

Improving Perception of Binocular Stereo Motion on 3D Display Devices

Petr Kellnhofer^a Tobias Ritschel^{a,b} Karol Myszkowski^a Hans-Peter Seidel^a

^aMax-Planck-Institut für Informatik, Campus E1.4, Saarbrücken, Germany;

^bSaarland University, Uni-Campus Nord, Saarbrücken, Germany

ABSTRACT

This paper investigates the presentation of moving stereo images on different display devices. We address three important issues. First, we propose temporal compensation for the Pulfrich effect when using anaglyph glasses. Second, we describe, how content-adaptive capture protocols can reduce false motion-in-depth sensation for time-multiplexing based displays. Third, we conclude with a recommendation how to improve rendering of synthetic stereo animations.

Keywords: Stereoscopic 3D, Pulfrich effect, Protocols, Motion in depth, Warping

1. INTRODUCTION

Stereoscopic 3D imaging is nowadays a wide-spread and affordable means to achieve a convincing game or movie experience. The human visual system (HVS) uses a combination of different perceptual cues to estimate spatial layout from 3D images. Different from common luminance imaging, stereo 3D display technology provides additional binocular disparity cues. An extensive body of work has investigated various static properties of binocular stereo content, such as manipulation and editing to achieve viewing comfort both in terms of technical requirements as well as in faithful perceptual modeling. In this work, we describe the interplay of binocular stereo motion with different display and rendering technologies. Temporal disparity changes can introduce conflicts with other cues that hamper scene understanding, e.g. occlusion if apparent depth order is altered. This increases the difficulty of depth-oriented tasks in simulations and games where the HVS combines various depth cues to estimate the spatial layout of objects and their motion. We find that motion can have a strong influence on perceived depth, especially in connection with limitations of display devices in everyday use. Such display devices vary in several key properties:

- **Spatial resolution.** Displays of high spatial resolution can present smaller changes of motion, be it in the screen plane or in depth, resulting in smoother motion.
- **Temporal resolution.** The display can repeat frames if its refresh frequency is higher than the one of the image data source.
- **Temporal protocol.** The left and right eye images can be presented simultaneously or sequentially as well as it can be presented continuously (hold-type LCD) or in flashes (cinema projector).
- **Multiplexing.** Polarization, color coding or parallax barriers are commonly used to separate images between the left and right eye if presented simultaneously.

In physical reality the viewing angle which is different for each eye guarantees perception of a pair of stereo images. Conventional displays show the same content to both eyes and therefore cannot reproduce binocular disparity. Therefore “multiplexing” has to be added between the display and the viewer. Such multiplexing is most commonly done using, either

Further author information: P. Kellnhofer: E-mail: pkellnho@mpi-inf.mpg.de

- **Color.** Color of left and right eye image is modified before presentation and color filters in glasses then separate two images.
- **Polarization.** Polarizer layer on the display changes polarity of emitted light for the left and right eye, and polarization filter in glasses then isolates the proper signal for each eye.
- **Parallax barrier and lenslet arrays.** Opaque barriers or tiny lens on the display ensures that individual pixels are only visible to one or the other eye.
- **Time-sequential presentation.** Left and right eye images are presented sequentially and glasses with active shutter are used to block inactive eye. Passive polarized glasses can alternatively be used to bring the active element of polarizer to the displaying device itself.

We investigated how these properties interact with each other in real display devices and how the perception of stereo 3D is altered. We conclude with recommendations for stereo 3D content optimization for some specific display technologies. We suggest the following three improvements:

- Compensation of false motion in depth for anaglyph display (Sec. 2.2).
- Compensation of false motion in depth for time-sequential displays (Sec. 3.2).
- Improved rasterization in the presence of disparity manipulation (Sec. 4).

We believe that such problems are yet researched by the display and computer graphic community, and addressing them will improve the viewing experience for dynamic 3D content at only a small computational cost and implementation effort.

2. TIME-CORRECTION FOR ANAGLYPH DISPLAYS

In this section, we describe an approach for compensation of false motion-in-depth for the anaglyph glasses.

2.1 Anaglyph display

The anaglyph glasses are a popular and affordable technology suitable for occasional 3D content viewing and for quick content presentation. It uses simultaneous presentation of images to both eyes where separation is provided by color filters. Frequency characteristics of these filters differ between individual technologies and it affects mainly the ability to reproduce original chromacity and the amount of crosstalk between eyes.¹ The most common type uses red filter for the left eye and cyan filter for the right eye. The big advantage over other technologies is that stereo images can be printed and thus is not limited to active, electronic display devices.

Left and right eye filters have not only different dominant color but also level of transparency across the visible light spectrum.¹ This adds up with non-uniform sensitivity of human visual system to different wavelengths² and makes each eye perceive different level of illumination. In case of red-cyan glasses the left eye’s red image is perceived darker than the right eye’s cyan one. In summary, one image of the stereo pair appears brighter than the other.

2.2 The Pulfrich effect

A darker signal takes the human visual system longer to process which causes the so-called “Pulfrich” effect. Consequently, a pair of stereo images with different brightness is therefore apparently shifted in time relatively to each other as described by Howard.³ If the user watches a moving object, the “bright eye” sees the object in position at time t while the “dark eye” sees it in a delayed position at time $t - \delta T$. This creates image disparity that is interpreted as an additional depth offset (Fig. 1a). If the motion speed is not constant the resulting disparity changes over time, introducing a false motion-in-depth cue. This principle has been used as a cheap and simple way of 2D-to-3D movie conversion.⁴

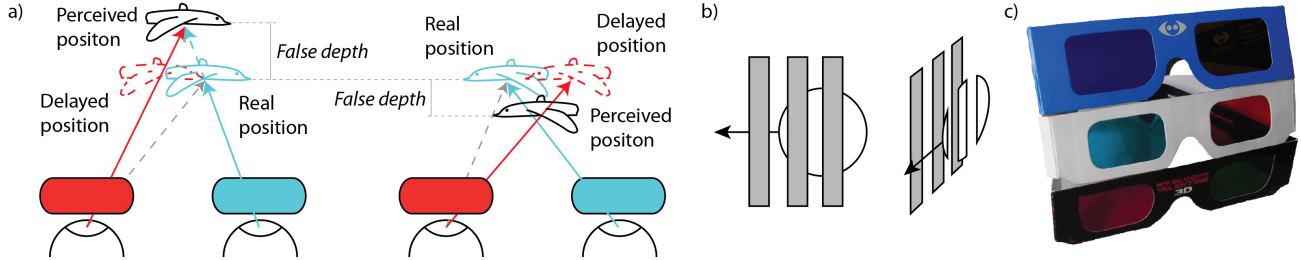


Figure 1: *a)* Pulfrich effect on a stereoscopically viewed horizontally moving bird. Red-cyan anaglyph colors are used to code individual eye images. The right eye image represents the real image-space position of the bird (cyan), the left eye image is perceived delayed (red). This results in the final percept being shifted in depth (black). *b)* Disparity-occlusion clue conflict. *c)* Anaglyph glasses used in our experiment.

It, however, does not provide any control over the amount or sign of generated depth and therefore it is likely to conflict with other cues, such as occlusion (Fig. 1b). If there is an object moving from right to left behind a fence, it will appear to be shifted toward the viewer wearing darker filter on his left eye, e.g., when using cyan-red glasses. Consequently, the disparity cue will predict the object’s depth to be in front of the fence violating the occlusion depth cue.⁵ The effect is not symmetric as the opposite motion will generate opposite depth shift. Horizontal motion is, however, just one special case of general motion. Vertical motion will introduce vertical disparities which might reduce viewing comfort or even prevent fusion.⁶ A typical example for this motion is falling rain.

2.3 Compensating for the Pulfrich effect

We measured attenuation for several types of anaglyph glasses (Fig. 1c) using luminance meter Minolta LS-100 (See Tab. 1). For most widely spread red-cyan glasses we got relative attenuation of the left eye as 0.55 log units. This is enough to create a delay of 5 to 10 ms according to Howard, Section 23.1.⁷ Even ColorCode amber-blue glasses which are well-known for good reproduction of chromacity show similar magnitude of left- versus right-eye attenuation, but with opposite sign. Therefore producing the opposite shift in depth. Some glasses do not show large differences between eyes such as the green-magenta combination in our tests.

Type	Luminance [cd/m^2]		Attenuation [log]			
	No filter	Left	Right	N/L	N/R	R/L
Red-cyan	218.10	21.82	76.64	1.00	0.45	0.54
Green-magenta	218.10	34.86	28.32	0.80	0.89	-0.09
Amber-blue	218.10	15.35	1.567	1.15	2.14	-1.0

Table 1: Our measurements of anaglyph glasses filter attenuations in log units. Common red-cyan glasses, Trioscopics green-magenta glasses and ColorCode 3-D amber-blue glasses.

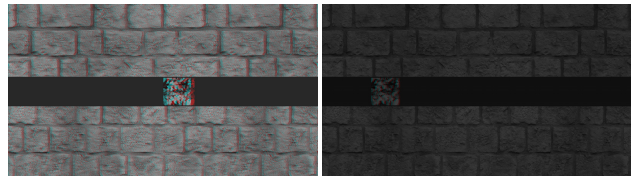


Figure 2: Experimental measurement of Pulfrich effect on red-cyan anaglyph glasses. User adjusts the delay of the right eye with lighter filter to negate introduced motion in depth.

We measured the impact of the Pulfrich effect on the stereo perception in a perceptual experiment. We asked 5 subjects with tested stereo vision to watch a textured square target stimulus moving horizontally in a sinusoidal motion (Fig. 2) on a Samsung SyncMaster 2233RZ 120 Hz LCD with the luminance range from 20 cd/m^2 to 200 cd/m^2 under normal office lighting conditions. The motion pattern covered viewing angle of 33 degrees. The stimulus depth was constant and in the plane of surrounding texture. Due to the Pulfrich effect, the target appeared to move in depth while moving left and right. Participants were instructed to adjusted the setup by pressing the “left” or “right” key, until the stimulus remains stable in depth. We experimented with 4 different stimulus motion frequencies varying from 0.3 to 1.2 Hz, 3 different disparity magnitudes (crossed, zero and uncrossed) and 2 different brightness levels within the range of the display. The theoretical model did not predict dependency on any of these attributes.

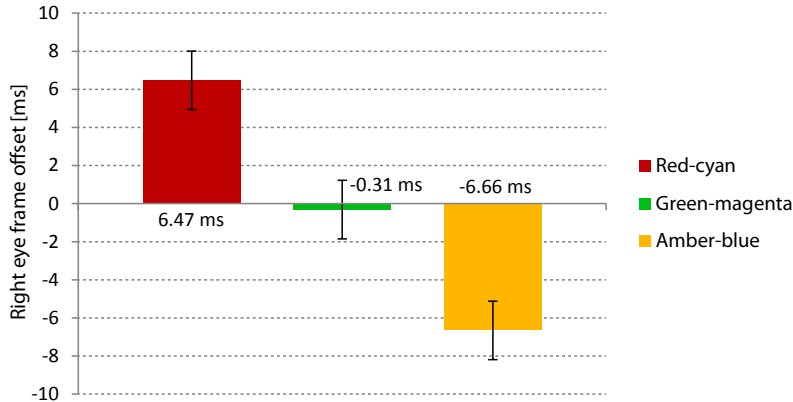


Figure 3: Time offsets in milliseconds introduced as delays for right eye frames to compensate for Pulfrich effect. Error bars denote confidence intervals for $p = 0.01$ according to Tukey HDS test.

The graph in Fig. 3 shows average time offsets for the right eye necessary to compensate each glass type. We have run one-way ANOVA with Tukey HDS post-hoc test and found that all pairs of mean values are different with $F(2, 102) = 81.06$, $p < 0.01$. The mean offset for red-cyan glasses 6.5 ms lies in the interval predicted by the physical measurement of filter densities.

As predicted, there was no significant difference detected among other experiment attributes. That can be due to insufficient extent of the study. Larger study would be required to prove that differences of means are negligible.

Our results confirm that the magnitude of the Pulfrich effect strongly depends on the type of glasses technology. Our supplemental material provides a simple HTML5 applet to measure your own equipment and visual system in a web browser (available on <http://resources.mpi-inf.mpg.de/TemporalStereo#spie2014>). We recommend applications using stereo to allow for adjustment of a delay of one eye. The delay can be either provided for most common anaglyph technologies or tuned by user using interface similar to our applet. Similar procedure is common for the contrast adjustment in graphical applications, e. g., computer games. Our measurements provide useful compensation values for some examples of a 3D equipment.

3. PROTOCOL CORRECTION FOR EYE PURSUIT MOTION

In this section, we devise an approach to switch between alternative stereo capture and presentation protocols, based on a novel content-dependent prediction of eye pursuit motion probability.

3.1 Protocols

Capture or *presentation protocols* describe how image data for the left and right eye are captured or presented in time. The two basic types are *simultaneous* and *alternating* protocols. With simultaneous protocols both images either describe the world at the same time in case of capture protocol or are displayed at the same time in case of presentation protocol. With alternating protocol images are either captured or displayed at alternating time sequences. We follow the notation on Hoffman et al.⁸ and denote the combination using abbreviations of capture protocol and presentation protocol in that order, e. g., SIM/ALT for “simultaneous capture with alternating presentation”.

The multiplexing technology usually determines the type of presentation protocol of choice. For time-sequential displays the alternating presentation is the only option. If simultaneous capture protocol was chosen it would produce conflict of time as the image presented in the second eye would be delayed with respect to its capture time. The resulting effect would introduce additional disparity, hence false motion-in-depth similarly as in the Pulfrich effect. Additionally, it can give raise to vertical disparity for vertical motion, reducing viewing comfort. Therefore matching the capture and presentation protocols, i. e., using SIM/SIM or ALT/ALT, is recommend.

However, as it was shown by Hoffman et al.,⁸ the situation changes, when frame repetition is required. Such repetition is employed commonly in the cinema where multiple flashes of identical frames are presented to reduce flickering for movies with relatively low frame rates. It was pointed out that especially repeating frames with an alternating presentation protocol might introduce a false motion-in-depth sensation. The proposed model assumed that the time offset capture that matches to ALT/ALT for analogical single flash presentation should minimize depth perception issues. We will further use just ALT/ALT for such protocol with frame repetition. It was, however, observed that this choice is appropriate only for slow angular speeds of video content motion and SIM/ALT is a better choice for speeds above some threshold. This observation is explained by the *temporal disparity-gradient limit* $|\Delta\delta/\Delta t|$, a change of disparity over time, after which the HVS' disparity estimation fails. Hoffman et al.⁸ conclude with the recommendation to select between the two protocols discussed based on the probability of frame repetition.

Protocol choice might be further complicated if the capture frame-rate is changing over time which is a typical case in real time rendered content. Time-sequential presentation based stereo 3D technologies using active shutter glasses typically operates at 120 Hz to avoid flickering. Therefore rendering at 60 Hz for each eye is necessary in order to avoid frame repetition. That is too much for most of the current middle-range consumer HW and most up-to-date games. We can save some performance by generating every other frame using warping of the previous frame. Didyk et al.⁹ used blurring to hide the resulting warping artifacts. They argued that blurring of in one eye does not reduce the overall sharpness. However, warping might not always be sufficient cure for the performance problem, as one would start to see quality decrease if too many interpolated frames were inserted. In this case frame repetition is required and used. It means that the preferred capture protocol might change over time according to conclusions of Hoffman et al.⁸

3.2 Time-sequential displays and presentation protocols

A common way how to display 3D content on conventional displays is based on time-sequential presentation of the left and right eye image. Active shutter glasses are used to occlude the eye that is currently not required. The resulting ALT presentation protocol raises depth perception problems if frame repetition is used as discussed in the previous section.

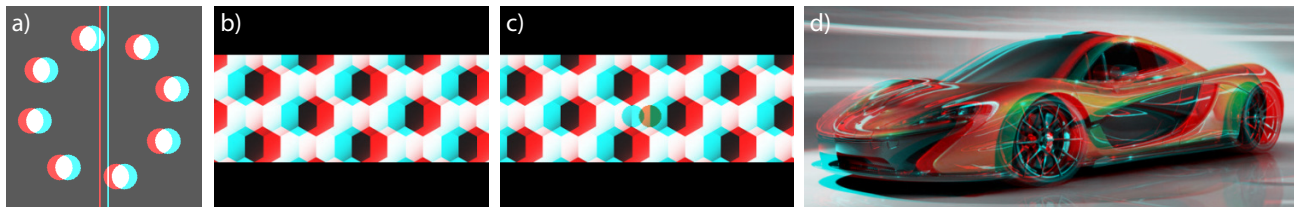


Figure 4: Representative examples of stimuli used in our experiment. *a)* Rotating stimuli used to reconstruct experiment of Hoffman et al.⁸ *b)* Periodic stimuli where eye pursuit motion fails at higher angular velocities. *c)* The same stimuli with an additional feature that improves eye pursuit motion. *d)* Stimuli with an easily trackable picture of a car.

We reproduced the fourth experiment of Hoffman et al.⁸ where depth stability was measured with either the ALT or SIM capture protocol for the ALT presentation protocol with frame repetition. We used green-magenta anaglyph glasses which were shown to have minimal Pulfrich effect to simulate the protocol on our 120 Hz LCD display. Our observations led to the same conclusion when applied to a rotating circle stimulus (Fig. 4a) as described. However, different conclusions have to be drawn from observations made when using more complex, 3D-rendered stimuli as found in interactive applications such as computer games. For example, we introduced periodical horizontal motion in the scene and used the ALT presentation of the left and right anaglyph image to simulate time-sequential display. We always compared the measured multi-flash protocol with a ground-truth SIM capture and presentation where no depth distortions are expected.⁸ We then studied the relative motion-in-depth between the reference and multi-flash stimuli shown on the same display.

For slow motion our observations matched those of Hoffman et al.⁸ When the motion speed increased, the effect started to vary between images and some became unstable in depth for the ALT capture protocol. For

other stimuli however, perception remained stable relative to a reference image and followed the model for slow speeds even at high speeds. We found, that the model for slow speeds is not valid for images with highly periodic texture patterns without significant features such as mosaics or rocks (Fig. 4b). For photographs of cars or people (Fig. 4d), the slow speed model was followed even at high speeds. We suspect that the reason for this difference is the inability to correctly pursue moving objects with smooth pursuit eye motion, when no visually significant and unique features can be distinguished on the object. Therefore, the measurement done for periodic textures were actually done without eye pursuit motion even though subjects were instructed to pursue the moving objects.

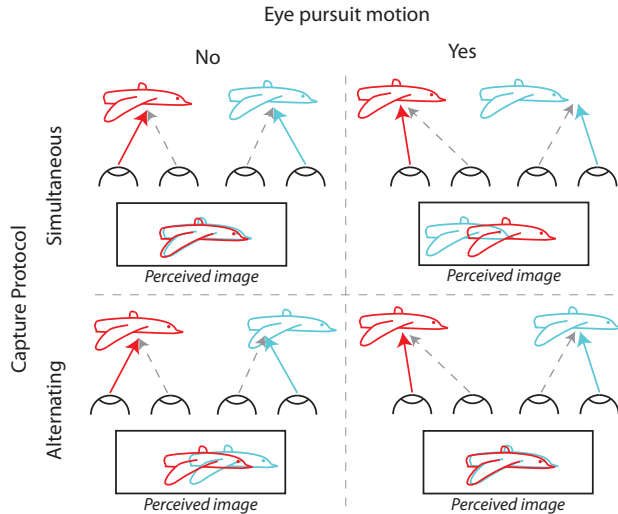


Figure 5: Perception of flying bird with zero disparity presented using ALT protocol without frame repetition. Red and cyan colors denote the eye active in the frame.

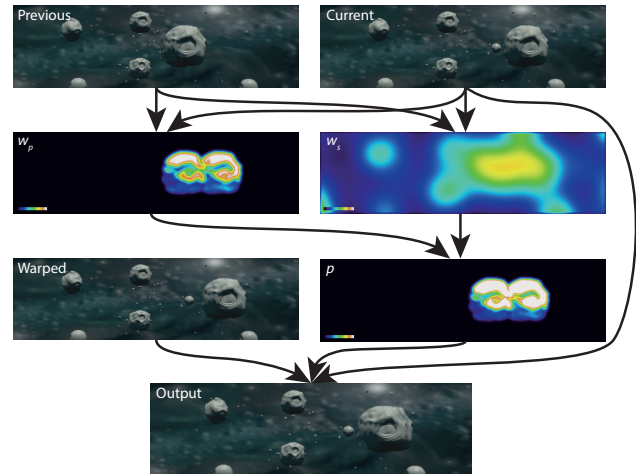


Figure 6: Diagram of our protocol for frame repeating ALT presentation. Previous and current left frames as inputs to our framework. Optimized left frame as output.

Fig. 5 explains the geometry of different presentation protocols and different eye pursuit motion conditions on perceived disparity. It shows that when assuming perfect estimation of smooth eye pursuit the temporarily accumulated disparity error is zero for alternating capture protocol with eye pursuit motion, or simultaneous capture protocol without eye pursuit motion. There is a false depth from disparity created with both other combinations.

To verify this assumption, another experiment was done using the same configuration and a periodic monochromatic image. We inserted a single unique feature into the pattern, a green colored circle, so that it became easier to pursue (Fig. 4c). We then found that the observation of apparent motion in depth became consistent with that for non-periodic images. We also did similar experiment with the original rotational setup. Here, we found difficulty to evaluate perceived motion in depth. We observed, that even though given point is properly pursued on its circular trajectory, it seems to move back and forth in depth with respect to the middle bar. However, the reason for this effect is found in the middle bar itself: When we instructed the participants to pursue the rotation motion, users lost their track with the static vertical bar. So it was the bar that exhibited screen space motion relative to the eye and therefore it moved in depth. As viewers were attracted to the moving circle, they did not observe its own motion and only saw different relative position at the transition phase of circle above the bar. This invoked the illusion of motion in depth for the circle.

We therefore generalize the recommendation given by Hoffman et al.⁸ and conclude that the rendering approach should be chosen not only based on capture rate but also on the content and expected attention of the viewer. We suggest a spatially adaptive approach based on a novel eye pursuit motion probability map. The offset-compensated capture protocol is advised by default as it works well when no frame repetition is involved⁸ and converges to the simultaneous capture protocol when motion speed goes to zero. Simultaneous

rendering should only be chosen for moving regions that are not pursued by viewer. Simultaneous rendering is simulated using motion-flow based warping of the frame to allow for local transformations. The eye pursuit motion probability map is driven by local blending between both protocols maintaining spatio-temporal smoothness of the original sequence (Fig. 6). The proposed approach combines benefits of both protocols for different types of content. It is conservative as in the worst case, it produces results identical to the less suitable of them and not worse.

To justify the blending between two protocols we estimate the disparity means for each of two extreme cases. Fig. 2 shows how additional disparity in relative units denoted as dx is cumulated through six screen states of a triple-flash ALT/ALT protocol with eye pursuit motion in the way as was depicted for single flash protocols in Fig. 5. Each eye receives image composed from several images at different time. As a result of predicted motion and smooth eye pursuit, retinal projections of these images are shifted by dx_i . The final image can be considered as a low-pass filtered with mean value approximated by average of dx_i . Then the disparity is difference of means for left and right eye. In the case described in Fig. 2 we get a mean as 2 units for both eyes and therefore no additional disparity. Values for all combinations are summed in Tab. 7. The difference sign denotes with the false additional disparity, which is observed in our experiments as a shift in depth.

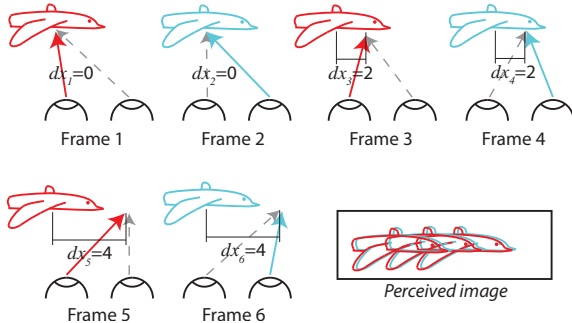


Figure 7: Cumulative disparity error during smooth pursuit of a flying target displayed using the ALT/ALT protocol with triple frame repetition. Red and cyan colors denote the left and right eye active in the respective frame.

Cap. protocol	Eye purs.	Sequence		Mean		Disp.
		Left	Right	Left	Right	
Simultaneous	No	0, 2, 4	0, 2, 4	2	2	0
Simultaneous	Yes	0, 2, 4	1, 3, 5	2	3	-1
Alternating	No	0, 2, 4	1, 3, 5	2	2	0
Alternating	Yes	0, 2, 4	0, 2, 4	2	3	-1

Table 2: The mean values of spatial offsets between the retina projection and expected motion trajectory and resulting false disparity for various capture protocols and eye pursuit motion assumptions. The relative units are multiples of product of spatial speed and frame duration.

For the method description to follow, we assume without loss of generality to display the left frame first. First the right frame is rendered at the simulation time and left frame is rendered at a time decreased by the delay of the right frame’s first presentation (e. g., 8.3 ms for 120 Hz display). That matches the ALT/ALT protocol with an offset described by Hoffman et al.⁸ The left frame is displayed without correction.

We compute an eye pursuit motion probability map for the right eye using a combination of dynamic saliency and similarity between frames. The saliency tells us if the user is motivated to pursuit given part of the image while the similarity tells us if user is likely to determine the motion flow and to perform the pursuit properly. We use the *Phase Spectrum of Quaternion Fourier Transformation* method¹⁰ to find a spatio-temporal saliency map $w_s(\mathbf{x}_i)$. To detect the similarity between frames we compare matching samples of current and previous frame $f_t(\mathbf{x}_i)$ and $f_{t-1}(\mathbf{x}_i)$ and we get the periodicity weight $w_p(\mathbf{x}_i)$ as:

$$w_p(\mathbf{x}_i) = \min(\max(1 - 2|f_t(\mathbf{x}_i) - f_{t-1}(\mathbf{x}_i)|, 0), 1) \quad (1)$$

Finally, a Gaussian low-pass filter of radius 30 arcmin (typically 16 pixels) is applied to achieve spatial smoothness. The final eye pursuit motion probability map then is

$$p(\mathbf{x}_i) = (1 - w_p(\mathbf{x}_i)) \cdot w_s(\mathbf{x}_i). \quad (2)$$

We produce the final image by local blending between the original frame and the warped frame. We may not want to introduce any compensation and possible artifacts into nearly static scenes where the default

rendering protocol is sufficient. Therefore we take the motion speed into account. The largest speed where the slow-speed-model holds was measured by Hoffman et al.⁸ as

$$s = \frac{C}{2f} \quad (3)$$

where f is the number of frame repetitions and C is the temporal disparity-gradient limit of $|\Delta\delta/\Delta t|$ approximated as 950 arcmin/sec of change of viewing angle at the eye. Therefore we obtain limit speed s as 237.5 or 158.3 arcmin/sec for double or triple frame repetition. This way speed coefficient $w_c(\mathbf{x}_i)$ is derived:

$$w_c(\mathbf{x}_i) = \max\left(\frac{|\Delta\delta(\mathbf{x}_i)|}{s\Delta t}, 1\right) \quad (4)$$

where Δt is time difference between consequent frames and $\Delta\delta(\mathbf{x}_i)$ is the angular difference between the position of pixel \mathbf{x}_i in the previous and the current frame, approximated for a perpendicular viewed distant display as

$$\Delta\delta(\mathbf{x}_i) = \tan^{-1}\left(\frac{\|\mathbf{x}_i - \mathbf{A}_f\mathbf{x}_i\| \cdot P}{d}\right) \quad (5)$$

where P is matrix size of pixel and d the screen distance. The permutation matrix \mathbf{A}_f describes the local motion flow image f . Then we can derive the final blending weight $w(\mathbf{x}_i)$ as

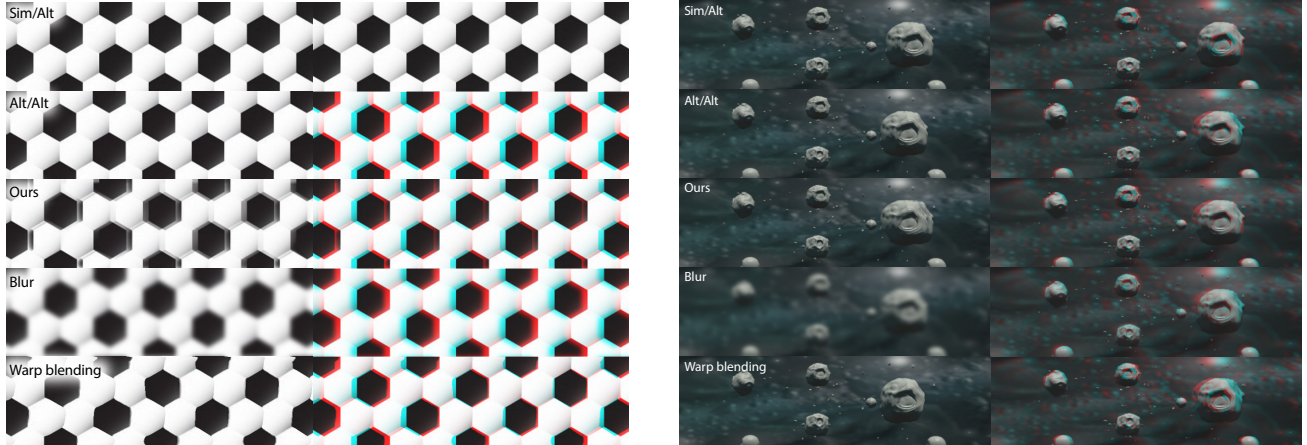
$$w(\mathbf{x}_i) = \text{clamp}(2 \cdot w_c(\mathbf{x}_i) \cdot p(\mathbf{x}_i), 0, 1) \quad (6)$$

We warp the frame f using warping map \mathbf{B}_f based on the motion flow \mathbf{A}_f to approximate the frame that would be rendered by the SIM capture protocol:

$$\mathbf{B}_f = \frac{1}{2f} \cdot \mathbf{A}_f \quad (7)$$

The final frame is then generated by blending between the alternating and simulated simultaneous capture protocols:

$$\hat{f}_t(\mathbf{x}_i) = (1 - w(\mathbf{x}_i))f_t(\mathbf{x}_i) + w(\mathbf{x}_i)f_t(\mathbf{B}_f\mathbf{x}_i) \quad (8)$$



(a) Simple periodic texture.

(b) Rendered 3D scene.

Figure 8: Left column shows frames processed by our method. Right column shows overlaid left and one consequent right frame which simulates viewing without eye pursuit motion. The image was presented in the screen plane without disparity.

We compared this proposed approach with a simple Gaussian blurring of the frame and with blending of in-between frame warping map \mathbf{B}_f instead of blending of warped images. Fig. 8 shows application examples

for simple periodic texture and rendered 3D scene. Simple blurring with symmetrical kernel does not change the mean value of disparity distribution and therefore was ineffective in improving depth stability. Blending of warping maps reproduced depth stability comparable to the proposed method but exhibited artifacts perceived as deformations. The proposed method also leads to visual artifacts which are perceived as double edges when observed statically, however such edges are blurred with previous frame repetitions in animation sequence which results in overall smoother appearance than the ALT/ALT protocol.

4. RENDERING FOR DISPARITY MANIPULATION

The 3D stereo content both real and synthetic is often manipulated in order to achieve better viewing experience on nowadays displays. This is only an approximation of light field we perceive in real life. One typical limitation is the amount of disparity that conventional displays can show due to the conflict of vergence and accommodation cues.¹¹

In order to cover scenes with wider depth extent the disparity is often manipulated. Simplest way is a global disparity compression,¹² which might, however, make important parts of the scene look flat and lead to the cardboard effect. Therefore, analogically to approaches used in HDR tone mapping, local operators are often applied to better distribute the depth budget.^{13,14} Another reason for local approach might be perceptual scaling of disparity in combination with luminance signal.¹⁵

Applied on static content, such remapping of disparity might heavily compress depth zones without important content. In dynamic scenes a moving object can easily enter or leave such a zone, which leads to temporal artifacts in a disparity change over time that are visible in Fig. 9 as abrupt temporal changes in the disparity field.

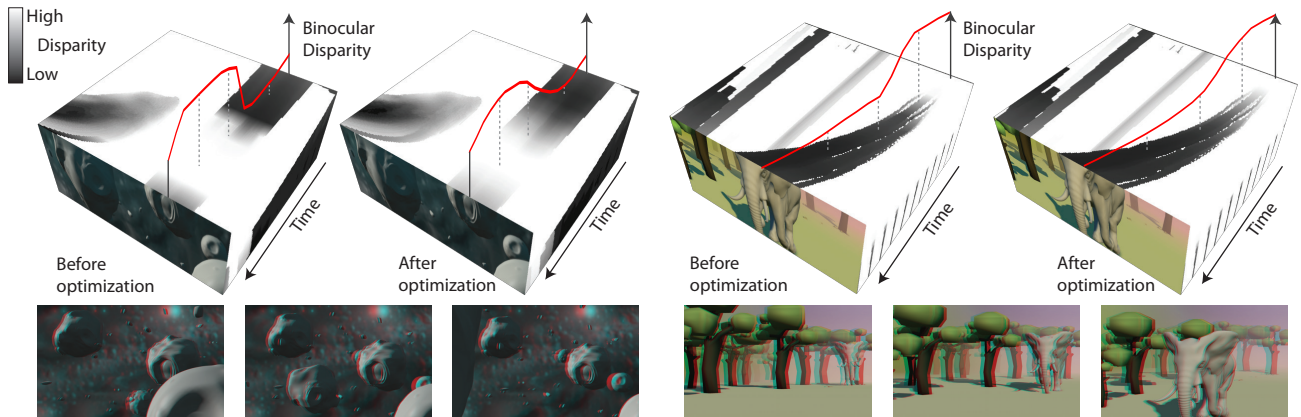



Figure 9: Two rendered scenes with complex deformation and camera motion used as stimuli for the user study.¹⁶ *Top*: Slices through manipulated spatio-temporal disparity cubes. Time of animation proceeds from the back to front side of the cube. Plot shows values of disparity in time for single object in the scene as it moves in 3D space.  *Bottom*: Three frames from each scene.

The cooperation of binocular cue of disparity change and monocular cues is important for accurate estimation of motion in depth.¹⁷ The disparity change cue is important for time to collision estimation at distances up to 75m.¹⁸ Its significance grows at higher motion speeds.¹⁹

The experiment with football like game simulation showed that aggressive disparity manipulation does indeed decrease the precision of trajectory estimation in 3D space.¹⁶ Optimization scheme to restore the motion in depth perception after an arbitrary disparity manipulation was suggested as a solution.

An energy function is built to describe similarity of the final disparity field with the user provided manipulated input while at the same time minimize the difference of the motion in depth compared to the original sequence without any manipulation. To capture general motion characteristics an acceleration of disparity is used to preserve high frequencies in the motion. This is an advantage over simple temporal smoothing applied in some other frameworks.^{13,14}

Both spatial and temporal undersampling is exploited to achieve a real time performance. The original high quality signal is then used as a guide in an upsampling reconstruction enabling recovery of high frequency details.

The method is applicable to any content where the disparity signal is available. It is especially effective in case of rendered images where the disparity can be retrieved directly from a depth map at no additional cost. The inverted problem of applying the modified disparity field back to the image content is also possible.

Conventional general image space warping techniques often lead to undesired disocclusion artifacts whenever the background is revealed by a motion parallax of a manipulated front object. This can be in some cases masked by inpainting techniques.

However, it is suggested, that for a specific case of rendered content, a better visual quality can be achieved by warping of the underlying 3D geometry before its stereoscopic projection to the 2D image space.¹⁶ The projection of each vertex using a central camera is used to find its location in the disparity field. If an occlusion is detected using an attached depth map a neighbourhood in the disparity field is searched to find a projection similar in both the 2D space and depth. A global curve reconstructed from the disparity field is used in the case of a failure.

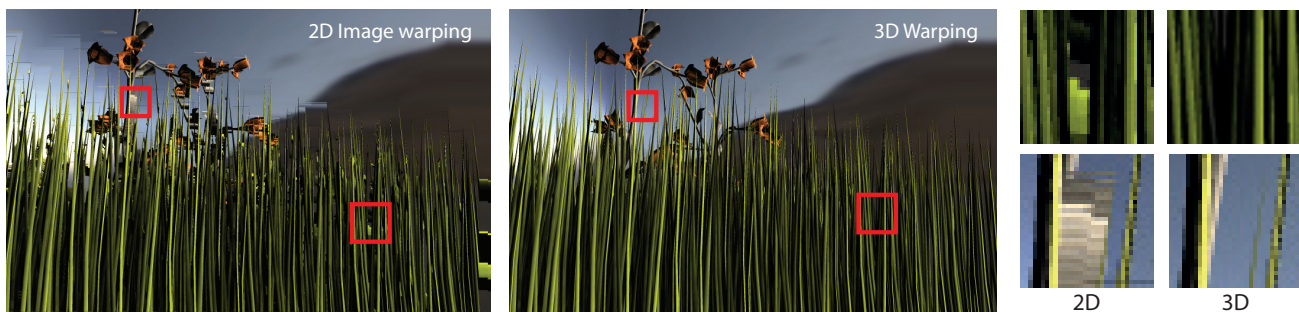


Figure 10: A comparison of a conventional 2D image space warping and the proposed 3D warping.¹⁶

Results of this method are visually superior to the image space warping, especially in challenging scenes with a high frequency content in both luminance and disparity, where disocclusions are frequent and hard to mask Fig. 10.

A user study comparing short computer generated sequences with object motion and compressed disparity (Fig. 9) has shown that the proposed motion optimization leads to preferred appearance and better precision in a task performance that requires depth judgements.¹⁶

5. DISCUSSION AND CONCLUSION

This paper focused on three specific issues related to generating and presenting a stereoscopic 3D content. It was shown that two of widely used stereoscopic display technologies are the subject of potential disparity distortion when a motion is introduced in the content.

First, the Pulfrich effect was analyzed for anaglyph glasses and then experimentally measured with several pieces of a consumer available 3D eye-wear. Our results lead to a suggestion of a temporal compensation for an anaglyph presented content.

Second, the left and right eye image capture and presentation protocols were discussed in the context of time sequential display technologies. Conditions leading to content dependent distortions of the disparity of moving objects were theoretically predicted and practically observed. A spatially adaptive saliency based approach was then proposed to combine advantages of two existing capture protocols and to minimize the disparity distortion for a dynamic 3D content on displays with a sequential presentation.

Finally, we also discussed difficulties of an application of stereo manipulation techniques in a production phase of temporarily changing scenes. We showed a technique minimizing the negative effect of a per frame manipulation of the disparity on the perception of the motion in depth. Such an optimized disparity field can then be applied to a high quality synthetic 3D rendering using a novel warping approach preventing common warping artifacts.

Both the topic of the motion in depth perception and the perception of depth in conditions of the motion have received a relatively less attention than the static 3D imaging in last years. A more comprehensive study and a wider choice of technologies should be included to create a complex recommendation for the 3D production and presentation. In future, we would like to focus on the interplay of disparity with other depth cues, such as texture, luminance, size or motion parallax itself. The stereo image capture processing is another challenging step.

REFERENCES

- [1] Woods, A. J. and Rourke, T., “Ghosting in anaglyphic stereoscopic images,” in [*Proc. SPIE 5291*], (2004).
- [2] Thornton, W. A., “Spectral sensitivities of the normal human visual system, color-matching functions and their principles, and how and why the two sets should coincide,” *Color Research & Application* **24**(2), 139–156 (1999).
- [3] Lit, A., “The magnitude of the pulfrich stereophenomenon as a function of binocular differences of intensity at various levels of illumination,” *The American Journal of Psychology* **62**(2), pp. 159–181 (1949).
- [4] Beard, T. D., “Low differential 3-d viewer glasses and method.” Patent (05 1991). EP 0325019 B1.
- [5] Mendiburu, B., [*3D movie making: stereoscopic digital cinema from script to screen*], Focal Press (2009).
- [6] Woods, A., Docherty, T., and Koch, R., “Image distortions in stereoscopic video systems,” in [*Stereoscopic Displays and Applications*], (1993).
- [7] Howard, I. and Rogers, B., [*Perceiving in Depth, Volume 2: Stereoscopic Vision*], Oxford Psychology Series, OUP USA (2012).
- [8] Hoffman, D. M., Karasev, V. I., and Banks, M. S., “Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth,” *J Soc Inf Disp* **19**(3), 271–297 (2011).
- [9] Didyk, P., Eisemann, E., Ritschel, T., Myszkowski, K., and Seidel, H.-P., “Perceptually-motivated real-time temporal upsampling of 3D content for high-refresh-rate displays,” *Computer Graphics Forum (Proceedings Eurographics 2010, Norrköping, Sweden)* **29**(2), 713–722 (2010).
- [10] Guo, C., Ma, Q., and Zhang, L., “Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform,” in [*2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA*], IEEE Computer Society (2008).
- [11] Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I., “Visual discomfort and visual fatigue of stereoscopic displays: A review,” *J Imag. Sci. and Tech.* **53**(3), 1–12 (2009).
- [12] Oskam, T., Hornung, A., Bowles, H., Mitchell, K., and Gross, M., “OSCAM-optimized stereoscopic camera control for interactive 3D,” *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* **30**(6), 189:1–189:8 (2011).
- [13] Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., and Gross, M., “Nonlinear disparity mapping for stereoscopic 3D,” *ACM Trans. Graph. (Proc. SIGGRAPH)* **29**(4), 75 (2010).
- [14] Yan, T., Lau, R., Xu, Y., and Huang, L., “Depth mapping for stereoscopic videos,” *International Journal of Computer Vision* **102**, 293–307 (2013).
- [15] Didyk, P., Ritschel, T., Eisemann, E., Myszkowski, K., and Seidel, H., “A perceptual model for disparity,” *ACM Trans. Graph. (Proc. SIGGRAPH)* **30**(4), 96:1–96:10 (2011).
- [16] Kellnhofer, P., Ritschel, T., Myszkowski, K., and Seidel, H.-P., “Optimizing Disparity for Motion in Depth,” *Computer Graphics Forum* **32**(4), 143–152 (2013).
- [17] Gray, R. and Regan, D., “Accuracy of estimating time to collision using binocular and monocular information,” *Vis. Res.* **38**(4), 499–512 (1998).
- [18] Cavallo, V. and Laurent, M., “Visual information and skill level in time-to-collision estimation,” *Perception* **17**(5), 623–32 (1988).
- [19] Regan, D. and Beverley, K., “Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage,” *Vis. Res.* **19**(12), 1331–1342 (1979).