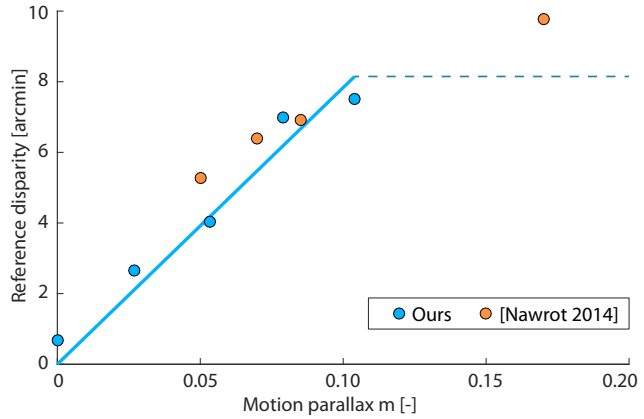

Figure 5: *Our measured data for the case of $d = 0$ and our fitted model* (blue) *compared to data of Nawrot et al. [2014, Fig. 6] for $\mathrm{d}\alpha/\mathrm{d}t = 3.3$ deg/s and viewing distance 54 cm* (orange). *Our model is clamped for $m > 0.1$ to avoid extrapolation beyond the measurement points.*

to one, our model provides safe extrapolation for larger disparities by matching each d to itself.

The comparison in Fig. 5 shows that our measurements covered a lower range of m values than those of Nawrot et al. [2014]. The observations are comparable for $0.05 < m \leq 0.10$, while a potential non-linearity was observed by Nawrot et al. [2014] for $m = 0.20$. Therefore, an extrapolation of our model for motion parallax magnitudes outside of the range we measured ($m > 0.10$) would be dangerous, as no conclusions about further behavior of the function can be made based on our measurements. For our applications, we clamp m to the range 0 to 0.10, which conservatively prevents a possible exaggeration of the contribution that larger parallax has to the depth perception.

Given the knowledge of absolute depth, we can convert m to the equivalent disparity m_e in arcmin and then find orthostereoscopic points where $m_e = d$. Such conditions remove depth cue conflicts, and therefore should create similar perception as the static stimulus.

Fig. 4b shows that the deviation between our measured model and identity to a static reference is within the range defined by disparity discrimination thresholds at different disparity levels for depth corrugation stimuli of 0.4 cpd [Didyk et al. 2011].

Near-threshold velocity Fig. 4c demonstrates the results of depth matching task between a static disparity signal and a stimuli containing motion parallax for three points in our stimuli space. The measurements show the inability of the HVS in perceiving depth from motion parallax for very low values of relative motion V. In such cases, binocular disparity acts as the only depth cue. Once a certain value of V is reached, motion is perceived, and motion parallax depth cue is triggered contributing to the perceived depth. We model this effect using a coefficient function $\Psi(.)$ that describes visibility of motion parallax. It scales its effect between zero (perceived depth is determined by binocular disparity alone) and our supra-threshold model:

$$\Phi(m, d, \Psi(.)) = d + \Psi(.) \cdot (\Phi_{\hat{V}}(m, d) - d) \tag{4}$$

We first approximate $\Psi(.)$ by fitting a sigmoid function to our measurements (Fig. 4c) as:

$$\Psi_0(V) = \frac{2}{1 + e^{-V \cdot c_4}} - 1 \tag{5}$$

with $c_4 = 1.553$ (DOF-adjusted $R^2 = 0.98$). Our velocity model saturates for $V > 3.0$ arcmin/s, which is similar to differential motion thresholds of 3.2 arcmin/s as measured for moving bars in the fovea [McKee and Nakayama 1984, Fig. 2]. As speed measurements close to zero are unreliable, the sigmoid function could lead to an overestimation of perceived depth for a near-static stimulus. To address this, we threshold the sigmoid at the 95th percentile and obtain:

$$\Psi_1(V) = \begin{cases} 1 & \text{when } V \geq v_c, \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $v_c \approx 2.36$ (Fig. 4c).

Because our matching experiment was performed for sinusoidal depth corrugations with spatial frequency of 0.4 cpd, function Ψ_1 is valid for the angular separation of $s = 1.25$ deg. In order to extend the model to arbitrary values of s, we assume that the detection

threshold of relative motion for a pair of targets is proportional to their angular separation. This is inspired by the study on discriminating separation differences between two dots [Yap et al. 1989]. Consequently, we model the velocity dependency as:

$$\Psi(V, s) = \Psi_1 \left(\frac{1.25}{s} V \right), \tag{7}$$

and state $\Psi(.) = \Psi(V, s)$.

3.3 Discussion

We compare our measurements to those in the existing literature that have analyzed perceived depth induced by motion parallax in the absence of binocular disparity ($d = 0$), and find that they are consistent. In particular, Nawrot et al. [2014] report a foreshortening of perceived depth in the presence of motion parallax for similar viewing conditions. Similarly, Durgin et al. [1995] describe foreshortening of depth from motion parallax in the absence of binocular disparity, albeit using a substantially different procedure.

The part of the space between $d = 0$ and $d = m$ (the left side with respect to the red line in Fig. 4a) corresponds to the region where binocular disparity is compressed with respect to the geometrically correct motion parallax; in this region, depth induced by motion parallax increases the perceived depth. This perceived depth enhancement can be used to reallocate the binocular disparity budget, which is typically limited in practical applications, as shown in Sec. 5. When we express *the relative depth from motion parallax* m by means of equivalent disparity m_e (the gray labels in Fig. 4a) we see that in an extreme case with zero binocular disparity ($d = 0$), motion parallax $m_e = 40$ arcmin alone induces depth corresponding to a static disparity of 7.5 arcmin or $40/7.5 \approx 20\%$ of veridical (geometric) depth.

The rest of the space, between $d = m$ and $m = 0$ (the right side with respect to the red line in Fig. 4a), depicts the case when disparity is expanded with respect to its original magnitude. In this case, motion parallax reduces perceived depth. Note that since motion parallax corresponds to the actual geometric configuration, it always drives the perceived depth towards its veridical value, irrespective of how binocular disparity is distorted. This is in accordance with the modified weak fusion model [Landy et al. 1995], in which relative weights of interacting cues are dependent on their reliability.

4 Motion parallax for points and pixels

The model presented in the previous section measures the contribution of motion parallax to the perceived depth of a scene in the presence of binocular disparity. In this section, we describe how this model, which was derived from simple stimuli, can be applied to obtain perceived depth between image pixels. The input is a single animation frame $I : \mathbb{N}^2 \rightarrow \mathbb{R}^3$, and as the output we seek a measure of perceived depth between pixels always assuming an observer fixation on the currently processed pixel. We demonstrate the method for computer-generated content, for which all scene information is available. The generalization of our technique to natural video streams is discussed in Sec. 7.

Motion parallax is well-defined for a pair of points (Fig. 2). However, its definition is usually limited to cases where the viewing position changes, yielding a globally-consistent motion of the scene. We extend this definition by observing that motion parallax introduces a depth impression in all regions of the scene that exhibit rigid motion. Therefore, to compute depth from both motion parallax and binocular disparity in our image I, we first need to determine whether two pixels, and thus two points in the scene, follow the same rigid transformation.

To detect whether two image pixels, $\mathbf{x}_A$ and $\mathbf{x}_B$, undergo the same rigid transformation, we first obtain for each pixel $\mathbf{x}$ its 3D world position $\mathcal{P}(\mathbf{x})$, and a transformation matrix $\mathcal{M}(\mathbf{x})$ describing its change in position over time. For the computer generated content, this can be easily done by storing a position buffer and transformation matrices for all pixels. For natural content, one needs to perform 3D scene reconstruction. In the following text, we assume that for each pixel in the image, we have given the two functions $\mathcal{P} : \mathbb{N}^2 \rightarrow \mathbb{R}^4$ expressed in homogeneous coordinates, and $\mathcal{M} : \mathbb{N}^2 \rightarrow \mathbb{R}^{3\times4}$ providing an affine transformation matrix over time between frames for each pixel.

Rigidity of the transformation of the two points $\mathbf{x}_A$ and $\mathbf{x}_B$ can be checked by comparing transformation matrices $\mathcal{M}(\mathbf{x}_A)$ and $\mathcal{M}(\mathbf{x}_B)$. However, a direct comparison of the two matrices is usually unstable and not robust due to small inaccuracies. Inspired by work of Braunstein et al. [1990], we propose a different measure of rigidity. Given two pixels $\mathbf{x}_A$ and $\mathbf{x}_B$, we compare the results of applying their transformation matrices $\mathcal{M}(\mathbf{x}_A)$ and $\mathcal{M}(\mathbf{x}_B)$ to a set of n different pixels $\mathcal{N} = \{\mathbf{x_i} \in \mathbb{R}^2 : 1 \leq i \leq n\}$, and compute:

$$\gamma(\mathbf{x}_A, \mathbf{x}_B) = \frac{1}{|\mathcal{N}|} \sum_{\mathbf{x}_i \in \mathcal{N}} \| \mathcal{M}(\mathbf{x_A})\, \mathcal{P}(\mathbf{x}_i) - \mathcal{M}(\mathbf{x_B})\, \mathcal{P}(\mathbf{x}_i) \|_2. \tag{8}$$

To reliably measure the difference between two transformations, $\mathcal{N}$ should contain at least three points that are not co-linear. In our case, we use a 3×3 pixel neighborhood of $\mathbf{x}_A$ as $\mathcal{N}$. We assume that the two matrices correspond to the same rigid motion if $\gamma(\mathbf{x}_A, \mathbf{x}_B)$ is small, thus defining the final rigidity measure for the pixels as:

$$\Gamma(\mathbf{x}_A, \mathbf{x}_B) = \begin{cases} 1 & \text{when } \gamma(\mathbf{x}_A, \mathbf{x}_B) < \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

This measure allows for a small amount of non-rigid motion, but in general, ϵ should be small to tolerate only small differences between the two transformations. In our applications, we use ϵ as 0.001 times the diameter of the $\mathcal{P}$ value domain bounding sphere.

Once we have determined rigidity Γ, in order to apply our model to the pixels, we need to know their velocities and spatial distances as well as their depth and pixel disparity. For the pixel disparity we are interested in the per-pixel angular measure (i.e., vergence angles, Fig. 2) $\mathcal{D}_d : \mathbb{N}^2 \rightarrow \mathbb{R}$. This is computed taking into account the viewing distance, and the interocular distance of a standard observer (6.4 cm). For velocities, we compute absolute velocities, $\mathcal{D}_v : \mathbb{N}^2 \rightarrow \mathbb{R}^2$, measured in visual angles, and similarly we express the angle between screen-space locations of both points $\mathcal{D}_s : \mathbb{N}^4 \rightarrow \mathbb{R}$. The linear depth $\mathcal{D}_m : \mathbb{N}^2 \rightarrow \mathbb{R}$ is only used to express the motion parallax m and thus needs only to be known up to a factor.

Given these, we can apply our parallax visibility model to a pair of pixels to obtain the contribution of motion parallax between them to perceived depth as:

$$\zeta(\mathbf{x}_A, \mathbf{x}_B) = \Psi(\Gamma(\mathbf{x}_A, \mathbf{x}_B) \cdot \Delta\mathcal{D}_v, \Delta\mathcal{D}_s), \tag{10}$$

and our full model to obtain the perceived depth between them as:

$$\Theta(\mathbf{x}_A, \mathbf{x}_B) = \Phi\left(\Delta\mathcal{D}_m, \Delta\mathcal{D}_d, \zeta(\mathbf{x}_A, \mathbf{x}_B)\right), \tag{11}$$

where:

$$\begin{aligned} \Delta\mathcal{D}_m &= \frac{|\mathcal{D}_m(\mathbf{x}_A) - \mathcal{D}_m(\mathbf{x}_B)|}{\max\left(\mathcal{D}_m(\mathbf{x}_A), \mathcal{D}_m(\mathbf{x}_B)\right)}, \\ \Delta\mathcal{D}_d &= |\mathcal{D}_d(\mathbf{x}_A) - \mathcal{D}_d(\mathbf{x}_B)|, \\ \Delta\mathcal{D}_v &= ||\mathcal{D}_v(\mathbf{x}_A) - \mathcal{D}_v(\mathbf{x}_B)||, \\ \Delta\mathcal{D}_s &= \mathcal{D}_s(\mathbf{x}_A, \mathbf{x}_B). \end{aligned} \tag{12}$$

Note that, according to the definitions above, $\Delta\mathcal{D}_d$ is the disparity between $\mathbf{x}_A$ and $\mathbf{x}_B$ (with d in the model as given by Eq. 4), $\Delta\mathcal{D}_m$ is our measure for motion parallax between both pixels (m in the model), the relative velocity $\Delta\mathcal{D}_v$ between them (V in the model) is a difference of per-pixel optical flow vectors and $\Delta\mathcal{D}_s$ between them (s in the model) is directly their angular distance.

With the definition in Eq. 11 we can compute the relative depth between any two pixels of the input image sequence. The application of this to complex images for disparity manipulation is presented in the next section.

5 Parallax-aware disparity manipulations

Many solutions exist for manipulating disparity signal in stereoscopic content. [Lang et al. 2010; Didyk et al. 2011; Didyk et al. 2012b; Masia et al. 2013a]. Some methods aim specifically at enhancing apparent depth without expanding overall disparity range [Chapiro et al. 2015; Didyk et al. 2012a]. Despite of the similar goals of our technique, none of the previous works accounts for motion parallax cue which, as shown in our experiment, greatly contributes to the perceived depth. The contribution is especially significant for compressed disparities where the contribution of motion parallax and binocular disparity to perceived depth is comparable. To exploit this observation, we propose a new disparity mapping operator (Fig. 6) which takes advantage of our motion parallax model. The goal of the technique is to compress disparity range for regions that are supported by motion parallax, and use the additional disparity budget in static regions. As our goal is only to reallocate disparities, the technique can be easily combined with other existing disparity mapping approaches by applying them to our input disparity map $\mathcal{D}_d(\mathbf{x})$ beforehand.

5.1 Overview

The input to our technique is a stereoscopic sequence with disparities mapped to fit a desired range. In the first step, we extract optical flow and 3D transformation matrices for each pixel of the input S3D sequence (Fig. 6a) either by means of computer vision or by directly dumping necessary information from the rendering pipeline (Fig. 6b). Second, we detect whether the transformation between neighboring pixels is both rigid and sufficiently fast according to our velocity model. This way, the visibility of motion parallax for each pair of neighboring pixels is predicted (Fig. 6c). Next, we use our model to estimate the contribution of the parallax to the perceived depth, and redistribute the depth budget by scaling local disparity gradients accordingly and storing them in a histogram (Fig. 6d). We then reconstruct a global disparity mapping curve from the histogram (Fig. 6e). As the last step, we remap the input disparity, and use an image warping technique to generate a new S3D sequence (Fig. 6f). In the following subsection, we describe details of each of these steps.

5.2 Parallax map

The parallax map encodes the relative visibility of motion parallax for a pair of pixels (Fig. 6c). Such pixels must have both a rigid mutual transformation and a relative motion speed above the visibility threshold.

For the purpose of this technique, we consider all effects in a single-scale fashion, and process local disparities and local parallax as both horizontal and vertical gradients $\Delta\mathbf{x}$ between neighboring pixels $\mathbf{x}_A$ and $\mathbf{x}_B$. Consequently, we refer to differential measures from Eq. 12 using a short notation $\Delta\mathcal{D} = \Delta\mathcal{D}(\mathbf{x}_A, \mathbf{x}_B)$. We process both vertical and horizontal gradients in the same way, and omit this information further on.

The rigidity of two neighboring pixels can be directly evaluated as $\Gamma(\Delta\mathbf{x}) = \Gamma(\mathbf{x}_A, \mathbf{x}_B)$ (Eq. 9). However, the visibility of motion between the two neighboring pixels may be too small to qualify for a visible parallax according to $\Psi_V(V)$. This, however, is not in line with motion perception of larger objects. The largest sensitivity for differential motion was detected for targets of 1 visual degree size [McKee and Nakayama 1984]. Therefore, instead of evaluating rigidity only for neighboring pixels, for each pixel $\mathbf{x}$, we search a local neighborhood (1 visual degree) for a rigidly connected location $\mathbf{x}'$ with the highest visibility of motion parallax $\zeta(\mathbf{x}, \mathbf{x}')$ (Eq. 10):

$$\mathbf{x}' = \max_{\mathbf{x}'} \zeta\left(\mathbf{x}, \mathbf{x}'\right). \tag{13}$$

The parallax map M at location $\mathbf{x}$ is then a product of both the local rigidity between neighboring pixels and the maximum motion parallax visibility (Fig. 6c):

$$M(\mathbf{x}) = \Gamma(\Delta\mathbf{x}) \cdot \zeta\left(\mathbf{x}, \mathbf{x}'\right). \tag{14}$$

5.3 Disparity scaling

We seek a disparity manipulation that provides images such that when both binocular disparities and motion parallax information are combined, they give depth perception specified by the input disparities $\Delta\mathcal{D}_d(\mathbf{x})$. This can be formally expressed by the following constraint:

$$\forall_{\mathbf{x}}\, \Delta\mathcal{D}_d(\mathbf{x}) = \Phi\left(\Delta\mathcal{D}_m(\mathbf{x}), \Delta\mathcal{D}'_d(\mathbf{x}), M(\mathbf{x})\right), \tag{15}$$

where $\Delta\mathcal{D}'_d$ are disparities of the output image. Although enforcing this constraint compresses only disparities that are supported by motion parallax, this increases the part of the disparity range dedicated to static parts. It is, therefore, sufficient to rescale the resulting disparity range to the original one at the end of the process. We use Eq. 15 to compute the target output gradients $\Delta\mathcal{D}'_d(\mathbf{x})$. As the function Φ is monotonic and all other variables are fixed, this can be done by numerically inverting it. The reader may refer to the supplemental materials for a numerical example of the described algorithm.

5.4 Curve construction

Once the constraint in Eq. 15 is enforced, a new vergence map $\mathcal{D}'_d$ can be recovered from disparities $\Delta\mathcal{D}'_d$ by solving a Poisson equation in a similar way as was done in [Fattal et al. 2002] for luminance. However, such a solution does not guarantee depth ordering preservation and temporal coherence. To address these issues, we propose to construct a global mapping curve based on $\Delta\mathcal{D}_d$ and $\Delta\mathcal{D}'_d$ using a custom histogram of gradients. To this end, we construct a histogram $H : \mathbb{I} \to \mathbb{R}$ that contains $n = 1024$ bins which split the range of $\mathcal{D}_d$. We found that a lower number of bins does not provide a measurable performance improvement and a higher number does not yield a perceptible quality increment. Each bin H_i stores information about values between vergence a_i and b_i where:

$$\begin{aligned} a_i &= \min(\mathcal{D}_d) + \frac{i}{n} \cdot [\max(\mathcal{D}_d) - \min(\mathcal{D}_d)] \\ b_i &= a_i + \frac{1}{n} \cdot [\max(\mathcal{D}_d) - \min(\mathcal{D}_d)]. \end{aligned} \tag{16}$$

We construct the histogram H by distributing each gradient $\Delta\mathcal{D}'_d$ to all bins covered by the interval between values $\mathcal{D}_d(\mathbf{x})$ and $\mathcal{D}_d(\mathbf{x}) +$

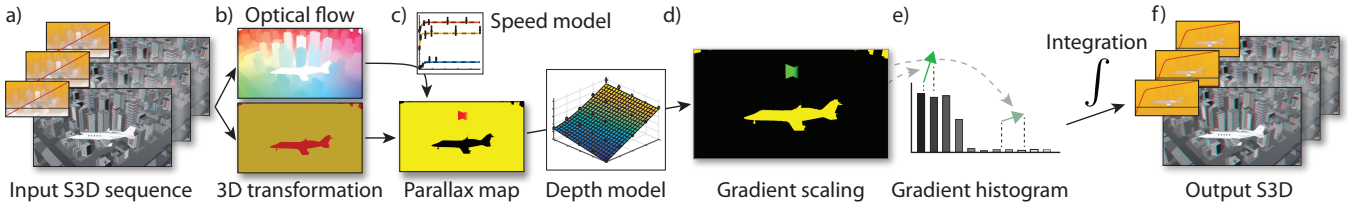

Figure 6: *Our disparity mapping algorithm. Optical flow and 3D transformations (b) are extracted for the input S3D sequence (a) and used together with our relative motion model to predict perceptibility of motion parallax in the X and Y directions (c). Our model of perceived depth is then sampled to estimate necessary scaling of each disparity gradient (d). Scaled gradients are accumulated in a histogram (e) and integrated to construct a new mapping curve that is applied to produce the final S3D content (f).*

Step	Time [ms]
Rendering	1.5
Parallax map (Sec. 5.2)	24.5
Gradient scaling (Sec. 5.3)	0.9
Curve construction (Sec. 5.4)	4.3
Image warping	3.9
Total	35.1

Table 1: *Performance break down of our non-optimized implementation computing a single frame at 720p resolution on a computer with Intel Xeon E5 and GeForce GTX 980 Ti.*

$\Delta\mathcal{D}_d(\mathbf{x})$:

$$H[i] = \sum_{\mathbf{x}\in R(i)} \frac{\Delta\mathcal{D}'_d(\mathbf{x})}{\Delta\mathcal{D}_d(\mathbf{x})}, \quad i \in 0, 1 \ldots |\Delta\mathcal{D}'_d| - 1, \tag{17}$$

such that:

$$R(i) = \{\mathbf{x} : \mathcal{D}_d(\mathbf{x}) < b_i \text{ and } \mathcal{D}_d(\mathbf{x}) + \Delta\mathcal{D}_d(\mathbf{x}) \geq a_i\}.$$

Effectively, we add to each of these bins the slope of the future mapping curve $\Delta\mathcal{D}'_d(\mathbf{x})/\Delta\mathcal{D}_d(\mathbf{x})$. For each gradient, we also accumulate weights w into an equivalent histogram $W : \mathbb{I} \to \mathbb{R}$. They correspond to the sensitivity of the HVS to the changes in disparities. This favors preserving small gradients. Formally, to each gradient $\Delta\mathcal{D}_d$, we assign the following weight :

$$w(\Delta\mathcal{D}_d) = \frac{1}{thr(\Delta\mathcal{D}_d)}, \tag{18}$$

where $thr(d)$ is a disparity discrimination threshold for pedestal disparity d [Didyk et al. 2011, Sec. 4]. The idea is that compression of large disparities has a smaller effect on the overall depth perception as these are likely to stay noticeable. For our method this also gives additional freedom for redistributing depth between distant objects.

After all gradients are accumulated in H, we normalize the value in each bin by the accumulated weights in W. Then, we compute the remapping between input pixel disparities $\mathcal{D}_d$ and new disparities $\mathcal{D}'_d$ (both expressed in vergence angles) as a cumulative sum of the histogram, which is normalized to the range of input disparities $\mathcal{D}_d$. Formally, the remapping function R can be defined as follows:

$$R(d) = [\max(\mathcal{D}_d) - \min(\mathcal{D}_d)] \cdot \frac{\sum_{i=0}^{h(d)} H_i/W_i}{\sum_{i=0}^{n-1} H_i/W_i} + \min(\mathcal{D}_d), \tag{19}$$

where the function $h(d)$ provides the index of the bin that corresponds to the value of d. Please note that if no gradients were overlapping in the histogram, the mapping constructed in such a way would provide $\Delta\mathcal{D}'_d$ that fulfill our initial constraint (Eq. 15). However, due to a large number of gradients which overlap, the mapping only tries to satisfy all of them. After the vergence information is remapped according to the above equation, we recompute pixel disparity information and use it for modifying the input frames. This last step is done by using image-based warping technique to generate new stereoscopic images [Didyk et al. 2010]. The result examples are presented in Fig. 7. The entire algorithm in our non-optimized implementation is executed in real-time for the input resolution of 1280×720 pixels. A breakdown of times for individual steps in Table 1 reveals that a naïve implementation of neighborhood sampling from Eq. 13 is a major bottleneck and a candidate for optimization.

Although histogram-based approaches typically feature better temporal coherence than local processing, we noticed that minor depth oscillation may be observed due to the normalization step. The coherence is also affected by objects leaving the scene. We improve the temporal coherence by an additional filtering of the remapping curves over time. Usually, a window of the ten last frames is used. Similar techniques have been applied before in a similar context [Oskam et al. 2011; Didyk et al. 2012b].

5.5 Application to autostereoscopic displays

One of the main drawbacks of current autostereoscopic displays is a significant angular aliasing when a large depth range is shown on such a screen [Zwicker et al. 2006]. The aliasing originates from the limited angular resolution of the display, and it reveals itself as an unfavorable ghosting artifact which breaks the stereoscopic effect (Fig. 8a). We can utilize our disparity manipulation technique to improve the image quality on such screens. To this end, instead of reallocating disparity signal as proposed before, we simply remove it from locations where motion parallax sufficiently supports depth perception. This reduces the required disparity range and thus the amount of visible aliasing. We follow the same algorithm as before with the exception of the final histogram construction. We do not map the resulting curve into the fixed depth range, but directly use the accumulated gradient values that encode the disparity just needed to deliver perception equivalent to a static image. We then construct the mapping curve directly as:

$$R_a(d) = [\max(\mathcal{D}_d) - \min(\mathcal{D}_d)] \cdot \sum_{i=0}^{h(d)} H_i/W_i + \min(\mathcal{D}_d). \tag{20}$$

We have tested our approach on a Full HD display Tridelity MV2600va that utilizes a parallax barrier to deliver 5 views. Please refer to Fig. 8 and the supplementary video for a captured screen comparison. We compared a linear mapping without and with our manipulation on top. Disparities of both results are centered around the screen to minimize the aliasing of the display [Zwicker et al. 2006]. While the reference footage suffers from aliasing in the front and far parts of the scene, our output was able to convey a very

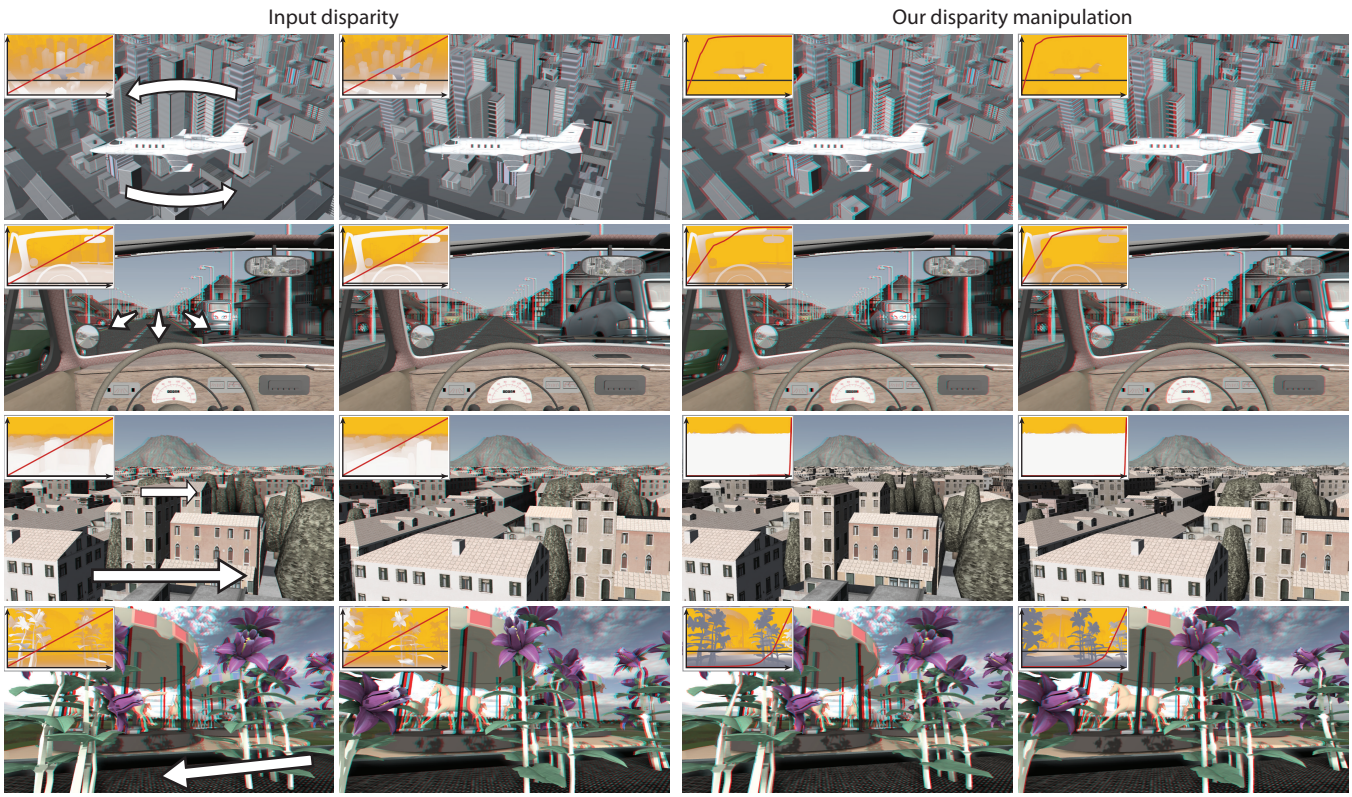

Figure 7: *Results of our parallax-aware disparity manipulation applied to four video sequences used in our user study (Sec. 6). The mapping curves along with the output disparity maps are shown in the insets. The blue color marks crossed disparity, yellow uncrossed disparity and white zero disparity. The arrows show the local motion direction.*

similar depth impression without violating the usable depth range and introducing visual artifacts due to severe aliasing.

6 Evaluation

Our experiment in Sec. 3 accounts for the combination of motion parallax and disparity cues but omits other depth cues that are present in complex images [Cutting and Vishton 1995]. To test the applicability of our model in such conditions we have performed a validation user study. We check if our method increases the 3D appearance of stereoscopic content by showing complex video sequences to study participants.

6.1 Method validation

Stimuli Four short video sequences (5-10 seconds) with a camera or scene motion (see Fig. 7) were used as stimuli. The stereoscopic videos that had disparities processed by our technique and the original input videos were played in a loop simultaneously side-by-side in a random order.

Task Participants were given unlimited time to compare both videos and answer the question "Which sequence is more 3D?" Each pair was shown twice, resulting in 8 trials.

Equipment The stimuli were presented using the polarized glasses technology on a $24''$ Zalman ZM-M240W display with a resolution of 1920×1080 pixels, at a viewing distance of 60 cm under normal, controlled office lighting. We used this display technology as it does not introduce temporal disparity artifacts potentially caused by a time-sequential presentation as described by Hoffman et al. [2011]. A chin rest was employed to prevent the participant from moving away from the optimal viewing angle.

Participants 12 participants (3 F, 9 M, 20 to 30 years old) took part in the study. All of them had normal or corrected-to-normal vision and passed a stereo-blindness test by describing the content of several random dot stereograms (RDS). The subjects were naïve to the purpose of the experiment.

Results The results of the study are presented in Fig. 9a. We observe a statistically significant (binomial test, $p < 0.05$) 60.0% preference of the 3D reproduction in our manipulated results. This confirms that our method succeeds in providing more depth impression by redistributing the same disparity budget. The results are affected by the poor performance in the *City flight* scene (Fig. 7, 1st row). We suspect that the relatively small screen coverage of the plane and great attention given to the city in the background caused many people to ignore the depth extension of the plane completely, and instead, focus on detecting small differences in the city. As our method aims to redistribute the disparity but does not change the parallax, the total sum of both cues is indeed smaller in the moving part if investigated in isolation, which was further facilitated by side-by-side presentation of the stimuli. A saliency predictor could possibly tackle this problem, but we decided not to include it into our pipeline to keep it clear and focused.

Figure 8: *Example of disparity compression for autostereoscopic displays. a) The original linear compression with the corresponding disparity map and mapping curve shown in the inset. b) Our manipulation applied on top of (a) to compress stronger regions that benefit from motion parallax. c) A photo of our display. Aliasing artifacts are pronounced in the content with only linear mapping (insets).*

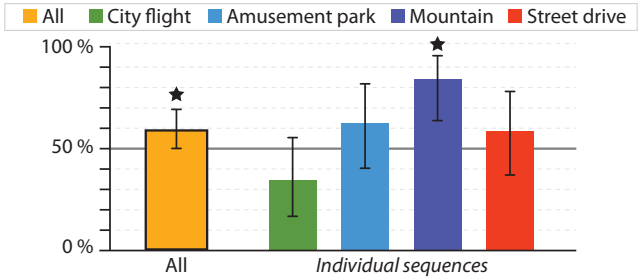

Figure 9: *Results of the first validation study comparing a linear mapping and our proposed method (Sec. 6.1). Confidence intervals are shown by the error bars and a significance in binomial test for $p < 0.05$ by star symbols.*

6.2 Modification

Next, we tested the significance of the role of our model in the proposed method by experimentally reducing its effect. As the mapping curve predicted by our method forced some regions of the scenes to be almost flat, we wanted to validate whether such behavior is appropriate. The goal of the following experiment was to see whether the scene can benefit if a small amount of disparity information is present everywhere. To this end, we repeated the validation study with a modified version of our method. Inspired by Larson et al. [1997], we adapted their iterative histogram adjustment algorithm, and used their constraint to guarantee that our mapping curve will never be less steep than a linear mapping to a target disparity range of 2.3 arcmin. This choice was motivated by the fact that most people (97.3%) have stereoacuity below this value [Coutant and Westheimer 1993]. This effectively guaranteed that most participants of our experiment should experience at least some stereo impression for such stimuli. In practice, a small disparity gradient was added to originally flat regions (Fig. 10).

Stimuli, Task, Equipment and Participants Stimuli presentation, task and equipment stayed the same as in Sec. 6.1 up to the modification to our method described above. A partially overlapping group of 12 participants (4 F, 8 M, 20 to 30 years old) took part in the study.

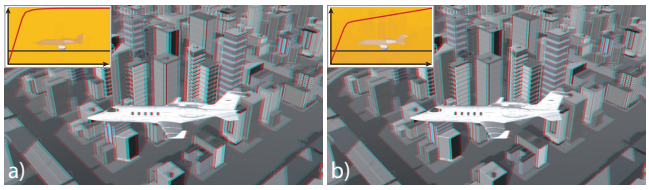

Figure 10: *Comparison of our proposed algorithm (a) and its modification evaluated in Sec. 6.2 (b).*

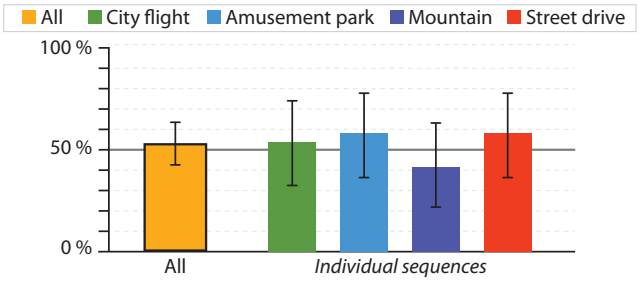

Figure 11: *Results of the second validation study comparing a linear mapping and the gradient adjusted mapping (Sec. 6.2). No significance was observed according to a binomial test for $p < 0.05$.*

Results The results are closer to chance level than previously (Fig. 9b). This demonstrates that the very small amount of disparity present in some of our stimuli in the original experiment was justified. For most of the scenes, the depth impression was reduced when the just-noticeable disparity was added to some regions, which effectively reduced this information in other parts of the scene and limited the effect of our approach. The differences with respect to the reference became too small to detect. This suggests that motion parallax not only enhances the perceived depth, but also elevates detection thresholds. Our model could capture the balance between motion parallax and disparity well, and therefore, we recommend our original method (Sec. 6.1) for practical applications.

7 Discussion and limitations

Our experiment in Sec. 3 was performed for isolated motion parallax and disparity cues, while we apply its outcome for complex images. Following the modified weak cue integration theory [Landy et al. 1995], which conforms well with many experimental data involving motion parallax [Howard and Rogers 2012, Ch. 30.2], the agreement of motion parallax cue with 2D pictorial cues should result in its possible promotion. This makes our model conservative in the sense of predicting the minimal amount of perceived depth enhancement due to motion parallax. The dependency on luminance contrast was previously integrated into the disparity model of Didyk et al. [2012b]. The validity of a similar relation for motion parallax still needs to be investigated.

Our model is also conservative with respect to our strong assumption on the transformation rigidity (Sec. 4), which ignores any motion parallax that can arise from non-rigid relative object motion [Luca et al. 2007]. We relegate to future work the extension of our experimental model (Sec. 3) to handle such non-rigid relative motion.

Motion parallax is just one of many monocular cues used by the HVS to estimate depth. In future work, similar depth equivalency models

could be derived for each of them, allowing for a more efficient disparity allocation. Additionally, this work is only concerned with the most common lateral motion parallax. Other parallax types have been described [Howard and Rogers 2012, Sec. 28.1.3] and their contribution to depth perception should be evaluated.

The *City flight* scene used in our user study (Sec. 6) demonstrated that enhancing the disparity gradient in non-salient regions at the cost of salient regions may not lead to an increase in the overall perceived depth. A possible extension of our method could account for this finding by integrating a saliency estimate into the weighting function in Eq. 18.

We described our method for detection and modeling depth from motion parallax for cases where all information about the scene is provided as an input. It is also possible to apply our technique to natural stereoscopic video sequences, which requires deriving all needed information using computer vision methods. For example, pixel disparity can be estimated using standard stereo correspondence methods such as [Brox et al. 2004; Zimmer et al. 2011]. Computing other information, such as depth, positions, a motion flow, is related to structure-from-motion techniques. While computing this information is non-trivial, a number of techniques exist which could be used to estimate the information [Szeliski 2010]. We leave a deeper investigation of the suitability of such methods for our purposes for future work.

Our model has been measured assuming a perfect match between the virtual camera and actual observation setups in terms of the field of view, the observer distance, and so on. In case of significant departures from such measurement conditions, e.g., in cinematographic applications, our model might not optimally predict the magnitude of perceived depth, and possibly new measurements would be required to accommodate relevant view configurations as well as disparity and motion parallax ranges in such conditions. Nevertheless, as discussed in Sec. 3.2, our model is protected against such out-of-range queries and should always lead to predictable behavior.

8 Conclusions and future work

We presented a method for predicting the impact of motion parallax on scene layout perception. To this end, we first conducted a psychovisual experiment in which we measured depth induced by motion parallax and related it directly to depth obtained from binocular presentation of static stimuli. Based on these measurements, we proposed a computational model that predicts the induced depth for complex image sequences. To our knowledge, this is the first work that tries to analyze and quantify structure perception from motion parallax for complex image sequences. A big advantage of our model is the compatibility with previously proposed image and video processing techniques. As a result, it can be easily incorporated into those techniques. We demonstrated this on several examples. Our model is also a significant step towards better understanding of perception for new output devices such as head-mounted or lightfield displays where motion parallax is an inherent cue obtained from observers' head movements.

Acknowledgements

We would like to thank Junaid Ali, Sumit Shekhar, and the anonymous subjects who took part in our perceptual studies. This work was partly supported by the Fraunhofer and the Max Planck cooperation program within the framework of the German pact for research and innovation (PFI).

References

BISTA, S., DA CUNHA, Í. L. L., AND VARSHNEY, A. 2016. Kinetic depth images: Flexible generation of depth perception. *The Visual Computer*, 1–13.

BRADSHAW, M. F., AND ROGERS, B. J. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research 36*, 21, 3457–3468.

BRADSHAW, M. F., AND ROGERS, B. J. 1999. Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research 39*, 18, 3049–56.

BRADSHAW, M. F., HIBBARD, P. B., PARTON, A. D., ROSE, D., AND LANGLEY, K. 2006. Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax. *Vision Research 46*, 17, 2636–2644.

BRAUNSTEIN, M. L., HOFFMAN, D. D., AND POLLICK, F. E. 1990. Discriminating rigid from nonrigid motion: Minimum points and views. *Perception & Psychophysics 47*, 3, 205–214.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, vol. 3024 of *Lecture Notes in Computer Science*, 25–36.

CAMPBELL, F., AND MAFFEI, L. 1981. The influence of spatial frequency and contrast on the perception of moving patterns. *Vision Research 21*, 5, 713–721.

CHAPIRO, A., HEINZLE, S., AYDIN, T. O., POULAKOS, S., ZWICKER, M., SMOLIC, A., AND GROSS, M. 2014. Optimizing stereo-to-multiview conversion for autostereoscopic displays. *Comp. Graph. Forum (Proc. Eurographics) 33*, 2, 63–72.

CHAPIRO, A., O'SULLIVAN, C., JAROSZ, W., GROSS, M., AND SMOLIC, A. 2015. Stereo from shading. In *Proc. of EGSR*.

COUTANT, B. E., AND WESTHEIMER, G. 1993. Population distribution of stereoscopic ability. *Ophthalmic and Physiological Optics 13*, 1, 3–7.

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 (Handbook Of Perception And Cognition)*, Academic Press, W. Epstein and S. Rogers, Eds., 69–117.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2010. Adaptive image-space stereo view synthesis. In *Vision, Modeling and Visualization Workshop*, 299–306.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2011. A perceptual model for disparity. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4, 96.

DIDYK, P., RITSCHEL, T., EISEMANN, E., MYSZKOWSKI, K., AND SEIDEL, H.-P. 2012. Apparent stereo: The cornsweet illusion can enhance perceived depth. In *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 82910N–82910N.

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. (Proc. SIGGRAPH) 31*, 6, 184.

DOMINI, F., CAUDEK, C., AND TASSINARI, H. 2006. Stereo and motion information are not independently processed by the visual system. *Vision Research 46*, 11, 1707–1723.

DURGIN, F. H., PROFFITT, D. R., OLSON, T. J., AND REINKE, K. S. 1995. Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance 21*, 3, 679.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Trans. Graph. (Proc. SIGGRAPH) 21*, 3, 249–256.

HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision 8*, 3, 33–33.

HOFFMAN, D. M., KARASEV, V. I., AND BANKS, M. S. 2011. Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth. *Journal of the Society for Information Display 19*, 3, 271–297.

HOWARD, I., AND ROGERS, B. 2012. *Perceiving in Depth*. Oxford University Press.

KRAL, K. 2003. Behavioural–analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behavioural Processes 64*, 1, 1–12.

LAMBOOIJ, M., FORTUIN, M., HEYNDERICKX, I., AND IJSSELSTEIJN, W. 2009. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology 53*, 3, 30201–1.

LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., AND YOUNG, M. 1995. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research 35*, 3, 389–412.

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. *ACM Trans. Graph. (Proc. SIGGRAPH) 29*, 4, 75.

LARSON, G. W., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics 3*, 4, 291–306.

LUCA, M. D., DOMINI, F., AND CAUDEK, C. 2007. The relation between disparity and velocity signals of rigidly moving objects constrains depth order perception. *Vision Research 47*, 10, 1335–1349.

MASIA, B., WETZSTEIN, G., ALIAGA, C., RASKAR, R., AND GUTIERREZ, D. 2013. Display adaptive 3D content remapping. *Computers & Graphics 37*, 8, 983–996.

MASIA, B., WETZSTEIN, G., DIDYK, P., AND GUTIERREZ, D. 2013. A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics 37*, 8, 1012 – 1038.

MCKEE, S. P., AND NAKAYAMA, K. 1984. The detection of motion in the peripheral visual field. *Vision Research 24*, 1, 25–32.

NAWROT, M., AND STROYAN, K. 2009. The motion/pursuit law for visual depth perception from motion parallax. *Vision Research 49*, 15, 1969–1978.

NAWROT, M., RATZLAFF, M., LEONARD, Z., AND STROYAN, K. 2014. Modeling depth from motion parallax with the motion/pursuit ratio. *Frontiers in Psychology 5*.

ONO, M. E., RIVEST, J., AND ONO, H. 1986. Depth perception as a function of motion parallax and absolute-distance information. *Journal of Experimental Psychology: Human Perception and Performance 12*, 3, 331.

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

PROFFITT, D., AND BANTON, T. 1999. Perceived depth is enhanced with parallax scanning. Tech. rep., University of Virginia-Cognitive Science Department.

ROGERS, B., AND GRAHAM, M. 1982. Similarities between motion parallax and stereopsis in human depth perception. *Vision Research 22*, 2, 261–270.

STROYAN, K. 2010. Motion parallax is asymptotic to binocular disparity. *arXiv preprint arXiv:1010.0575*.

SZELISKI, R. 2010. *Computer Vision: Algorithms and Applications*. Springer.

TURNER, J., BRAUNSTEIN, M. L., AND ANDERSEN, G. J. 1997. Relationship between binocular disparity and motion parallax in surface detection. *Perception & Psychophysics 59*, 3, 370–380.

ULLMAN, S. 1983. Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. *Perception 13*, 255–74.

V3© IMAGING, 2015. www.inv3.com.

WALLACH, H., AND O'CONNELL, D. N. 1953. The kinetic depth effect. *Journal of Experimental Psychology 45*, 4, 205.

WETZSTEIN, G., LANMAN, D., HIRSCH, M., AND RASKAR, R. 2012. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4, 1–11.

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

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

YAP, Y. L., LEVI, D. M., AND KLEIN, S. A. 1989. Peripheral positional acuity: Retinal and cortical constraints on 2-dot separation discrimination under photopic and scotopic conditions. *Vision Research 29*, 7, 789 – 802.

YOUNG, M. J., LANDY, M. S., AND MALONEY, L. T. 1993. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research 33*, 18, 2685–2696.

ZIMMER, H., BRUHN, A., AND WEICKERT, J. 2011. Optic flow in harmony. *International Journal of Computer Vision 93*, 3, 368–388.

ZWICKER, M., MATUSIK, W., DURAND, F., AND PFISTER, H. 2006. Antialiasing for automultiscopic 3D displays. In *Proc. of EGSR*, The Eurographics Association.