Dynamic Horizontal Image Translation in Stereo 3D

von der Fakultät

für Elektrotechnik und Informationstechnik

der Technischen Universität Dortmund

genehmigte

Dissertation

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

von

Stefan Eickelberg

Dortmund, 2017

Tag der mündlichen Prüfung: 06.04.2017

Hauptreferent: Prof. Dr.-Ing. Rüdiger Kays

Korreferent: Prof. Dr.-Ing. Ulrich Reimers

Dortmunder Beiträge zur Kommunikationstechnik

Band 15

Stefan Eickelberg

Dynamic Horizontal Image Translation in Stereo 3D

D 290 (Diss. Technische Universität Dortmund)

Shaker Verlag
Aachen 2017

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Zugl.: Dortmund, Technische Univ., Diss., 2017

Printed in Germany.

ISBN 978-3-8440-5361-6
ISSN 1863-9054

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen
Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9
Internet: www.shaker.de • E-Mail: info@shaker.de

Vorwort

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Kommunikationstechnik der Technischen Universität Dortmund im Zeitraum von 2011 bis 2016.

Meine Faszination für Stereo 3D und die Wissenschaft war schon in sehr frühen Jahren gegeben. Ich möchte mich daher an dieser Stelle von ganzem Herzen bei Lehrstuhlinhaber Herrn Prof. Dr.-Ing. Rüdiger Kays dafür bedanken, dass er mir in solchem Umfang ermöglichte, meine Kindheitsträume zu verwirklichen. Seine Betreuung meiner Doktorarbeit war stets hilfreich und motivierend.

Herrn Prof. Dr.-Ing. Ulrich Reimers vom Institut für Nachrichtentechnik der Technischen Universität Braunschweig danke ich für das Interesse an meiner Arbeit, seine wertvollen Kommentare und für die Übernahme des Korreferates.

Ich möchte ebenfalls meinen ehemaligen Arbeitskollegen danken, die im Laufe der Jahre zu Freunden geworden sind und die Zeit am Lehrstuhl zu einem nicht nur fachlich bedeutsamen Lebensabschnitt für mich machten. Dabei bin ich insbesondere meinem langjährigen Schreibtischnachbarn Herrn Dr.-Ing. Matthias Brüggemann für die zahlreichen interessanten Diskussionen und unterhaltsamen Momente zu Dank verpflichtet. Ich möchte mich ferner bei allen Studierenden, deren Abschlussarbeiten ich betreut habe, und allen Hilfskräften bedanken. Auch ihr Einfluss trug selbstverständlich zum Gelingen dieser Arbeit bei.

Zu guter Letzt möchte ich mich bei meiner Mutter und meinem Vater sowie allen anderen Familienangehörigen und Freunden dafür bedanken, dass sie mich stets unterstützt haben.

Dortmund, Juni 2017

Stefan Eickelberg

Contents

Kurzfassung

Im Bereich Stereo 3D (S3D) bezeichnet „Dynamic Horizontal Image Translation (DHIT)“ das Prinzip, die S3D-Ansichten einer Szene horizontal in entgegengesetzte Richtungen zu verschieben, wodurch die dargestellte Szene in der Tiefe verschoben wird. Dies wird vor allem im Kontext von „Active Depth Cuts“ eingesetzt. Hier werden die S3D-Ansichten vor und nach einem Szenenschnitt so verschoben, dass es nicht zu starken, störenden Tiefensprüngen kommt.

Die menschliche Wahrnehmung der DHIT wurde experimentell untersucht. Eine der wichtigsten Erkenntnisse war, dass es starke individuelle Unterschiede in der Empfindlichkeit gegenüber der DHIT gibt. Daher wird empfohlen die Verschiebungsgeschwindigkeit einer S3D-Ansicht nicht höher als 0,10 °/s bis 0,12 °/s zu wählen, sodass Zuschauerinnen und Zuschauer nicht von der DHIT gestört werden.

Bei der DHIT kommt es zu einer Verzerrung der dargestellten Szenentiefe. Dies wird bei dem vorgeschlagenen Ansatz „Distortion-Free Dynamic Horizontal Image Translation (DHIT+)“ kompensiert, indem der Abstand zwischen den S3D-Kameras durch Verfahren der Ansichtensynthese angepasst wird. Dieser Ansatz zeigte sich signifikant weniger störend im Vergleich zur DHIT. Die Ansichten konnten ohne Wahrnehmungsbeeinträchtigung etwa 50 % schneller verschoben werden.

Ein weiteres vorgeschlagenes Verfahren ist „Gaze Adaptive Convergence in Stereo 3D Applications (GACS3D)“. Unter Verwendung eines Eyetrackers wird die Disparität des geschätzten Blickpunkts langsam über die DHIT reduziert. Dies soll die Ermüdung des visuellen Systems mindern, da die Diskrepanz zwischen Akkommodation und Konvergenz reduziert wird. In einem Experiment mit emuliertem Eye-Tracking war GACS3D signifikant weniger störend als eine normale DHIT. Im Vergleich zwischen dem kompletten GACS3D-Prototypen und einer Bildsequenz ohne jegliche Verschiebungen konnte jedoch kein signifikanter Effekt auf den subjektiven Betrachterkomfort registriert werden. Eine Langzeituntersuchung der Ermüdung des visuellen Systems ist nötig, was über den Rahmen dieser Dissertation hinausgeht. Da für GACS3D eine hochgenaue Schätzung der Blickpunktdisparität benötigt wird, wurde die „Probabilistic Visual Focus Disparity Estimation“ entwickelt. Bei diesem Ansatz wird die 3D-Szenenstruktur in Echtzeit geschätzt und dazu verwendet, die Schätzung der Blickpunktdisparität deutlich zu verbessern.

Abstract

Dynamic horizontal image translation (DHIT) denotes the act of dynamically shifting the stereo 3D (S3D) views of a scene in opposite directions so that the portrayed scene is moved along the depth axis. This technique is predominantly used in the context of active depth cuts, where the shifting occurs just before and after a shot cut in order to mitigate depth discontinuities that would otherwise induce visual fatigue.

The perception of the DHIT was investigated in an experiment. An important finding was that there are strong individual differences in the sensitivity towards DHIT. It is therefore recommended to keep the shift speed applied to each S3D view in the range of 0.10 °/s to 0.12 °/s so that nobody in the audience gets annoyed by this approach.

When a DHIT is performed, the presented scene depth is distorted, i.e., compressed or stretched. A distortion-free dynamic horizontal image translation (DHIT+) is proposed that mitigates these distortions by adjusting the distance between the S3D cameras through depth-image-based rendering techniques. This approach proved to be significantly less annoying. The views could be shifted about 50 % faster without perceptual side effects.

Another proposed approach is called gaze adaptive convergence in stereo 3D applications (GACS3D). An eye tracker is used to estimate the visual focus whose disparity is then slowly reduced using the DHIT. This is supposed to lessen visual fatigue since the infamous accommodation vergence discrepancy is reduced. GACS3D with emulated eye tracking proved to be significantly less annoying than a regular DHIT. In a comparison between the complete prototype and a static horizontal image translation, no significant effect on subjective visual discomfort could be observed, however. A long-term evaluation of visual fatigue is necessary, which is beyond the scope of this work. In GACS3D, highly accurate visual focus disparity is required. Therefore, the probabilistic visual focus disparity estimation (PVFDE) was developed, which utilizes a real-time estimation of the 3D scene structure to improve the accuracy by orders of magnitude compared to commonly used approaches.

1 Introduction

1.1 Motivation

In a 2010 study with almost 8000 subjects, 30 % of the subjects reported eyestrain and 6 % reported headaches after watching a **stereo 3D (S3D)** movie at a cinema [Ber10]. In a more recent study in 2015 with about 400 subjects, 10 % of the subjects reported headaches or eyestrain after being exposed to the movie “Toy Story” (2009 3D re-release) on an S3D TV [Rea15]. There are many factors that could explain the differing numbers in these reports, but they concur that a significant portion of the population suffers from negative side effects when watching S3D content. The exhaustion of the **human visual system (HVS)** is called visual fatigue, and its reduction is mandatory for the success of S3D, especially in a work environment.

Two of the numerous sources for visual fatigue are the motivational foundation for this work: the **accommodation vergence discrepancy (AVD)** [Lam09b, Shi11, Hof08, War95] and excessive depth movement of objects inducing exhausting vergence movements of the eyes [Kim14, Lam09b, Tod04]. The discrepancy between accommodation and vergence distance is given when the eyes of an observer converge on a stimulus behind or in front of an S3D display, while the imagery, .i.e., the accommodation stimulus is located at the display distance. Small discrepancies between these distances do not induce visual fatigue so that S3D scenes are usually limited to a zone of comfort [Lam09b, Shi11]. This can be achieved by two steps:

1. Adjust the distance between the S3D cameras, i.e., the camera baseline such that the resulting disparity range is small enough to fit inside the zone of comfort.
2. Place the recorded scene in the zone of comfort by applying a **horizontal image translation (HIT)** of the S3D views in opposite directions [Bro11,

Men11]. Thereby, the portrayed scene is shifted along the depth axis.

Regarding the latter source for visual fatigue, a reduction of excessive depth movement of objects can be achieved simply by reducing object speed onset. However, excessive vergence movements can also be induced by depth discontinuities at a cut between two shots. A widespread solution to this problem is the active depth cut [Dve10, Men09] using a **dynamic horizontal image translation (DHIT)**: The S3D views are shifted in a temporally dynamic manner just before and after the cut in such a way that the objects of interest of both shots are located at approximately the same depth when the cut occurs, thereby eliminating exhausting vergence movements. The shifting operation is supposed to be done so slowly that it cannot be perceived.

At the beginning of this research project, there was the idea to use an eye tracker in order to reduce the AVD at the visual focus, i.e., at the point of regard by shifting the scene back or forth continuously using the DHIT. The idea led to the approach called **gaze adaptive convergence in stereo 3D applications (GACS3D)** [Eic13c], which represents one of the main contributions of this work. Other groups have later proposed similar automated systems [Ber14, Han14]. Instead of eye tracking, visual saliency calculations have also been used to automatically design the DHIT [Cha10, Han14]. But how is the DHIT supposed to be parametrized in such automated approaches? Usually, it is designed heuristically by a stereographer, but manual inspection is impossible for automated DHIT approaches. There are just a very limited number of publications in this field, and only vague recommendations are given rather than actual numbers. The goal of implementing and parameterizing the automated, gaze adaptive DHIT approach and the lack of publications was the motivation to conduct further research on the perception of DHIT. Specifically, how fast can the DHIT be performed without annoying the observers or being noticed by them, what factors influence the perception of DHIT, and how can the DHIT be improved to further reduce visual fatigue? These are the main research questions of this thesis.

One property of HIT, and therefore also of the DHIT, is that there is a certain distortion of depth [Smi12]. If an S3D scene is displayed completely in front of the display, it can only occupy the space between the observer and the display. However, if the same scene is shifted completely behind the display, the scene can occupy a space from the display to infinity. This stretching or compressing of scene depth is temporally dynamic in the case of DHIT, and its effect on the human visual system is unknown. The **distortion-free dynamic horizontal image translation (DHIT+)** is another major contribution of this work. The approach is based on the idea that a constant scene depth can be ensured by dynamically adjusting the camera baseline, i.e., the disparities of the scene according to the chosen DHIT parameters. This is perceptually similar to a camera movement along the depth axis.

1.2 Outline and Contributions

Following the introduction, this thesis starts with a description of the HVS in **chapter 2**. Knowledge about the HVS is important for two reasons. On the one hand, eye tracking is used in the proposed approach GACS3D, and respective filter techniques had to be developed. Therefore, the physiology of the human eye as well as its types of movements have to be known. On the other hand, the reader needs to know how depth is estimated and perceived by the HVS in order to be able to understand the concepts of S3D and visual fatigue.

Chapter 3 starts with the basics on S3D, where a simple S3D processing chain is explained, and the distortion of scene depth is mathematically derived so that the reader can fathom the proposed DHIT+. Afterwards, visual fatigue is described along with its origins and methods to measure it. Here, a more detailed description of the statements and concepts in the motivation above is given. The methods of measurement are important in the context of the experiments conducted for this work.

In **chapter 4**, the first two contributions of this work are described. At first, a brief introduction to eye tracking is given. However, the focus of this chapter is on filtering techniques in 2D and 3D space, which are needed for GACS3D. After a review of 2D filtering techniques, a Kalman filter tailored to *gaze-directed human-machine-interaction* is derived. This type of filter is capable of detecting outliers and extrapolating missing or rejected samples, which is very useful considering the erroneous results of common eye trackers. This filter utilizes a new kind of eye movement event detection that is compatible with missing samples. This event detection represents the first contribution of this work. Afterwards, a review of 3D visual focus estimation techniques is given since a highly accurate respective approach is needed in order for GACS3D to work. However, the common approaches rely on gaze data exclusively and yield insufficient accuracy due to eye tracker inherent system noise and oculomotor noise, i.e., involuntary, unconscious eye movements. Therefore, the second contribution of this work is proposed: The **probabilistic visual focus disparity estimation (PVFDE)** utilizes a real-time estimation of 3D scene information in order to improve accuracy.

Chapter 5 represents the main matter of this work. A detailed review of DHIT related works is given at the beginning. Afterwards, two major contributions of this work in the form of the aforementioned approaches DHIT+ and GACS3D are described. The latter comes with an automated floating window algorithm, which could be singled out as another major contribution because it is generally applicable to all automated DHIT approaches. The perceptual properties of the DHIT are investigated in a subjective experiment, and the performance of the two mentioned approaches is evaluated in three further experiments.

This work is finally concluded in **chapter 6**, where an outlook on future research is also given.

2 Human Visual System

In this chapter, some aspects of the **human visual system (HVS)**, that are relevant to this work, are described.

2.1 Physiology and Oculomotor Functions of the Human Eye

A top view drawing of the right human eye is displayed in **figure 2.1**. The light enters the eye through the **pupil**. The lens optically images the light on the **retina** in a horizontally and vertically flipped manner. The retina contains **cones** and **rods**, which are light sensitive cells used for color vision in bright surroundings and gray scale vision in dark surroundings, respectively. In addition to these different receptors, **adaption** to different lighting conditions is carried out by the **iris**, which functions as an aperture with adjustable diameter. The focal length of the optical imaging is defined by the cornea and the lens, and it can be adjusted by stretching the latter using the ciliary muscle. This **oculomotor** function is called **accommodation**. When a human looks at an object, its image is projected on the **fovea**, which is the central point on the retina with a very high density of cones. Here, the cones are about $1'$ apart, which defines the resolution of the human eye. The fovea only spans less than $2°$ of the visual field [Hol11], and the density of cones decreases drastically with increasing eccentricity. The line connecting the fovea, the center of the pupil, and the observed object is the **visual axis** of the eye, and the object or point gazed upon is the **visual focus**. The place on the retina where the optic nerve connects to the eye is the **blind spot** because there are neither cones nor rods located there. Due to the axial symmetry of the two eyes, humans do not actually perceive this blind spot in regular viewing conditions.

Muscles connected to the outside of the eyeballs enable yaw, pitch, and torsional

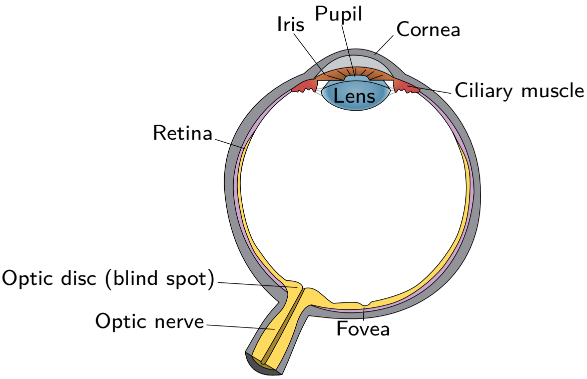


Figure 2.1: Physiology of the Human Eye (From Wikipedia Commons).

rotation, resulting in some specific kinds of eye movements, i.e., oculomotor functions of the eyes. These can be categorized into unconscious involuntary eye movements and conscious eye movements. **Table 2.1** lists all eye movements and their properties. Most of the time, a human observer will fixate on an object. The duration of such a **fixation** can vary between orders of milliseconds and seconds. The volatile movement to another object is called **saccade**. Saccades are the fastest eye movements with speeds up to 900 °/s, and they last less than 100 ms [Gol02]. This eye movement occurs mostly voluntarily, but can also be triggered instinctively, for example when an object is approaching the observer very fast. The saccadic reaction time is 240 ms on average [Joo03], but it can also be faster. Gezeck et al. have distinguished three modes of saccades, and the express mode exhibits reaction times as fast as 90 ms to 120 ms [Gez97]. Shortly before, after, and during a saccade, perception is drastically suppressed, which is called **saccadic suppression** [Gol02]. If an object of interest is moving, the observer will instinctively track the object with the eyes. This **smooth pursuit** cannot occur in absence of a moving stimulus, and the stimulus speed dictates the tracking speed, which can be as high as 100 °/s [Kor78]. If the stimulus is faster than that, additional saccades are used in order to track the stimulus. The last conscious eye movement happens in conjunction with saccades or smooth pursuit when the stimulus depth changes between time instances. The eyes rotate in opposite directions in order to place the target stimulus, located at the new depth, on the foveae of both

Table 2.1: Typical eye movement parameters [Gol02, Hol11, Kor78].

Type	Conscious	Duration (ms)	Amplitude (°)	Speed (°/s)
Fixation	✓	50 - 1000	-	-
Saccade	✓	30 - 80	4 - 20	30 - 900
Smooth pursuit	✓	-	-	10 - 100
Vergence	✓	-	-	1 - 20...
Drift	-	200 - 1000	1/60 - 1	1/10 - 5/12
Glissade	-	10 - 40	0.5 - 2	20 - 140
Microsaccade	-	10 - 30	1/6 - 2/3	15 - 50
Tremor	-	-	$<1/6$	$<1/3$

eyes. This is called **vergence movement**. The comparatively low speed of this kind of eye movement is dependent on the desired change in **vergence**. Speeds as low as 20 °/s have been reported [Kor78], but the speed is likely higher for bigger changes in vergence.

Even during a fixation, the eyes never remain completely still but exhibit some unconscious micro movements, which are listed in **table 2.1** as well. There is some jitter with very low amplitudes of $10'$ at high frequencies of 70 Hz to 90 Hz [Kor78]. This is called **tremor** and it is likely caused by imperfect muscle control [Gol02, Hol11, Kor78]. Furthermore, the eyes slowly **drift** away from the visual focus. In order to maintain the fixation, a **micro saccade** is carried out after some 100 ms to 200 ms [Kor78], to correct the drift. These two processes are believed to ensure high contrast output from the retinal receptors by continuously delivering fresh excitations to them [Gol02, Hol11, Kor78]. Finally, there is also an unconscious eye movement that occurs during saccades. Saccades may be programmed imperfectly and end up at the wrong position. In order to correct that, another form of micro saccades called **glissades** are used.

2.2 Perception of Depth

The HVS derives the 3D structure of an observed scene from numerous depth cues. These pictorial and oculomotor depth cues as well as their combination by the HVS are described in the following sections.

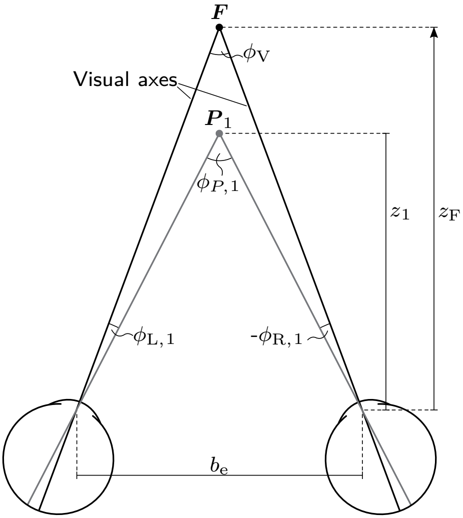


Figure 2.2: Retinal Disparity of a point $\boldsymbol{P}_1$ while fixating on $\boldsymbol{F}$ [How95].

2.2.1 Binocular Pictorial Depth Cue: Retinal Disparity

The perception of depth from binocular vision is called **stereopsis**. Due to the spatial offset of the eyes, a given scene is observed from two vantage points. This means that objects at different depths exhibit different spatial offsets in those views. The HVS analyses these offsets and combines the two views to a single 3D view of the scene through a process called **fusion**. This 3D view is also known as **cyclopean perception** because it is located right in the middle between the eyes [Men11, How95]. The process of fusion involves a correspondence analysis in which the **retinal disparity** D_i° of any given point $\boldsymbol{P}_i$ is estimated, which is linked to the depth of an object relative to the fixation point $\boldsymbol{F}$. The concept of retinal disparity is illustrated in **figure 2.2**. The visual axes intersect under an angle ϕ_{V}, which represents the **vergence angle** of the eyes. The point $\boldsymbol{P}_1$ is located at a different depth so that its images are located at a certain distance away from the left and right fovea. This distance can be expressed independently of eyeball size by the angles $\phi_{\mathrm{L},1}$ and $\phi_{\mathrm{R},1}$ between the **visual lines** of $\boldsymbol{P}_1$ and the visual axes. Generally, the angles $\phi_{\mathrm{L},i}$ and $\phi_{\mathrm{R},i}$ are positive if $\boldsymbol{P}_i$ is located on the right side of

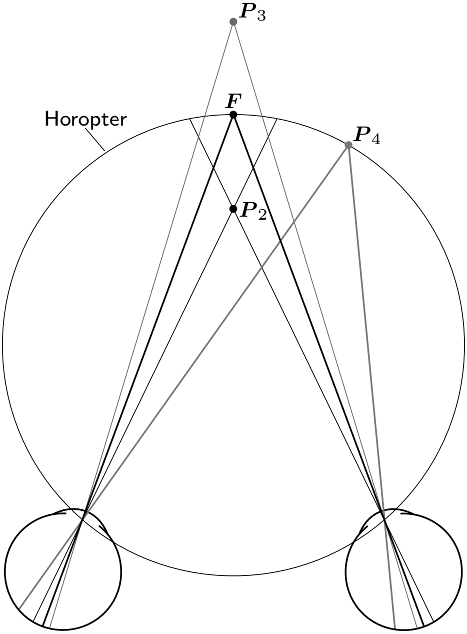


Figure 2.3: Points exhibiting crossed ($\boldsymbol{P}_2$) and uncrossed retinal disparity ($\boldsymbol{P}_3$), and a point $\boldsymbol{P}_4$ placed on the horopter, so that retinal disparity is zero [How95].

the respective visual axis, which means that $\phi_{\mathrm{L},1} = -\phi_{\mathrm{R},1}$ in the symmetric arrangement in **figure 2.2**. The retinal disparity D_i° of a point $\boldsymbol{P}_i$ is given by

$$D_i^\circ = \phi_{\mathrm{L},i} - \phi_{\mathrm{R},i}. \tag{2.1}$$

Since retinal disparity is given in angles, it is also called **angular disparity**. Another frequently used term is **binocular disparity**. It can also be calculated from the vergence angle ϕ_V and the angle $\phi_{P,i}$ between both visual lines of $\boldsymbol{P}_i$:

$$D_i^\circ = \phi_{P,i} - \phi_\mathrm{V}. \tag{2.2}$$

The retinal disparity is obviously zero for the fixated point, but it is also zero for all other points that meet $\phi_{\mathrm{L},i} = \phi_{\mathrm{R},i}$. These points are located on the

Vieth-Müller circle, i.e., horizontal **horopter** [How95], like $\boldsymbol{P}_4$ in **figure 2.3**. The circle connects the nodal points of the eyes and the fixation point $\boldsymbol{F}$, which means that $\boldsymbol{F}$ dictates the radius of the circle. The horopter also has a vertical component [How95], which is not relevant to this work, however. The retinal disparity of point $\boldsymbol{P}_2$ is positive and it is called crossed retinal disparity, because its visual lines intersect inside the horizontal horopter. Conversely, point $\boldsymbol{P}_3$ yields an uncrossed retinal disparity, which is negative.

Retinal disparity is a relative depth cue because the depth z_i of $\boldsymbol{P}_i$ and z_F may be scaled by any real factor and still yield the same retinal disparity, as can be seen in **figures 2.2 and 2.3**. For a symmetric arrangement like the one in **figure 2.2**, the retinal disparity is given by

$$D_i^\circ = \phi_{P,i} - \phi_\mathrm{V} = 2 \cdot \arctan\left(\frac{b_\mathrm{e}/2}{z_i}\right) - 2 \cdot \arctan\left(\frac{b_\mathrm{e}/2}{z_\mathrm{F}}\right) \tag{2.3}$$

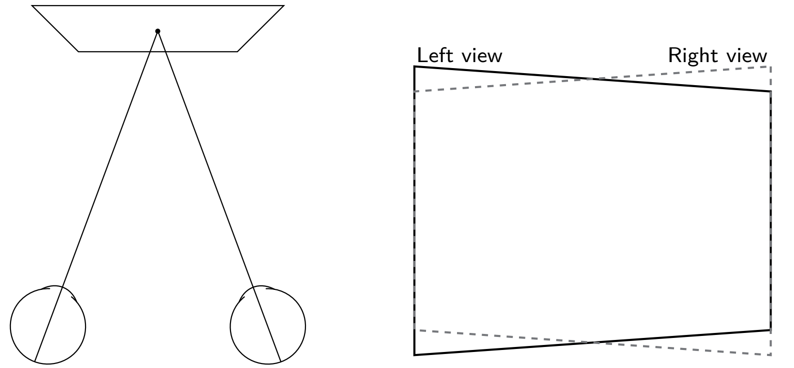
$$= 2 \cdot \arctan\left(\frac{\frac{b_\mathrm{e}/2}{z_i} - \frac{b_\mathrm{e}/2}{z_\mathrm{F}}}{1 + \frac{(b_\mathrm{e}/2)^2}{z_i \cdot z_\mathrm{F}}}\right) , \tag{2.4}$$

where b_e is the distance between the eyes, i.e., the interpupillary distance. The equation can be rearranged to retrieve the depth

$$z_i = \frac{b_\mathrm{e} \cdot z_\mathrm{F} - {b_\mathrm{e}}^2/2 \cdot \tan\left(D_i^\circ/2\right)}{2 \cdot z_\mathrm{F} \cdot \tan\left(D_i^\circ/2\right) + b_\mathrm{e}} . \tag{2.5}$$

in dependency on fixation distance z_F.

Retinal disparity usually refers to the horizontal, depth dependent, retinal offset of images described above. However, there is also **vertical retinal disparity**, as illustrated in **figure 2.4**. Due to the vergence angle of the eyes, an object in the periphery of the visual field appears bigger in one eye than in the other. The vertical size ratio of the images of that object is dependent on its eccentricity and point symmetric around 0° eccentricity[All04, Dve10]. If the rectangle in **figure 2.4** was moved closer to the eyes, they would converge further and the vertical disparities would increase. Hence, the horizontal gradient of vertical disparities can be used to estimate the vergence angle [Ban12, Wat05], which in turn yields the fixation distance z_F to scale the horizontal retinal disparities.



(a) Eyes fixating on the center of a rectangle.

(b) Flat projection of resulting images exhibiting vertical disparities.

Figure 2.4: Origin of vertical retinal disparity [Ban12, Wat05].

2.2.1.1 Limits of Fusion

If retinal disparity exceeds a certain value, fusion fails so that double vision, i.e., **diplopia** occurs. The small range around the horopter where fusion succeeds is called **Panum's fusional area** [Pan58]. A very detailed overview over the limits of fusion, i.e., the diplopia threshold is given by Howard and Rogers [How95]. The fusional limit is defined as the radius of Panum's fusional area and can be very small. For foveal stimuli, the fusional limit was found to be about $D^\circ = \pm 10'$ (minutes of arc) on average [Pal61]. However, the limit varies individually and is not always symmetrically distributed over crossed and uncrossed retinal disparity. Furthermore, according to Howard and Rogers, it is dependent on many factors, namely retinal eccentricity, spatial frequency, and the presence of other stimuli [How95]. The latter can reduce the fusional limits. For decreasing spatial frequencies, the fusional limits are increased and move asymptotically towards a linear function of spatial frequency [Sch84]. Furthermore, for an eccentricity of 6°, an increase of the fusional limit to about $\pm 30'$ was reported [Pal61] and after 10°, a linear increase of the fusional limit by 6 % to 7 % of the eccentricity angle was observed by numerous groups [Ogl64, Cro73, Ham83]. Since it is not easy to grasp how these fusional limits relate to three-dimensional space, the diameter of Panum's fusional area

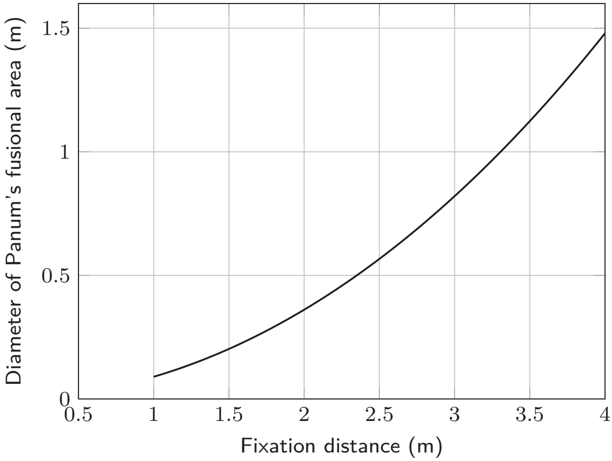


Figure 2.5: Diameter of Panum's fusional area as a function of fixation distance, for foveal stimuli and an interpupillary distance $b = 6.5\,\text{cm}$.

for foveal stimuli as a function of fixation distance, which can be computed via **equation (2.5)**, is plotted in **figure 2.5**. For a fixation distance of 1 m the diameter is as small as 9 cm. In other words, when fixating on a point $\boldsymbol{F}$ in a distance of 1 m, another closely located point $\boldsymbol{P}_i$ can only be fused if it is less than ±4.5 cm in depth away from $\boldsymbol{F}$, which can be easily verified in a self-experiment. However, this remarkably small fusional range can be accounted for by appropriate vergence movement.

2.2.2 Monocular Pictorial Depth Cues

In a survey by Richards [Ric70], about 4 % of a comparatively young test group was stereoblind, which means that stereopsis fails and no central 3D view is generated. For elderly people, the rate is increased to 14 % [Rub97]. Stereoblind individuals have to rely on monocular (and oculomotor) depth cues. Some are described in the following list [Dve10, Rei10], ordered by degree of relevance.

Occlusion

Whenever one object occludes a background object, the occluder is instantly perceived as in front of the background object. The HVS performs depth ordering, but the distance between objects cannot be estimated. This depth cue is also known as interposition [Dve10], and it can also be temporally dynamic.

Motion parallax

When the observer moves, the images of static objects move at different speeds across the retina. The speed depends on the depth offset between object and fixation point $\boldsymbol{F}$: the smaller the offset, the lower the speed. Objects farther than $\boldsymbol{F}$ move in the same direction as the observer, whereas the opposite is true for nearer objects.

Kinetic depth effect

This is the analogon to motion parallax but related to object motion rather than observer motion.

Linear perspective

Parallel lines converge to a vanishing point at infinity.

Texture gradient

The frequency response of any non-flat texture is dependent on observation distance. Furthermore, the texture gradient is altered by changes in surface orientation.

Depth of field

The **depth of field (DOF)** represents the distance between the nearest and farthest object that is perceived sharply. It is dependent on the accommodation state of the eyes, i.e., the fixation distance z_F as well as adaption state, specifically the pupil diameter. The DOF has been measured by numerous groups [Cha77, Cam57, Mar99]. For a pupil diameter of 3 mm, the DOF is about ±0.3 dpt (diopters, 1 dpt = 1/m), which corresponds to

$$z_i = \frac{1}{1/z_\mathrm{F} \pm 0.3\,\mathrm{dpt}} \ . \tag{2.6}$$

The amount of blur of a given point outside the DOF is dependent on the depth difference between that point and the visual focus.

Lighting and shading

The way objects react to light obviously provides a lot of information about their 3D structure, mainly surface orientation.

Relative size

The size difference of the images of two objects of the same kind makes estimation of the distance between those objects possible.

Known size

Knowing the size of an object enables estimation of the absolute distance to it.

Aerial perspective

Distant objects lose contrast and are colorized due to the atmosphere.

2.2.3 Oculomotor Depth Cues

The HVS can derive depth from the following three oculomotor functions.

Accommodation

The focal length of the eyes is obviously related to stimulus depth. So, the degree of contraction of the ciliary muscle yields information about the accommodative state of the eye.

Vergence angle

The relation between vergence angle and stimulus depth has already been explained in **section 2.2.1**. The vergence angle can be estimated from the respective muscle contractions and from the analysis of the horizontal gradient of vertical retinal disparities [Ban12, Wat05].

Pupil size

The correlation of pupil size with depth is not as obvious as with the other two cues. When fixating on very near objects, the pupil size will decrease in order to increase the very shallow DOF [Rei10].

2.2.4 Depth Cue Combination and Vetoing

Most depth cues are ambiguous, when observed separately. Here are just a few of many examples:

- A specific amount of blur due to optical imaging can be generated in front of or behind the visual focus.
- There is an infinite number of depth hypotheses that would exhibit an observed distribution of light and shade [Tod04].
- Simultaneous observer and object motion yield an infinite number of depth hypotheses.

It is obvious that ambiguous depth hypotheses have to be constrained by a priori knowledge and other depth cues in a cue combination process by the HVS somehow. However, research in the field of depth cue combination is not easy because depth cues cannot be isolated completely. A detailed review of previous works is given by Watt et al. [Wat05], which shall be summarized and extended in this section.

It has been shown empirically that the final depth sensation is a linear combination of the depth cues, weighted according to their estimated reliabilities [Wat05, Wis10]. However, the weighting of depth cues differs individually [Ban12]. While the 3D structure of an object is perceived very accurately, the absolute depth differences of a relief are usually underestimated by 38 % to 75 % [Tod04]. If no depth cue is reliable in a region, depth may be extrapolated from more reliable neighbor regions through a priori knowledge [Tod04]. In case of conflicts, some depth cues veto others [Wis10]. A famous example is the Randot-stereogram, with which Julesz proved that retinal disparity is in fact analyzed by the HVS [Jul60]: When viewed monocularly, the images just looked like random noise, but viewing it binocularly yielded a 3D sensation due to disparities contained in the pseudo-random images. While this minimalistic test stimulus served its purpose, the results of this experiment also mean that monocular and binocular depth cues were in conflict. Specifically, the retinal disparity cue vetoed the texture gradient cue [Wat05] because the latter suggests that the surface being looked at is flat rather than 3D. In general, largely conflicting depth cues annoy the observer, the fusion time is increased [Wat05], and in some cases bistability of the depth perception might even occur [Ee03]. The resulting distortion of depth may furthermore appear unnatural in certain conditions.

Retinal disparity is one of the strongest depth cues in that it vetoes most of the others. The dominance over the texture gradient cue has already been outlined above. However, as described in **section 2.2.1**, retinal disparity is a relative depth cue that needs to be scaled with fixation distance z_{F} in order to calculate the absolute depth of objects. Because of that, retinal disparity yields comparatively big estimation errors [Dve10], especially so at longer distances [Lam07, Wat05]. On the other hand, the reliability of the texture gradient cue is independent of fixation distance, so that this cue can outweigh retinal disparity in certain situations. The reliability of the retinal disparity cue is furthermore increased with high spatial frequencies and observation time [Rei10]. The DOF cue and its reliability is also dependent on fixation distance. The absolute fixation distance is predominantly estimated from the vergence angle and, according to Watt et al., possibly also the state of accommodation. The reliability of the accommodation cue is relatively low [Wat05] and so is the weight of these oculomotor depth cues in the cue combination process. In case of conflict, they are vetoed by retinal disparity. For example, when the visual axes are parallel, the vergence distance is at infinity where no retinal disparity is given in the real world. If retinal disparity has been artificially generated at virtual infinity, an observer will estimate depth much nearer than infinity [Sta97]. The only depth cue that always vetoes retinal disparity is occlusion. An object occluding another object is perceived nearer, regardless of possibly conflicting retinal disparity cues. The resulting depth distortion is generally deemed very annoying [Dve10, Hak11]. The occlusion cue is especially strong in a temporal manner, as in the case of motion parallax or the kinetic depth effect [Lam07].

2.3 Conclusion

The movements of the eye can be divided into conscious and unconscious types. The prior are comprised of fixations, saccades, smooth pursuit, and vergence. The distinction between conscious and unconscious eye movements as well as the properties of the different conscious eye movements are important when developing filters for *human-machine-interaction* using an eye tracker, as in **chapter 4**.

However, this chapter mostly established the theoretical background for understanding the concepts of S3D and visual fatigue by explaining the perception of depth by the **human visual system (HVS)**. The HVS analyzes binocular, monocular, and oculomotor depth cues. Most of them are inherently ambiguous and have to be combined by the HVS to construct the final depth perception. The cues are weighted according to their estimated reliability, but some depth cues veto others. The most prominent example is how occlusion vetoes disparity in the context of window violations in **stereo 3D (S3D)**, as described in the next chapter.

The binocular depth cue retinal disparity is one of the strongest depth cues. All points located on the horopter, a circle connecting the fixation point and the nodal points of the eyes, exhibit zero retinal disparity. If the visual lines of a non-fixated point cross in front of the horopter, its retinal disparity is positive and called “crossed”, whereas otherwise, it is negative and called “uncrossed”. This concept is adopted in the description of planar disparity in S3D. The limits of fusing retinal disparities are remarkably small. In analogy to this fusion, the stereoscopic fusion range is introduced in the next chapter as well, which represents the limits of fusion supported by vergence movements, i.e., motoric fusion, when watching S3D stimuli.

3 Stereoscopy and Visual Fatigue

In the last chapter, the basics of the **human visual system (HVS)** have been introduced. It has been pointed out that incongruent depth cues can alter the perceived depth significantly and possibly even annoy the observer. These factors have to be taken into consideration, when creating 3D content and technology. In the following sections, the basics of stereoscopy and its effect on the HVS are described.

3.1 Stereoscopy

Stereoscopy, also known as **stereo 3D (S3D)**, denotes the act of presenting two views of a scene to the observer in order to achieve a 3D sensation. The views are recorded or rendered using an S3D camera setup, where the cameras are horizontally offset by a certain amount, just like the human eyes. The recorded 2D images are presented to each eye separately by means explained in **section 3.1.2**. The HVS then fuses the views to a 3D sensation.

3.1.1 Basic Processing

Similarly to the concept of angular disparity described in **section 2.2.1**, the images of objects recorded by an S3D camera setup exhibit a certain offset that is dependent on depth. However, the scene is now projected onto planar camera sensors rather than spherical retinas. The resulting (planar) **disparity**, sometimes also referred to as **parallax**, can therefore best be expressed as a measure of distance. For reasons that are explained in **section 3.2.3.6**, a rectified camera setup is usually used[1], which means that the disparity of $\boldsymbol{P}_i$

[1] The cameras are aligned parallel and lens distortions have been corrected.

is purely horizontal and can be expressed in pixels by

$$D_i = x_{\mathrm{L},i} - x_{\mathrm{R},i}, \tag{3.1}$$

with the real pixel-coordinates $x_{\mathrm{L},i}$ and $x_{\mathrm{R},i}$ of the left and right view images of $\boldsymbol{P}_i$. In analogy to the concept of the horopter in **section 2.2.1**, there is a locus of points with zero disparity in stereoscopy. For the rectified parallel camera setup, these points in *3D scene space* are located in a plane at infinity that is parallel to the sensor planes and known as the **convergence plane**. All nearer points exhibit positive disparity so that the disparity range $[D_{\mathrm{min}}, D_{\mathrm{max}}]$ of the raw S3D views is contained in

$$0 \le D_{\mathrm{min}} < D_{\mathrm{max}} < \infty \ . \tag{3.2}$$

In *3D observer space*, the convergence plane is equivalent to the display plane, which means that the unprocessed recorded scene is perceived completely in front of the display. In order to shift the scene partially behind the display, a **horizontal image translation (HIT)** is applied to the S3D views [Bro11, Men11]. If one object of interest is supposed to be placed in the display plane, its disparity is nulled by the HIT. That disparity is called **convergence disparity** D_{conv}, with

$$D_{\mathrm{min}} \le D_{\mathrm{conv}} \le D_{\mathrm{max}} \ . \tag{3.3}$$

The convergence disparity is also known as the **zero parallax setting (ZPS)**. The HIT is applied by shifting the left view to the left by $D_{\mathrm{conv}}/2$ and the right view to the right by the same amount [Bro11]. In this way, all disparities D are transformed into the *shifted domain*

$$\tilde{D} = D - D_{\mathrm{conv}} \ . \tag{3.4}$$

The absolute disparity budget

$$D_{\mathrm{B}} = D_{\mathrm{max}} - D_{\mathrm{min}} = \tilde{D}_{\mathrm{max}} - \tilde{D}_{\mathrm{min}} \tag{3.5}$$

obviously remains unchanged by this process. Negative disparities are introduced if $D_{\mathrm{conv}} > D_{\mathrm{min}}$. Just like in **section 2.2.1**, disparities are called *crossed* if the visual lines cross in front of the stereoscopic horopter-equivalent, i.e.,

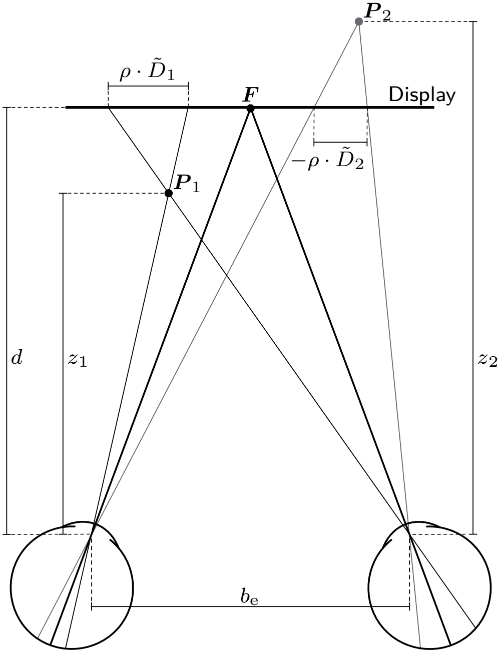


Figure 3.1: Illustration of crossed ($\tilde{D}_1$) and uncrossed disparity ($\tilde{D}_2$). In this example, the observer fixates on a point in the convergence plane.

the display plane, or uncrossed otherwise, see **figure 3.1**. As can be deduced from **equation (3.1)**, crossed disparities are positive. Technically, D_{conv} does not actually need to satisfy **equation (3.3)** because the scene could be shifted far behind or in front of the display. This is usually unwanted, however. In 3D scene space, the HIT moves the convergence plane from infinity closer to the cameras. Since the convergence plane is moved, this process is sometimes also referred to as **reconvergence** [Men11].

Considering **figure 3.1**, the depth z_i of $\boldsymbol{P}_i$ can be computed from its disparity through a simple application of the intercept theorem, which yields

$$\frac{z_i}{b_\mathrm{e}} = \frac{d - z_i}{\rho \cdot \tilde{D}_i} \qquad \text{for } z_i \geq 0 \qquad (3.6)$$

$$\Leftrightarrow z_i = \frac{d \cdot b_\mathrm{e}}{b_\mathrm{e} - \rho \cdot \tilde{D}_i} \qquad \text{for } \tilde{D}_i \geq -b_\mathrm{e}/\rho \ , \qquad (3.7)$$

where d is the display distance and ρ is the pixel pitch, i.e., the distance between two pixels on the display screen. If the restriction in **equation (3.7)** is not met, eye divergence occurs, which should be avoided when designing the HIT because it is visually unpleasant [Bro11, Men11]. The **depth budget** is defined as

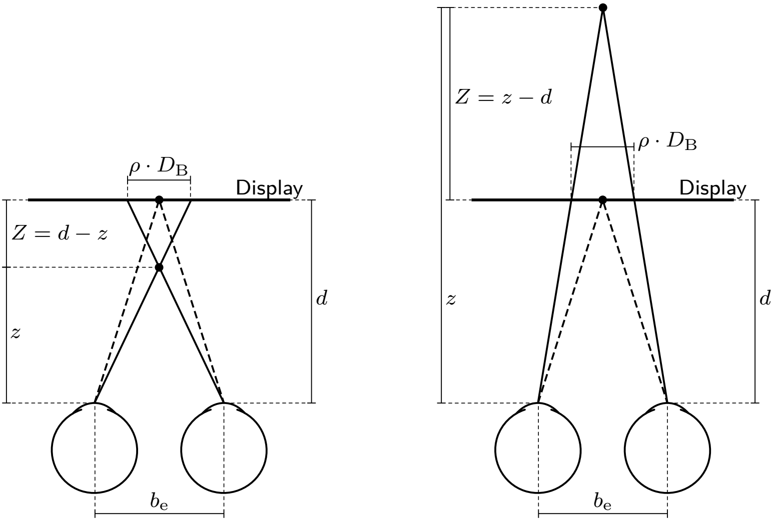
$$Z = z_{\mathrm{far}} - z_{\mathrm{near}} \ , \tag{3.8}$$

where z_{near} is the nearest observed depth, as computed from $\tilde{D}_{\mathrm{max}}$ using **equation (3.7)**, and z_{far} is the farthest observed depth corresponding to $\tilde{D}_{\mathrm{min}}$. Hence,

$$Z = \frac{d \cdot b_{\mathrm{e}}}{b_{\mathrm{e}} + \rho \cdot \tilde{D}_{\mathrm{min}}} - \frac{d \cdot b_{\mathrm{e}}}{b_{\mathrm{e}} + \rho \cdot \tilde{D}_{\mathrm{max}}} \ . \tag{3.9}$$

As can be seen, Z is indirectly dependent on the convergence disparity D_{conv} via $\tilde{D}_{\mathrm{min}}$ and $\tilde{D}_{\mathrm{max}}$. This means that the depth budget is not constant for different convergence disparities in contrast to the **disparity budget**. As an example, consider **figure 3.2a**: An infinitely deep scene is displayed without any HIT or sensor shift, so that the displayed disparity range is $[0, D_{\mathrm{B}}]$. In **figure 3.2b**, this scene is shifted completely behind the display by applying the HIT using $D_{\mathrm{conv}} = D_{\mathrm{B}}$, which alters the displayed disparity range to $[-D_{\mathrm{B}}, 0]$. The disparity budget is obviously the same in both cases, the depth budget Z, however, changes strongly between these figures. A similar description of this effect based on the “shape ratio” has been given by Smith and Collar [Smi12]. The worst-case scenario is given for $\rho \cdot D_{\mathrm{B}} = b_{\mathrm{e}}$, which simply means that the metric disparity is equal to the distance between the eyes. Using **equation (3.9)**, this yields depth budgets $Z = \infty$ and $Z = d/2$, respectively.

Due to the translation and the limited recorded image width, black borders may appear on opposite sides of the views. This can be avoided by image acquisition with some extra width or rescaling of the S3D images [Bro11]. However, instead of applying an HIT in post-production, the sensors behind the lenses may be shifted to achieve the same effect without having to acquire extra width [Sch05]. The disadvantage of the sensor shift approach is that lens distortions are stronger on the outer regions of the lens [All04]. Broberg



(a) An unprocessed scene is perceived completely in front of the display.

(b) The same scene is shifted completely behind the display. The depth budget is increased.

Figure 3.2: Illustration of the distortion of the depth budget due to HIT. The disparity budget is the same in both cases.

offers some further guidance to the design of HIT [Bro11]. There is also a temporally dynamic HIT that is explained in **section 3.2.3.2.1**.

3.1.2 Fundamentals of Stereo 3D Presentation

The first stereoscope was invented by Wheatstone in 1838. It consists of two mirrors that direct the visual axes of the eyes to two separate 3D drawings[2] exhibiting some disparity [Whe38]. Going directly to the present, a straight adoption of Wheatstone's concept is the usage of two separate displays placed close to the eyes behind lenses in a head-mounted setup, i.e., head-mounted displays. Both of these stereoscopes can only be used by a single individual at a time. Obviously, multiuser concepts using only one display are preferred in many

[2] Photography had not been invented yet.

applications. There are two basic single display S3D presentation protocols: synchronous and time-sequential. While the synchronous presentation of S3D views is natural, the time-sequential approach does exhibit an unwanted side effect. When an observer is tracking an object, the eyes move continuously and the object is expected to be at different positions in each frame. However, since the S3D views have been recorded synchronously, but are presented sequentially to each eye, the object does not actually move between those two time instances. The spatial offset between the expected and the presented position is perceived as a disparity due to the presentation to different eyes. This means that the perceived depth of a tracked object is distorted. This phenomenon is also known as the Mach-Dvorak effect [Hof11]. The distortion increases with increasing object speed and decreasing capture rate.

Despite this side effect, the time-sequential approach is very wide spread due to its cheap and easy implementation. In the case of projection screens, the time-sequential presentation can be implemented by using a high frame rate projector in conjunction with active shutter glasses that darken a single eye synchronously to the projector presentation, e.g., *XPAND Active Shutter 3D* [XPAa]. Alternatively, passive polarization filter glasses can be used with a synchronized polarization modulator in front of the projector, e.g., *XPAND Passive 3D Polarization Modulator Gen2* [XPAb] or *RealD XL* [Rea]. The same glasses can also be used in a synchronous presentation approach, either with two projectors or with a single projector top-and-bottom arrangement of the S3D views, which are projected via two separate objectives, e.g., *Sony LKRL-A502* [Son]. A less commonly used approach is the wavelength multiplex, where the views for each eye are presented with slightly offset color primaries. Special color filter glasses are used to separate the views again, e.g., *Dolby 3D* [Dol]. For direct view displays, the synchronous approach is implemented by using polarization filter glasses in combination with an alternating polarization rotation of each line such that all uneven lines are only visible to the left eye and all even lines to the right eye, e.g., *LG Electronics* [Isr11]. This effectively halves the vertical resolution presented to each eye. The time sequential approach is implemented using high frame rates and shutter glasses, e.g., *Samsung* [Sam15]. Due to the frame rate upconversion of modern displays, the aforementioned distortion of depth is not as severe as in the case of projection screens. However, the frame rate upconversion may introduce new artifacts.

3.1.3 Perception of 2D vs. Stereo 3D

The perception of S3D is more than just the added 3D sensation. In 2D content, the monocular depth cues are in conflict with the absent disparity cue. This is especially problematic when object motion and temporally dynamic occlusion is involved: the occluded object exhibits the same disparity as the occluder. In regular 3D viewing conditions, this can only ever happen at infinity. Hence, the observer feels distanced from the scenery and becomes a passive observer [San12b]. An S3D display, however, is perceived as a window, i.e., **stereoscopic window** into another 3D world. The personal space of the observer is penetrated by the scenery with the effect of heightened immersion. The emotional effect of S3D can best be observed in movies featuring 3D shock effects, where an object moves rapidly towards the observer who instinctively reacts by ducking out of its way before even identifying it [San12a]. The increased intensity of S3D has also been shown in a study by Ujike et al., where visually induced motion sickness in S3D and 2D viewing conditions was measured physiologically and subjectively [Uji11]. So apparently, the generally more intense perception of S3D also applies to problematic content, which carries the potential to exhaust the HVS.

3.2 Visual Discomfort and Visual Fatigue in Stereo 3D

3.2.1 Definition

Visual discomfort and visual fatigue are frequently used as synonyms in the literature, but Lambooij et al. suggested a distinction of these two terms [Lam07], which is adopted throughout this work.

Visual discomfort denotes the subjective sensation of discomfort during and after exposure to problematic S3D content.

Visual fatigue refers to the objectively measurable exhaustion of the HVS during and after exposure to problematic S3D content.

3.2.2 Properties and Measurement

The symptoms of visual discomfort and visual fatigue include, but are not limited to: headaches or pain in other areas like neck and shoulders, eyestrain, dry eyes, blurred vision, uncomfortable vision, reduced oculomotor mobility and loss of concentration [Lam07]. While visual discomfort may be induced immediately by inappropriate S3D stimuli, most symptoms are accumulated over time so that they are only really assessed after prolonged exposure [Iat14, Lam09a]. However, the occurrence and severity of symptoms varies individually [Shi11]. Subjects exhibiting visual degradations are generally more prone to visual discomfort and visual fatigue than healthy subjects [Lam09a].

Lambooij et al. have given a detailed overview about subjective and objective evaluation methods [Lam09b]. Visual discomfort is mostly evaluated through subjective assessments. The most common test methods are outlined in the respective ITU recommendations [ITU12a, ITU12b]. In these methods, the visual discomfort induced by brief stimuli is rated individually or in a pair comparison paradigm. For prolonged stimulus exposures, the severity of the aforementioned symptoms can be evaluated using a questionnaire [Iat14, She03].

The evaluation of visual fatigue is more elaborate. Its symptoms can be quantified using optometric and brain activity measurement methods [Lam09b]. However, due to the individual symptomatic differences, all symptoms would have to be measured. This is problematic because each measurement consumes a certain amount of time and the HVS can partially recover quickly [Lam09a], which introduces increasing systematic uncertainty to the later measurements. Furthermore, since most symptoms do not occur early on, an evaluation of visual fatigue is unsuitable for experiments with brief stimuli, especially in the case of pair comparison approaches.

Some groups have tried to combine subjective and a small subset of objective measurements in order to develop a quick and easy evaluation approach for visual discomfort and visual fatigue. Lambooij et al. have used questionnaires and eight optometric tests before and after relatively short but stressful stimuli [Lam09a]. They have used an algorithm to classify the susceptibility of each subject towards visual fatigue. A meaningful alteration of fusion range has been discovered for the susceptible group of subjects only, whereas all other optometric tests have yielded no significant alteration. This might have been due to the comparatively short stimulus exposures. Iatsun et al. have evaluated visual discomfort and visual fatigue in 2D and S3D viewing conditions using questionnaires and eye tracking features like the number of blinks and saccades per time interval [Iat14]. The total time of stimulus exposure has been 60 min, but the test has been halted every 10 min to collect visual discomfort ratings for each time slot. The visual discomfort ratings have steadily increased over time. Interestingly, they have also increased in the 2D viewing condition, albeit at a much lower pace. The eye tracker results have exhibited a decline in the number of saccades after 50 minutes of exposure, but a strong dependency on content has also been found.

In conclusion, the evaluation of visual fatigue remains largely unsolved. It is therefore generally recommended to evaluate visual discomfort using the methods outlined above.

3.2.3 Causes and Solutions

In this section, the sources of visual discomfort and visual fatigue are described, as well as methods to prevent them.

3.2.3.1 Accommodation Vergence Discrepancy

There is an inherent problem with all S3D displays [Lam09b]: The disparity of a stereoscopic stimulus might suggest that it is located behind or in front of the display, while the imagery is actually presented at the display distance, as illustrated in **figure 3.3**. Hence, the accommodation and vergence cues differ, which is unnatural and known as the **accommodation vergence discrepancy (AVD)** [War95] or **accommodation vergence conflict** [Lam09b, Shi11, Hof08]. Now, there are two basic theories how the HVS handles this issue.

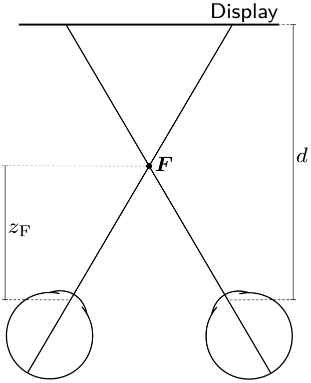


Figure 3.3: Accommodation vergence discrepancy: Vergence or fixation distance z_{F} is not equal to the distance to the true imagery, i.e., display distance d.

Hoffmann et al. and other groups have argued that the eyes fixate on the stereoscopic stimulus through vergence movements, while accommodation actually remains fixed on the display distance [Hof08]. Hence, the AVD is supposedly always present when watching S3D content, but tolerable to a certain degree.

More recently, other groups have claimed that accommodation actually shifts away from the display along with the vergence movement [Lam09b, Shi11] because these mechanisms are neurally cross-connected [Shi11]. According to these groups, the discrepancy between display distance and accommodation distance has basically no effect on the HVS as long as the display plane is contained in the **depth of field (DOF)** surrounding the fixation point $\boldsymbol{F}$. Otherwise, blur is perceived, which triggers the adjustment of accommodation and vergence towards the display plane, while retinal disparity still triggers vergence in the opposite direction. Thereby, an unstable system is created.

Regardless of which theory is correct, researchers concur that big AVDs are one of the main sources for visual discomfort and visual fatigue [Hof08, Lam09b, Shi11]. The basic approach to solve this issue is to limit disparities to a zone of comfort. This can be achieved by

1. choosing appropriate capturing parameters so that the disparity budget is small enough to fit inside the zone of comfort,

2. applying the HIT or sensor shift, as described in **section 3.1.1**, to place the scene in the zone of comfort and objects of interest near the display plane.

3.2.3.1.1 Stereoscopic Fusion Range

If the AVDs become too big, blur is perceived or fusion fails completely, so that diplopia occurs, which is uncomfortable. This leads to the concept of the **stereoscopic fusion range (SFR)**, also known as the **zone of clear single binocular vision**. In contrast to Panum's fusional area, the SFR does not represent a range of disparities that can be fused simultaneously, but rather those that can be fused on an S3D display after appropriate vergence movements, which is also known as **motoric fusion**. Due to the vergence movements, the SFR is a lot bigger than Panum's fusional area. Furthermore, it varies individually and can be increased by training [Lam07]. It is measured subjectively by having the subjects report when either blur occurs or fusion breaks, i.e., measuring blur or break points, respectively. In a 1990 experiment with 8 subjects, the SFR was measured using break points, resulting in a range of $-1.57°$ to $4.93°$ in retinal disparity [Yeh90]. The SFR is displayed in **figure 3.4** as a function of viewing distance, along with some other graphs, which are explained in the next section. Most other experiments of this kind have been conducted using prisms [Emo05, She34]. However, using prism glasses perceptually differs from S3D. With prism glasses, the AVD is constant over the whole visual field and the HVS can adapt to that [Yan04]. In S3D, the AVD varies spatially according to the presented scene structure.

3.2.3.1.2 Zone of Comfort

The AVD is linked to the DOF. It is still unclear whether the AVD exists inside the DOF. However, experts concur that the problems outlined above occur for stimuli located on the outside [Ban12, Lam09b]. The inherent consequence is to limit the portrayed depth to the DOF. As mentioned in **section 2.2.2**, the DOF spans about $\pm 0.3\,\text{dpt}$ and the respective depth values are plotted in **figure 3.4**. Using **equation (2.4)** and **equation (2.6)**, the DOF corresponds

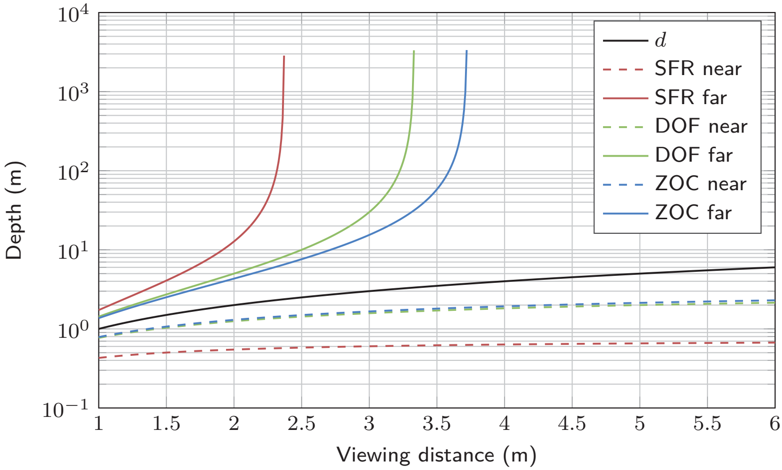


Figure 3.4: Plot of stereoscopic fusion range [Yeh90], depth of field (±0.3 dpt) [Cha77] and zone of comfort [Lam09b] as a function of viewing distance d for a pupil diameter of 3 mm and an interpupillary distance $b_\mathrm{e} = 6.5\,\mathrm{cm}$.

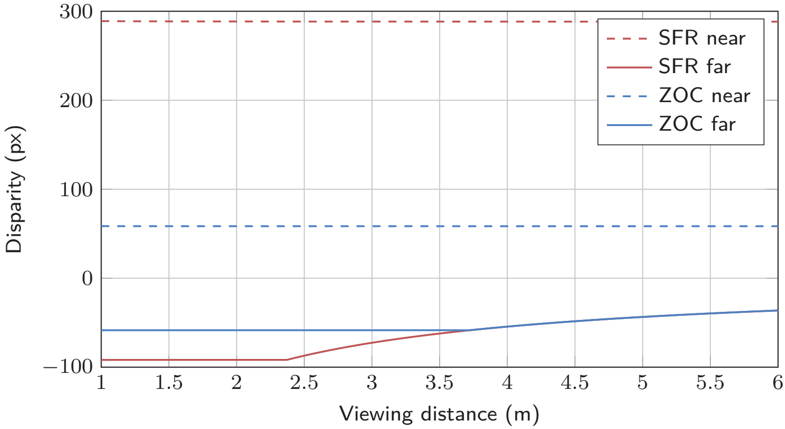


Figure 3.5: Plot of stereoscopic fusion range [Yeh90] and zone of comfort [Lam09b] in terms of disparity $\tilde{D}$ for a Full-HD display at design viewing distance $d = 3.1 \cdot H$ and an interpupillary distance $b_\mathrm{e} = 6.5\,\mathrm{cm}$.

to a retinal disparity of $D^\circ = 1.11°$. Since the DOF varies individually and is dependent on lighting conditions, a more conservative recommendation is to limit retinal disparities to $|D^\circ| \leq 1°$ [Lam09b]. The resulting range of depths that can be viewed comfortably is called the **zone of comfort (ZOC)** and is plotted in **figure 3.4** as well. Other groups have recommended similar values based on other assumed DOFs, e.g., $|D^\circ| \leq 0.8°$ due to a DOF of ±0.2 dpt [Yan04]. There have also been other recommendations like Percival's or Sheard's ZOC [She34], which are based on prism glasses and therefore not really appropriate for S3D, as mentioned in the last section.

The SFR and ZOC are also plotted in terms of disparity in pixels in **figure 3.5**. The values are computed for a Full-HD display at design viewing distance [ITU12a], i.e., $d = 3.1 \cdot H$, where H is the metric height of the stimulus display. The design viewing distance is an important concept here because it is the closest distance any observer should ever sit away from the screen. The observed disparities decrease with increasing viewing distance so that the design viewing distance serves as a worst case scenario as far as the AVD is concerned. In this setup, the ZOC constitutes ±58.4 px in disparity. Another, comparatively simple recommendation can be deduced from **figure 3.4**. The points where the graphs turn towards infinity exhibit a vergence angle of $\phi_\mathrm{V} = 0°$. So, a further increase of disparity would lead to eye divergence, which should be avoided [Bro11, Men11]. In **figure 3.5**, this constraint translates into the asymptotic increase of the uncrossed disparity limit towards 0 px.

In conclusion, the disparity budget must be small enough to fit inside the ZOC. This can be achieved by adjusting the camera baseline b_c. A widespread rule of thumb in stereography is to set b_c to 1/30th of the depth of the closest object in the scene [Men09]. This rule does not impose any restrictions on uncrossed disparity. However, as can be seen in **figure 3.4**, the ZOC extends to infinity for viewing distances $d > 3.7$ m, so that the lacking restriction does not pose an issue in most viewing conditions. The rule also does not specify focal length, though, which is problematic, because a wide-angle lens will yield much smaller disparities than a regular lens. A better option is the "Percentage-Rule" [Shi11], which directly states that crossed disparity should be smaller than 2 % to 3 % of the display width W, while for uncrossed disparities lower percentages of

1 % to 2 % are recommended. The 3 % recommendation is approximately the same as the $D^\circ \leq 1^\circ$ ZOC [Lam09b] on a Full-HD display at design viewing distance, since $0.03 \cdot 1920\,\text{px} = 57.6\,\text{px} \approx 58.4\,\text{px}$, see **figure 3.5**. There are many stereoscopy calculators available online [Tau], which help in implementing this recommendation by setting an appropriate camera baseline based on the distance to the foremost object and the focal length.

3.2.3.2 Oculomotor Stress

It is not surprising that excessive usage of the oculomotor system leads to an exhaustion of the visual system. It is predominantly gaze point motion along the z-axis, i.e., vergence movement that causes visual fatigue [Kim14, Lam09b, Tod04]. Vergence movements can be induced by S3D content in two ways.

Firstly, there is fast object motion along the z-axis. A fast motion towards the observer, like the 3D shock effects described in **section 3.1.3**, is very commonly used in S3D movies because of the strong and immediate emotional reaction by the audience. Limiting the amount of these 3D shock effects is an obvious method to reduce visual fatigue.

Secondly, vergence movement can also be induced by strong depth discontinuities at shot cuts. Not only do these depth discontinuities cause visual fatigue, but also loss of fusion for an extended time period due to the relatively slow vergence movement capabilities of the HVS.

3.2.3.2.1 Dynamic Horizontal Image Translation

The depth discontinuities at shot cuts can be mitigated by the **active depth cut** [Men09, Dve10]. An HIT is performed over a couple of seconds in a temporally dynamic manner just before and after the shot cut. This is done in such a way that the objects of interest of both shots are located at approximately the same depth during the cut. This **dynamic horizontal image translation (DHIT)** is done so slowly that it is supposedly not perceived. It is usually performed linearly on a range of convergence disparities, e.g., from $D_{\text{conv,min}}$

to $D_{\text{conv,max}}$ or vice versa, with

$$D_{\text{min}} \leq D_{\text{conv,min}} \leq D_{\text{conv}} \leq D_{\text{conv,max}} \leq D_{\text{max}} \ . \tag{3.10}$$

The absolute difference between start and end convergence disparity is called **shift budget**

$$\text{SB} = D_{\text{conv,max}} - D_{\text{conv,min}} \tag{3.11}$$

throughout this work.

The distortion of the depth budget described in **section 3.1.1** becomes temporally dynamic in the context of the DHIT. The effect of that distortion on the HVS is unknown. A distortion-free extension of the DHIT is proposed in **section 5.2.1**. Since the DHIT is the main topic of this thesis, it is described and analyzed in more detail in **chapter 5**.

3.2.3.3 Crosstalk

Crosstalk is a technology issue and denotes the phenomenon that the left view on an S3D display is partially visible to the right eye and vice versa. Measurement is usually done by displaying plain black on the intended S3D view and white on the unintended view and measuring the resulting leakage luminance L_{Leak}. The amount of crosstalk is then given by

$$\text{Crosstalk} = \frac{L_{\text{Leak}} - L_{\text{Black}}}{L_{\text{Signal}} - L_{\text{Black}}} \cdot 100\,\% \ , \tag{3.12}$$

where L_{Black} is the black level of the S3D display and L_{Signal} is the luminance of the white view [Woo12]. Woods has also described some other crosstalk evaluation metrics [Woo12].

Crosstalk is reportedly perceivable for values of 0.3 % to 2 % and visual discomfort is generated at 5 % [Kap07]. For polarization based S3D displays, crosstalk is very low in the range of 0.1 % to 0.3 % [Pas97]. In time-sequential S3D, weak synchronization for example can lead to very big amounts of crosstalk. In 1997, values of 20 % have been reported [Pas97], but more recently approximately 0.5 % crosstalk was measured [Bar11]. So, using a modern, high quality, preferably polarization based, S3D display ensures that crosstalk is kept at acceptable values.

3.2.3.4 Retinal Rivalry

Any differences between the images on the retinas of the two eyes, which are not due to disparity, cause **retinal rivalry**. It is sometimes also referred to as **binocular rivalry** and can cause the HVS to suppress one retinal image locally, thereby compromising fusion [How95]. It can furthermore cause visual discomfort and visual fatigue [Lam09b, Men11]. Retinal rivalries can occur in natural viewing, induced by iridescence, sparkle and occlusion [Men11]. The first two factors actually serve as depth cues, as described in **section 2.2.2**. Their extent is usually very small, which is why they are not really disturbing the HVS. Occlusion on the other hand can cause very big rivalries, depending on the distance between occluder and background. In order to avoid these visual discomfort inducing rivalries, the HVS tends to avoid fixating them, e.g., the background very close to an occluder.

In the context of S3D, retinal rivalries can be caused by numerous technological issues as explained in the following list:

Synchronization issues during recording

If the views are not perfectly synchronized, there will be small rivalries due to object or camera motion [Men11]. Vertical disparity is also introduced through vertical object motion, which distorts the depth perception [All04, Men11]. Therefore, precise camera synchronization is important in S3D production.

Color or luminance rivalry

Imperfect inter-view synchronization of Camera settings and post-processing as well as interocular lens asymmetries may alter color and luminance distribution locally or even globally, causing strong retinal rivalries [Men11]. Some of these errors can be corrected in post-production, but accurate capturing parameter synchronization is again of utmost importance as well as high quality matched lenses. A well-known example for color rivalry are anaglyph S3D glasses.

View misalignment during recording or playback

Small view misalignments can be compensated by the HVS through appropriate eye movements. If that fails, strong retinal rivalries are induced. A horizontal offset is harmless, since it is equivalent to the commonly used HIT approach, described in **section 3.1.1**. A vertical offset below 0.5° can be compensated by vertical vergence movements without visual discomfort [All04]. A much more critical view misalignment is the rotation around the display normal. The eyes compensate up to a few degrees of rotational misalignment through rotation of the eyes along the visual axis, which is called cyclovergence, but it causes strong visual discomfort and visual fatigue [Ban12]. However, all view misalignments can be fixed easily in post-production.

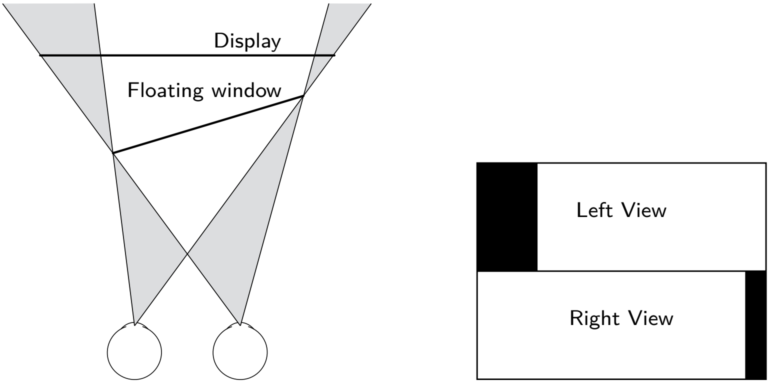
3.2.3.5 Window Violation

Window violations are a kind of artifact in S3D that is known under many synonyms including "breaking the proscenium rule" [Dve10], "frame cancellation" [Tod04], and "frame violation" [Hak11]. The artifact occurs when an object is stereoscopically displayed in front of the display plane, while it simultaneously reaches out of the display area at the left or right display border. The display border exhibiting zero disparity is perceived as an occluder of the foreground object. This indicates that the object is located behind the display border. However, the non-zero crossed disparity of the object indicates that it is actually located in front of the display border. Hence, there is a conflict between the depth cues occlusion and disparity and the prior vetoes the latter in the depth cue combination process described in **section 2.2.4**. The stereoscopic window is violated by conflicting disparity cues, giving this artifact its name. The conflict induces a distortion of depth and visual discomfort [Dve10, Hak11]. Therefore, it is to be avoided. Window violations do not occur at the top or bottom border of the display, since these do not exhibit any disparity cues. The stereoscopic window is simply perceived as if it were curved, which is not harmful [Dve10].

3.2.3.5.1 Floating Window

The method to avoid window violations is to move the stereoscopic window in front of the violating object by assigning a stereoscopic depth to it. The stereoscopic window then floats in front of the display rather than occupying the display plane, see **figure 3.6a**. This **floating window** can be generated simply by rendering black borders on opposing sides of the stereoscopic views, as illustrated in **figure 3.6b**. This effectively adds a specific amount of crossed disparity to the borders [Gar11, Dve10]. The floating window disparity must be at least as big as that of the violating object in order to resolve the cue conflict. Since the HIT, explained in **section 3.1.1**, might change the violation disparity, the floating window should be designed after applying the HIT. The floating window borders do not necessarily have to be symmetric on each side. In fact, the floating window can be tilted or even bent around all axes to enhance the emotional effect and overall story telling [Gar11]. For example, shifting the floating window away from the observer makes the scene objects stick out further and potentially more threatening. One thing to consider are the gray areas in **figure 3.6a**, which represent retinal rivalries because those areas are only visible to one eye. These rivalries are due to occlusion and, therefore, no stereoscopic deficits. However, as pointed out in **section 3.2.3.4**, they are still uncomfortable to look at and should be minimized for that reason. The floating window should be designed in such a way that window violations are prevented and only a small number of objects occupy the areas of retinal rivalry. There are also *dynamic* floating windows, which are simply animated. The motion of the floating window is said to be not perceivable if it follows the motion of the scene [Gar11].

In conclusion, floating windows are easy to create and can effectively remove window violations. However, certain properties of the underlying display technology may hinder the performance of this approach. Firstly, crosstalk is especially visible on the black floating window borders, which is very annoying [Gar11]. Secondly, in the case of a home viewing environment, the physical display frame is usually visible. This is problematic because one eye sees the display frame next to the imagery, whereas the other sees the black floating window border next to it. This generates annoying retinal rivalry. Because of



(a) Slanted floating window, as produced by **figure 3.6b**, with areas of retinal rivalry marked in gray. [Gar11]

(b) Non-symmetric floating window borders in a Top-and-Bottom S3D layout.

Figure 3.6: The principle of the floating window and retinal rivalry.

that, it is recommended to view S3D content in a dark room on a display with a black physical frame.

3.2.3.6 Notes on Vertical Disparity

In S3D, vertical disparity can be introduced globally by a view misalignment, as described in **section 3.2.3.4**. It can also be introduced locally varying, either by synchronization errors, see **section 3.2.3.4**, or by a camera toe-in analogously to **figure 2.4**. The latter induces vertical disparities at the outer regions of the recorded views. Allison has discussed this topic in detail [All04]. Research regarding the perception and fusion of vertical disparity is lacking, especially with respect to visual discomfort. It is known, however, that the vertical SFR is smaller than the horizontal SFR, which suggests that the same might be true for the vertical ZOC [All04]. But regardless of the examination of visual discomfort and visual fatigue, there are facts to consider that call for an elimination of all vertical disparity through the process of rectification.

Firstly, vertical retinal disparity in natural viewing exists only in peripheral vision. In S3D, the observer does not necessarily look at the convergence point of the cameras, which means that there is unnatural foveal vertical disparity. That disparity must not be bigger than 0.5°, or visual discomfort will occur [All04]. This limit is especially problematic for synchronization induced vertical disparity, which is dependent on the object speed and therefore not bounded.

Secondly, the vertical disparity distribution is superimposed on the natural vertical retinal disparity, which is an important depth cue. This means that depth perception is distorted [All04, Dve10].

3.3 Conclusion

A **horizontal image translation (HIT)** can be applied to the **stereo 3D (S3D)** views in opposite directions in order to shift the portrayed scene along the depth axis, e.g., so that it extends partially behind the display. There is also a temporally **dynamic horizontal image translation (DHIT)**. This technique is predominantly used in the context of active depth cuts, where strong depth discontinuities of objects of interest at shot cuts are mitigated by adjusting the depth of those objects before and after the cut. The discontinuities would otherwise induce oculomotor stress, which is a source for visual fatigue. There is a certain distortion of depth due to the HIT. This distortion becomes temporally dynamic in the case of the DHIT. Therefore, a distortion-free DHIT is proposed later, in **section 5.2.1**.

Visual fatigue represents the objectively measurable exhaustion of the human visual system due to prolonged exposure to S3D content. There is also a subjective counterpart called visual discomfort, which denotes the subjective and sometimes immediate sensation of discomfort. The symptoms of visual fatigue are manyfold and vary individually. Therefore, it is hard to measure it reliably. An established recommendation is to measure visual discomfort instead using questionnaires.

There are many sources for visual fatigue. Aside from the aforementioned oculomotor stress, this thesis is also focused on reducing the accommodation vergence discrepancy: The eyes of an observer converge on a stimulus behind or in front of an S3D display, while the imagery that the observer theoretically needs to accommodate to is located at the display distance. Small discrepancies between these distances do not induce visual fatigue so that S3D scenes are usually limited to a zone of comfort, e.g., the 1° limit on retinal disparity [Lam09b].

4 Eye Tracking

The GACS3D-prototype, described in **section 5.2.2**, involves **eye tracking**, i.e., eye-gaze tracking in a **stereo 3D (S3D)** consumer or work environment. The basics of eye tracking and problems that arise with it are outlined in the first two sections of this chapter. Afterwards, 2D gaze filtering methods are described in **section 4.3** and a Kalman filter approach with a new kind of event detection is proposed, that is used in GACS3D. GACS3D reduces the **accommodation vergence discrepancy (AVD)** by zeroing the disparity at the visual focus using **dynamic horizontal image translation (DHIT)**. Hence, a visual focus disparity estimation is necessary. Since disparity is linked to depth via **equation (3.7)**, a generalization of this procedure is the 3D visual focus estimation. In **section 4.4**, respective approaches are reviewed and a new approach for visual focus disparity estimation is proposed, that is used in the full GACS3D prototype.

4.1 Types of Eye Trackers

The first eye trackers were invasive approaches in a way that electrodes had to be attached to the face of the subject. A contact lens with a small wire loop has been used in the 1950s in conjunction with measuring the current induced by movement of the eye through magnetic fields [AN10, Ham08, Hol11]. These *electromagnetic coil systems* deliver the highest accuracy, but are known to alter saccades of a subject [Hol11]. Another approach is the 1970s *electro-oculogram* where electrodes pick up small changes of electrical potential generated by eye movements [AN10, Ham08, Hol11]. This type of eye tracker is very cheap, but comes with a degraded accuracy and suffers from electromagnetic noise from surrounding muscles [Hol11]. Both of these invasive approaches have the advantage that they are independent of lighting conditions and the state of the eyes or eye lids. However, wearing a device like that is not comfortable, which makes them unsuitable for discomfort research and consumer applications.

Nowadays, video-based eye trackers are mostly used. These devices utilize a camera directed at the subject to estimate the gaze direction. These eye trackers can be classified as *static* or *head-mounted* [Hol11]. The latter kind estimate gaze direction relative to the orientation of the wearable device, i.e., the head orientation. The fixed position of the cameras, relative to the eyes, is advantageous as far as accuracy is concerned. In order to convert the relative gaze data to absolute gaze data in 3D space, these devices are sometimes augmented with an additional *head tracker* to estimate the absolute position and orientation of the head.

Static eye trackers are placed in front of the subject and can be further divided into *tower-mounted* and *remote* devices [Hol11]. Tower-mounted devices establish a fixed spatial offset between the eyes and the cameras by fixating the subject's head using a bite-bar or forehead and chin rest. Remote eye trackers on the other hand discard the head fixation and allow for free subject movement in a limited range called *head box*. The increase in mobility and subject comfort is traded off with a decrease in accuracy and higher *recovery times*. The recovery time represents the time it takes to resume tracking after, e.g., prolonged blinks. These disadvantages can be mitigated by improved head tracking for example through infrared reflectors, however [Hol11]. For discomfort research and consumer applications, only remote eye trackers are an option because they do not induce discomfort by head fixation or by having to wear a device which is why the rest of this chapter focuses on remote eye tracking.

4.2 Principles and Properties of Remote Eye Tracking

Most remote eye trackers utilize one or two cameras directed at the subject's eyes in conjunction with some LEDs. The LEDs and the camera parameters are fully calibrated, so that the positions of the LED reflections on the cornea can be used to estimate the 3D coordinates of the eye ball center. This estimation is based on an eye ball model whose parameters have to be set individually by a calibration routine prior to operation. Furthermore, the pupil is detected in the eye image. The 3D coordinates of its center and the eye ball center specify the gaze direction. The constructed eye gaze vector is intersected with the calibrated display plane to obtain the 2D gaze coordinates.

The **accuracy** and **precision** of gaze data are degraded by noise, imperfect calibration and further issues [Hol11]. Accuracy denotes the **mean absolute error (MAE)** between the true gaze position and the measured gaze samples. A degradation is mostly caused by noise in the pupil or corneal reflections, or an eye ball model mismatch. Some external factors like varying pupil sizes, pupil color, heavy eye makeup and head movements are also problematic [Hol11]. Precision denotes the spread of the measured gaze samples, i.e., the ability of a device to reliably reproduce measurements. Precision is predominantly determined by the eye camera resolution and sensor noise level, and it can be traded off against sampling frequency, since a higher frequency increases noise. Precision is mostly evaluated via the standard deviation and measured using an artificial eye [Hol11, Tob11]. In this way, only the eye tracker inherent **system noise** is evaluated, while ensuring the exclusion of individually varying **oculomotor noise**, i.e., involuntary, unconscious eye movements. In practice, this leads to much worse precision values during actual operation compared to what the manufacturer measured under ideal conditions. Also, precision can vary depending on the artificial eye used [Hol11]. This means that precision values of different manufactures cannot be compared easily.

4.3 2D Gaze Filtering and Event Detection

The gaze data delivered by an eye tracker is affected by noise. Hence, filtering techniques are used to improve accuracy and precision. It should be noted that, while precision can basically be improved immensely using these filters, accuracy will eventually degrade. There are real-time and post-processing approaches. The prior are required for *gaze-directed human-machine-interaction*, which is the context of this chapter. Amongst the most widespread approaches are the Butterworth filter [Duc11, Wan14] and the Kalman filter [AA02, Koh09a, Kom07, Koh09b]. Špakov has done a comparison study with numerous approaches and has found out that all perform equally bad with respect to accuracy [Špa12], while noting that the specific Kalman filter [Kom07] they tested performed the worst. A comparatively small study has also been conducted for this work by Wermers in his supervised master's thesis [Wer16]. His study has revealed that the Kalman filter derived in the next section actually performed best. However, accuracy was only marginally improved compared to the raw gaze data. The conclusion is that the choice of an approach is predominantly determined by the required features and properties rather than the (equally bad) filter accuracy.

Kalman filtering [Kal60] is frequently used in 2D gaze filtering because of some very useful properties. A Kalman filter estimates the state of a process based on a defined statistical model and minimizes the **mean squared error (MSE)** between the estimation and the true state [Wel06]. During the estimation process, the gaze movement is predicted and can be used as filter output in case of a missing measurement or an outlier, which is a common issue in gaze filtering, e.g., due to blinks. The Kalman filter is furthermore recursive since it only depends on the previous state and the most recent measurement. This makes it applicable for real-time applications. Depending on the used statistical model, different kinds of eye movements can be passed through or removed and eye movement events or outliers can be detected [Koh09b]. These properties make the Kalman filter a good choice for gaze-directed human-machine-interaction.

4.3.1 Kalman Filter

In this section, the general concept of the Kalman filter is described. Afterwards, an approach tailored to gaze-directed human-machine-interaction is derived, that can be applied to each coordinate of the gaze samples separately.

4.3.1.1 General Filter approach

4.3.1.1.1 Model

A Kalman filter utilizes two statistical models [Wel06]: one for the process state[1]

$$\boldsymbol{s}_k = \mathcal{A} \cdot \boldsymbol{s}_{k-1} + \boldsymbol{w}_{k-1} \tag{4.1}$$

and one for the measurement

$$\boldsymbol{m}_k = \mathcal{H} \cdot \boldsymbol{s}_k + \boldsymbol{n}_k \ . \tag{4.2}$$

Here, $\mathcal{A}$ is the state transition matrix, which predicts the process state $\boldsymbol{s}_k$ at time instance k from the previous state $\boldsymbol{s}_{k-1}$. The vector $\boldsymbol{w}_{k-1}$ is the additive white gaussian process noise with covariance matrix $\mathcal{Q}$. The statistical variable $\boldsymbol{w}_{k-1}$ is used to model how much the state may change between two consecutive iterations in addition to the prediction induced change. The observation model matrix $\mathcal{H}$ transforms the process state into a measurement $\boldsymbol{m}_k$, which is again affected by an additive white gaussian noise vector $\boldsymbol{n}_k$. This noise vector models the measurement noise and is statistically independent of $\boldsymbol{w}_k$. It has a covariance matrix $\mathcal{R}$.

4.3.1.1.2 Concept

The true process state can never be known, but it can be estimated *a priori*, i.e., without knowledge of the current measurement, and *a posteriori*. In this work, all estimated variables are notated with a hat symbol ($\hat{}$) and *a priori* variables with a superscript minus ($^-$), whereas no superscript is used for *a posteriori* variables. The *a posteriori* state estimate

$$\hat{\boldsymbol{s}}_k = \hat{\boldsymbol{s}}_k^- + \mathcal{K}_k \cdot (\underbrace{\boldsymbol{m}_k - \mathcal{H} \cdot \hat{\boldsymbol{s}}_k^-}_{\boldsymbol{i}_k}) \tag{4.3}$$

is a linear combination of the *a priori* state estimate $\hat{\boldsymbol{s}}_k^-$ and the difference between the measurement $\boldsymbol{m}_k$ and the measurement prediction $\mathcal{H} \cdot \hat{\boldsymbol{s}}_k^-$ [Wel06], which is called **innovation** $\boldsymbol{i}_k$. Here, $\mathcal{K}_k$ is the *Kalman gain*. It minimizes the *a posteriori* variance of error $\mathrm{E}\left\{\|\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k\|^2\right\}$, i.e., the MSE between the true process state and the *a posteriori* process state estimate. This MSE is the

[1] The process state model actually also includes an optional control input vector [Wel06]. However, this vector is not utilized in this work, which is why it is not described any further.

trace of the (unknown) *a posteriori* estimation error covariance matrix

$$\mathcal{P}_k = \mathrm{E}\left\{(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k) \cdot (\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k)^\top\right\} \ . \tag{4.4}$$

The solution for the Kalman gain is given by [Wel06]

$$\mathcal{K}_k = \mathcal{P}_k^- \mathcal{H}^\top \left(\mathcal{H}\mathcal{P}_k^- \mathcal{H}^\top + \mathcal{R}\right)^{-1} \ , \tag{4.5}$$

where

$$\mathcal{P}_k^- = \mathrm{E}\left\{\left(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k^-\right) \cdot \left(\boldsymbol{s}_k - \hat{\boldsymbol{s}}_k^-\right)^\top\right\} \tag{4.6}$$

is the (unknown) *a priori* estimation error covariance matrix. For the interpretation of $\mathcal{K}_k$, it is helpful to consider the limiting values of $\mathcal{R}$ and $\mathcal{P}_k^-$ [Wel06]. For a comparatively small measurement noise covariance $\mathcal{R}$, the measurements are trusted more than the prediction, since

$$\begin{aligned}\lim_{\mathcal{R}\to 0} \mathcal{K}_k &= \mathcal{P}_k^- \mathcal{H}^\top \left(\mathcal{H}\mathcal{P}_k^- \mathcal{H}^\top\right)^{-1} = \mathcal{P}_k^- \mathcal{H}^\top \left(\mathcal{H}^\top\right)^{-1} \left(\mathcal{P}_k^-\right)^{-1} \mathcal{H}^{-1} \\ &= \mathcal{H}^{-1} \ ,\end{aligned} \tag{4.7}$$

which using **equation (4.3)** leads to

$$\lim_{\mathcal{R}\to 0} \hat{\boldsymbol{s}}_k = \hat{\boldsymbol{s}}_k^- + \mathcal{H}^{-1} \cdot \left(\boldsymbol{m}_k - \mathcal{H} \cdot \hat{\boldsymbol{s}}_k^-\right) = \mathcal{H}^{-1} \boldsymbol{m}_k \ . \tag{4.8}$$

Conversely, a small *a priori* estimation error covariance $\mathcal{P}_k^-$ means, that the prediction is trusted more, since

$$\lim_{\mathcal{P}_k^- \to 0} \mathcal{K}_k = 0 \ , \tag{4.9}$$

and therefore

$$\lim_{\mathcal{P}_k^- \to 0} \hat{\boldsymbol{s}}_k = \hat{\boldsymbol{s}}_k^- \ . \tag{4.10}$$

4.3.1.1.3 Algorithm

The recursive Kalman filter algorithm consists of two steps, forming a *predictor-corrector-loop*. In the *time update* step, the previous *a posteriori* estimates are mapped forward in time (prediction) to obtain new *a priori* estimates. In the *measurement update* step, these predictions are corrected by the new measurement to obtain the new *a posteriori* estimates that yield the minimum MSE. Not only the process state but also the *a priori* and *a posteriori* estimation error covariance matrices $\hat{\mathcal{P}}_k^-$ and $\hat{\mathcal{P}}_k$, respectively, need to be estimated because they are unknown and needed for the computation of the Kalman gain.

1) Time update:

$$\hat{\boldsymbol{s}}_k^- = \mathcal{A} \cdot \hat{\boldsymbol{s}}_{k-1} \tag{4.11}$$

$$\hat{\mathcal{P}}_k^- = \mathcal{A}\hat{\mathcal{P}}_{k-1}\mathcal{A}^\top + \mathcal{Q} \tag{4.12}$$

2) Measurement update:

$$\mathcal{K}_k = \hat{\mathcal{P}}_k^- \mathcal{H}^\top \mathcal{S}_k^{-1} \tag{4.13}$$

$$\hat{\boldsymbol{s}}_k = \hat{\boldsymbol{s}}_k^- + \mathcal{K}_k \cdot \left(\boldsymbol{m}_k - \mathcal{H} \cdot \hat{\boldsymbol{s}}_k^-\right) \tag{4.14}$$

$$\hat{\mathcal{P}}_k = (\mathcal{I} - \mathcal{K}_k\mathcal{H})\,\hat{\mathcal{P}}_k^- \tag{4.15}$$

Here, $\mathcal{I}$ is the identity matrix, and

$$\mathcal{S}_k = \mathcal{H}\hat{\mathcal{P}}_k^-\mathcal{H}^\top + \mathcal{R} \tag{4.16}$$

is the innovation covariance matrix.

4.3.1.1.4 Parametrization and Initialization

After choosing a statistical model, the performance of the Kalman filter can only be altered through the process noise covariance matrix $\mathcal{Q}$ and the measurement noise covariance matrix $\mathcal{R}$. As already shown in **equations (4.7) and (4.8)**, a small $\mathcal{R}$ leads to measurements being trusted more. That is, unless $\mathcal{Q}$ is even smaller, which, according to **equations (4.12), (4.13) and (4.15)**, leads to a small *a priori* estimation error covariance $\hat{\mathcal{P}}_k^-$ and, therefore, a suppression of measurements, see **equations (4.9) and (4.10)**. In other words, if the process state is allowed to change a lot while the measurement noise is comparatively low, the measurements are very trustworthy. On the other hand, if the measurement noise is high, relative to the allowed change of the process state, the measurement noise is suppressed, giving the filter a low pass character. The measurement noise can simply be measured prior to filter operation. Determining the process noise covariance is not as easy, however, since the process cannot be directly observed [Wel06]. Usually, $\mathcal{Q}$ is coarsely approximated and then heuristically tuned until the results are as desired.

The filter finally needs to be initialized. While the process state could theoretically be initialized to anything, the filter will converge faster if it is initialized using the first samples, e.g., $\hat{s}_0 = \mathcal{H}^{-1} m_0$. An initial $\hat{\mathcal{P}}_0 \neq 0$, e.g., $\hat{\mathcal{P}}_0 = \mathcal{I}$, will also lead to faster convergence. However, if $\mathcal{Q}$ and $\mathcal{R}$ are constant, so will $\hat{\mathcal{P}}_k$ and $\mathcal{K}_k$ be after a few iterations of convergence [Wel06]. Hence, the converged matrices can be used as constants after estimating them by applying the filter to training data.

4.3.1.2 Application Specific Filter Design

4.3.1.2.1 Model

The choice of a statistical model depends on the kind of application and the statistical properties of the measurement noise. In eye movement research, all kinds of eye movements have to be modeled so that only the system noise of the eye tracker is removed by the filter. In *human-machine-interaction* only fixations, smooth pursuit, saccades and, in the case of S3D, vergence movements are usually of interest. Here, the other eye movements are considered oculomotor noise that should be removed in conjunction with the system noise. This can be achieved by the model described in this section.

The state of a moving particle in one dimensional space, e.g., a gaze coordinate, can best be described by its position p and speed v, which may be altered through acceleration a via these well-known differential equations

$$a = \frac{\partial v}{\partial t} = \frac{\partial^2 p}{\partial^2 t} \; . \tag{4.17}$$

Applying this concept in a time-discrete manner to the process state model in

equation (4.1) yields

$$\begin{bmatrix} p_k \\ v_k \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}}_{\mathcal{A}} \cdot \begin{bmatrix} p_{k-1} \\ v_{k-1} \end{bmatrix} + \underbrace{\begin{bmatrix} \frac{1}{2}T^2 \\ T \end{bmatrix} \cdot a_{k-1}}_{\boldsymbol{w}_{k-1}} \ , \tag{4.18}$$

where T denotes the sampling interval, and the acceleration a_{k-1} is the statistical variable controlling the change of the process state over time. The process noise covariance matrix is then given by

$$\begin{aligned} \mathcal{Q} &= \mathrm{E}\left\{\boldsymbol{w}_k \cdot \boldsymbol{w}_k^\top\right\} \\ &= \begin{bmatrix} \frac{1}{2}T^2 \\ T \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2}T^2 & T \end{bmatrix} \cdot \mathrm{E}\left\{a_k \cdot a_k^\top\right\} \\ &= \begin{bmatrix} \frac{1}{4}T^4 & \frac{1}{2}T^3 \\ \frac{1}{2}T^3 & T^2 \end{bmatrix} \cdot \sigma_a^2 \ . \end{aligned} \tag{4.19}$$

This is the widespread *constant-velocity model* [BS04], commonly used in gaze filtering and many other applications. Since the eye tracker samples only consist of gaze positions and no speeds, the measurement equation, i.e., **equation (4.2)**, for this state vector is one-dimensional and given by

$$m_k = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\mathcal{H}} \begin{bmatrix} p_k \\ v_k \end{bmatrix} + n_k \ . \tag{4.20}$$

This means that the measurement noise is one-dimensional, too, so that its covariance matrix is given by

$$\mathcal{R} = \sigma_n^2 \ . \tag{4.21}$$

As mentioned before, the measurement noise here is a combination of the eye tracker system noise and oculomotor noise, which means that the requirement of a gaussian distribution does not necessarily hold. Furthermore, the whiteness of the process noise a_k is questionable. While optimal results are only ensured if all constraints are met, the Kalman filter is known to produce acceptable results for slight violations if enough uncertainty is injected into the process model via $\mathcal{Q}$ [Wel06].

4.3.1.2.2 Parametrization

The measurement noise variance σ_n^2 can be estimated by having a subject look at a target on the screen and calculating the variance of the resulting gaze samples around that known target position. In this way, eye tracker system noise and most of the oculomotor noise are automatically combined. An eye tracker might yield 1° sample spread radius during operation. Assuming normal distribution, 95 % of the samples are contained within $\pm 2\sigma_n$, so that σ_n could be approximated by 0.5°. At a viewing distance $d = 3.1H$, using a Full HD display, this angle corresponds to $\sigma_n^2 = 853.63\,\text{px}^2$.

The process noise variance σ_a^2 should be chosen in accordance with the respective eye movement properties in **table 2.1**. A big σ_a^2 would include the fast saccades in the model. However, as mentioned before, a relatively big $\mathcal{Q} \sim \sigma_a^2$ allows more noise to pass the filter. Kohlbecher and Schneider solve this issue by using three Kalman filters in parallel: a constant-position model for fixations and constant-velocity models with small and big process variance for slow eye movements and saccades, respectively [Koh09b]. The final filter output is taken from the model that fits the data best. This approach mainly serves as an eye movement classification, though. Utilizing it as a filter yields discontinuities due to the filter swapping operation, which might be unwanted behavior. A common approach without such discontinuities is to model only the slow vergence movements and smooth pursuit, and have the Kalman filter reinitialize on saccades, where the discontinuities are wanted. To do so, saccades have to be detected, which can be done as follows in the next section. In the experiments presented in this work, the measurement noise variance above and a comparatively low process noise variance of $\sigma_a^2 = 3.43 \cdot 10^5\,\text{px}^2/\text{s}^4$ produced good results.

4.3.1.2.3 Event Detection and Processing

Any violation of the above model is to be considered either an outlier or a saccade with the difference that multiple connected outliers occur in case of a saccade. A model violation is given if the total squared innovation $\boldsymbol{i}_k^\top \boldsymbol{i}_k$ is significantly bigger than its expected value as of the innovation covariance matrix $\mathcal{S}_k$. The respective normalized innovation squared

$$e_k^2 = \boldsymbol{i}_k^\top \mathcal{S}_k^{-1} \boldsymbol{i}_k \tag{4.22}$$

varies as a χ^2-distribution with $\dim(\boldsymbol{m}_k)$ degrees of freedom [BS04]. Hence, the model is valid as long as

$$e_k^2 \leq e_{k,\beta}^2 \ , \tag{4.23}$$

where $e_{k,\beta}^2$ is the critical value of the χ^2-distribution at the chosen confidence level β.

This concept can be also extended to check model validity for an analysis window bigger than one measurement [BS04], which is needed in order to distinguish saccades from regular outliers. However, eye trackers may occasionally deliver samples marked as invalid, e.g., due to brief tracking loss, so that no innovation can be computed. Because of that, a more pragmatic approach has been developed in this work. Each valid measurement, that does not satisfy **equation (4.23)**, is classified as a violation, and the number of connected violations n_V is counted. As soon as a measurement is not classified as violation, the counter is reset to zero. If the number of connected violations is above a certain threshold n_th, a saccade has been detected. Since invalid samples cannot be classified, they are processed as "Don't Cares" in the following way. Instead of just one counter, an upper bound counter $n_\mathrm{V,UB}$ and a lower bound counter $n_\mathrm{V,LB}$ is used. For any invalid sample interleaved in a series of violations, the violation counters are not reset to zero. Instead, $n_\mathrm{V,UB}$ is increased by 1, whereas $n_\mathrm{V,LB}$ is left "as is". This means that the true but unknown number of connected violations suffices $n_\mathrm{V} \in [n_\mathrm{V,LB}, n_\mathrm{V,UB}]$. Considering that the higher the number of invalid samples n_inv in the current series of connected violations, the less confident one can be about correctly identifying a saccade, the final violation count is estimated to

$$\hat{n}_\mathrm{V} = \begin{cases} n_\mathrm{V,LB} + \left(n_\mathrm{V,UB} - n_\mathrm{V,LB}\right)/n_\mathrm{inv} & \text{for } n_\mathrm{inv} > 0 \\ n_\mathrm{V,LB} & \text{else.} \end{cases} \tag{4.24}$$

This value is compared with the aforementioned threshold n_th.

The choice of n_th is a trade-off between robustness and filter delay. A comparatively big value will delay the saccade detection and, with that, the filter output. On the other hand, a too low n_{th} might reset the filter too early on a glissade accompanying the saccade. Glissades are to be considered oculomotor noise that is supposed to be removed completely. Hence, $n_{\mathrm{th}} \cdot T$ should be bigger than the saccade duration as of **table 2.1**. After a saccade has been detected, the filter may be reinitialized using multiple previous measurements if they appear to be stable.

Invalid measurements also affect the filter procedure in other ways than just the saccade detection. Since the innovation $\boldsymbol{i}_k$ cannot be computed for invalid measurements, the process state is just projected ahead in time, whereas the measurement update in **equations (4.13) to (4.15)** must be skipped. All outliers are furthermore processed in the same way because they affect filter performance, and therefore also saccade detection, negatively.

In the experiments presented in this work, $e^2_{k,\beta} = 5\,\mathrm{px}^2$ and $n_{\mathrm{th}} = 13$ produced good results.

4.4 3D Visual Focus Estimation

A very basic approach for 3D visual focus estimation is the geometric method, i.e., triangulation [Ess06, Pfe08, Wan14, Wib14]. The visual axes of the eyes are estimated based on the gaze data delivered by the eye tracker. The point where the visual axes intersect represents the 3D visual focus. However, the estimated visual axes possibly do not intersect due to noise so that the point nearest to both visual axes is used, or the vertical coordinates are equalized. In the latter case, the visual axes do intersect so that this approach is equivalent to estimating disparity directly from gaze data, and transforming it to depth using **equation (3.7)** [Duc14]. The disparity is estimated simply by the difference between the horizontal gaze coordinates of the two eyes using **equation (3.1)**. As an example, the results of this approach, denoted as “Raw”, are displayed in **figure 4.1**. A subject looking at a test image was asked to track a target pointer that was placed on a series of locations exhibiting different stereoscopic disparities[2]. As can be seen, the raw disparity differs strongly from the target

[2] The Full HD passive S3D display was located at a distance of $d = 3.1 \cdot H \approx 180\,\mathrm{cm}$ away from the subject.

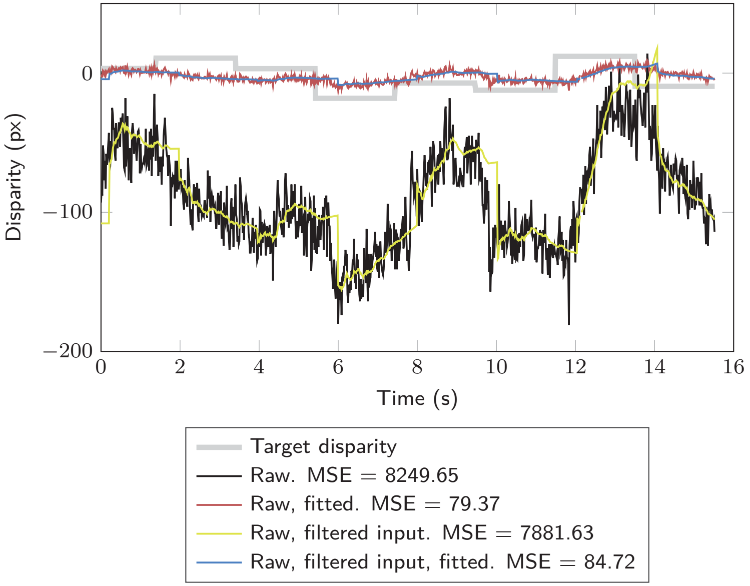


Figure 4.1: Disparity directly computed from gaze data of a subject. All MSE values in px^2.

disparity. This can be attributed to the fact that the system noise, and possibly also the oculomotor noise, is uncorrelated between both eyes. However, in the latter case, there are contradictory publications [Rol09]. Applying the Kalman filter described in **section 4.3.1** to the input, prior to the disparity calculation does not significantly improve the results, see **figure 4.1**. This shows again that the filter does not improve gaze data accuracy very much.

Because of these inaccuracies, a 3D calibration is usually performed, which is essentially a mapping operation typically involving first or second order polynomials for each spatial dimension [Duc14, Wan14, Wib14]. The mapping coefficients are calculated by minimizing the absolute differences between subject gaze data and some reference 3D calibration points using the Moore-Penrose-inverse. For a first order polynomial, the depth calibration equation for N data points is given by

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_1 \\ 1 & \hat{z}_2 \\ \vdots & \vdots \\ 1 & \hat{z}_N \end{bmatrix}}_{\mathcal{M}} \cdot \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} , \tag{4.25}$$

where z_i and $\hat{z}_i$ are the reference and estimated depth[3], respectively, of data point number i, and $\begin{bmatrix} c_0 & c_1 \end{bmatrix}^\top$ are the polynomial coefficients with the solution

$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \left(\mathcal{M}^\top \mathcal{M}\right)^{-1} \mathcal{M}^\top \boldsymbol{z} . \tag{4.26}$$

While tracking, every estimated depth value is corrected using these polynomial coefficients. This 3D calibration has been applied retroactively in **figure 4.1**, denoted as “Raw, fitted”, where a first order polynomial has been fitted to the “Raw” disparity sequence rather than to an actual calibration sequence. Hence, the graph represents the theoretical best-case calibration, which cannot be achieved in practice, especially not without a head rest because a recorded 3D calibration becomes inaccurate when a subject’s head moves [Hol11]. The MSE of the calibrated disparity sequence improves a lot and a slight correlation to the target disparity can be observed, but the results are still very inaccurate. Instead of applying the mapping operation to each dimension of the estimated 3D visual focus, the 2D gaze data can also be mapped directly into the 3D space, e.g., using the parametrized self-organizing maps approach [Ess06] or a support vector regression model [Toy14]. According to the authors, this does mildly improve accuracy. However, head movement still poses an issue in the same way as before.

While the previous approaches process the gaze coordinates delivered by the eye tracker in order to estimate depth, it can also be estimated directly in the eye tracker from different cues. When fixating on a stimulus, the vergence angle of the eyes is adjusted and it can be estimated via the distance between

[3] Depth is estimated based on the gaze data.

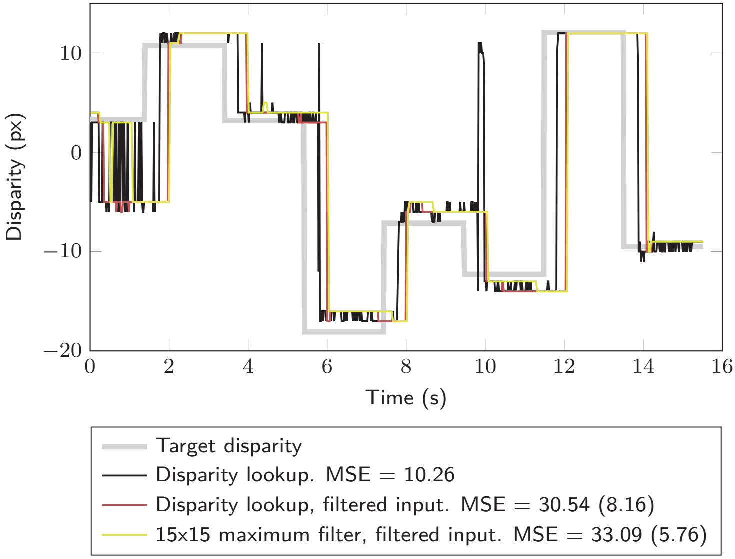


Figure 4.2: Basic utilization of a disparity map. All MSE values in px^2.

the pupil centers [Ki07, Alt14]. However, since that distance varies individually, a calibration procedure recording pupil distances at different reference depths is mandatory. As mentioned in **section 2.2.3**, the pupil size is also dependent on depth and can therefore be analyzed [Alt14, Lee12]. Again, a calibration is needed for this approach. Finally, there are multiple reflections of the eye tracker LEDs observable in the eye. They are called *Purkinje images*. The positions of the first and fourth Purkinje images vary with the accommodation state of the eyes and are therefore linked to fixation depth. Lee et al. combine the analysis of Purkinje images and pupil size to obtain a depth estimate [Lee12]. However, accommodation might be an unreliable cue in the context of S3D displays, see **section 3.2.3.1**.

The accuracy of all the approaches mentioned so far is approximately of the same order of magnitude, which is not sufficient for the application in this work. If gaze depth or disparity is supposed to be estimated in the context of S3D, an important source of information neglected by these approaches is the actual 3D structure of the presented scene. It can be represented by a disparity or depth map. A simple disparity lookup at the gaze coordinates already yields an improvement of the MSE by almost an order of magnitude, see **figure 4.2** compared to **figure 4.1**. As can be seen, there is a certain delay between the target pointer disparity and the straight lookup graph. It is due to the limited saccadic reaction time of about 240 ms [Joo03] as well as the saccade and vergence movement duration, which means that the target pointer disparity cannot be used as ground truth eye disparity in the MSE computation straight away. The ground truth is formed by delaying the target pointer disparity by 18 samples, i.e., 300 ms, and ignoring 6 samples after each saccade to account for the variability in human reaction delay[4]. A straight lookup at Kalman filtered gaze coordinates is displayed in **figure 4.2** as well. As already mentioned in **section 4.3.1**, there is an additional delay because of the saccade detection, which decreases the MSE a lot. In certain applications an additional delay is not harmful and only the stable end result is of importance. Because of that, another MSE, where the target pointer disparity is delayed by 31 samples[5] rather than just 18, has been computed. This value shall be called *delay-adjusted MSE* henceforth and it is displayed in brackets after the original MSE wherever applicable. However, there is still only little improvement over the original unfiltered result, which contains big errors at about 10 s. This shows again that the Kalman filter does not necessarily improve accuracy. Hanhart and Ebrahimi apply a 15 by 15 maximum filter to each disparity map at the respective filtered gaze coordinates. This filtering approach is based on the assumption that foreground objects are more likely to be looked at [Han14]. Afterwards, a fourth order moving median filter is applied to the sequence of disparities. The results of this technique are displayed in **figure 4.2** as well and in this example the delay-adjusted MSE shows a slight improvement[6]. However, a problem with this approach is that the analysis window is a lot smaller than the spread of samples due to system noise and oculomotor noise. This means that the true gaze coordinates may not be contained in that window and therefore neither the correct disparity. Furthermore, the approach relies on precomputed disparity maps and does not handle erroneous maps well. In this work, a new, high accuracy approach is proposed, which leverages these

[4] This has been done in **figure 4.1** as well.

[5] The delay is increased by 13 samples compared to the original MSE computation. This is exactly the minimum delay introduced by the saccade detection of the Kalman filter, see **section 4.3.1.2**.

[6] Please note that Hanhart and Ebrahimi applied a moving median filter to the 2D gaze data, while a Kalman filter has been used here for its believed superior performance.

problems. It is described in the next section.

4.4.1 Probabilistic Visual Focus Disparity Estimation

The developed approach follows the same basic idea as the one by Hanhart and Ebrahimi [Han14] in a way that a disparity map is utilized to find the disparity of the visual focus. In contrast to the referenced approach, the disparity map is estimated locally for a **region of interest (ROI)** in real-time using parallel computing on a **graphics processing unit (GPU)** in **CUDA C++**. The estimation is carried out from left to right view and from right to left view, denoted as $\hat{D}_{\mathrm{L}}$ and $\hat{D}_{\mathrm{R}}$, respectively. Furthermore, a weight is assigned to each candidate pixel in this ROI. The weight represents the confidence of identifying the correct disparity. Because of that, the approach is called **probabilistic visual focus disparity estimation (PVFDE)**. In order to calculate the weights, certain restrictions and assumptions are defined for the candidates. The appropriate candidate disparity is finally chosen using a weighted histogram filter. The processing steps are illustrated in **figure 4.3** and are described in more detail in the following sections, after a brief description of the utilized gaze data model.

4.4.1.1 Gaze Data Model

The eye tracker delivers 2D gaze data $\boldsymbol{p}_{\mathrm{G,L}}$ and $\boldsymbol{p}_{\mathrm{G,R}}$ for the left and right eye, respectively. The gaze data is affected by additive noise, which can be modeled for the left and right eye by

$$\boldsymbol{p}_{\mathrm{G,L|R}} = \boldsymbol{p}_{\mathrm{T,L|R}} + \boldsymbol{p}_{\mathrm{N,L|R}} \ , \tag{4.27}$$

where $\boldsymbol{p}_{\mathrm{T,L|R}}$ is the true gaze position in the display plane and $\boldsymbol{p}_{\mathrm{N,L|R}}$ is the additive noise term. In this application, only fixations, saccades, smooth pursuit and vergence movements are of interest. Hence, the noise term is a combination of system noise and oculomotor noise. The distribution of noise is approximated to be normal.

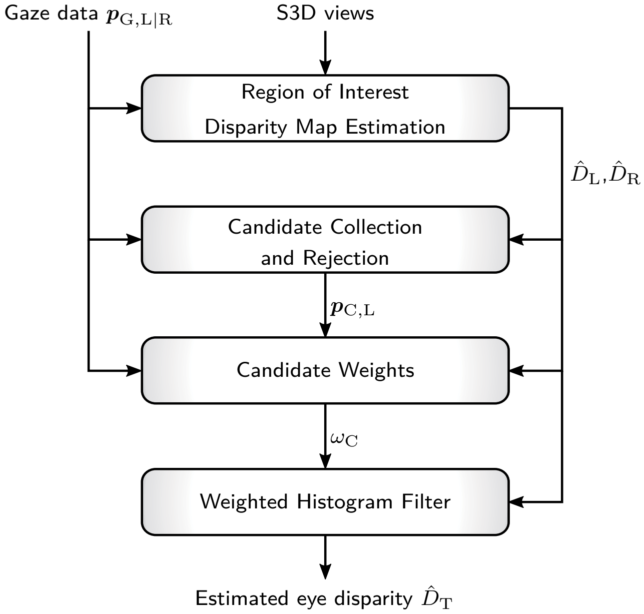

Gaze data $\boldsymbol{p}_{\mathrm{G,L|R}}$
S3D views
Region of Interest Disparity Map Estimation
$\hat{D}_{\mathrm{L}}$,$\hat{D}_{\mathrm{R}}$
Candidate Collection and Rejection
$\boldsymbol{p}_{\mathrm{C,L}}$
Candidate Weights
ω_{C}
Weighted Histogram Filter
Estimated eye disparity $\hat{D}_{\mathrm{T}}$

Figure 4.3: Processing steps of the PVFDE.

The aim is to find an estimate $\hat{D}_{\mathrm{T}}$ of the true disparity D_{T}

$$D_{\mathrm{T}} = D_{\mathrm{L}}\left(\boldsymbol{p}_{\mathrm{T,L}}\right) = D_{\mathrm{R}}\left(\boldsymbol{p}_{\mathrm{T,R}}\right) \tag{4.28}$$

at the visual focus, where D_{L} and D_{R} are error free disparity maps.

4.4.1.2 Region of Interest Disparity Map Estimation

In this first step, block matching is used to estimate the left and right disparity maps in real-time for square ROIs. These ROIs are centered on the left and right gaze coordinates, respectively. Generally, the block matching algorithm estimates correspondency vectors between temporally or spatially neighboring views of a scene. This is done by matching a reference block, typically of size

9 px by 9 px, against a spatially offset block in the neighboring view using some similarity criterion. Multiple offsets are tried in an iterative manner following a certain search strategy. The offset that maximizes block similarity is the estimated correspondency vector.

Numerous search strategies exist for different applications. In this case, since the S3D views are assumed to be rectified, disparity is purely horizontal. Hence, the search direction is horizontal only. All integer offsets, i.e., disparities in a configurable range are tested using the **sum of absolute differences (SAD)** similarity criterion. This non-conditional search strategy has the advantage that it can be parallelized well, which is mandatory for real-time performance. While there are other more sophisticated and robust similarity criteria, the SAD is used because of its computational simplicity. The decrease in robustness is mitigated by the confidence calculations in the subsequent steps of this algorithm, which are described in the next sections. Using **equation (3.1)**, for the estimation from the left view image I_L to the right view image I_R, the purely horizontal SAD of a block $B\left(\boldsymbol{p}_i\right)$, which is centered on the 2D image space coordinate $\boldsymbol{p}_i$, is given by

$$\mathrm{SAD}_{\mathrm{L}\rightarrow\mathrm{R}}\left(\boldsymbol{p}_i, D\right) = \sum_{\boldsymbol{p}\in B(\boldsymbol{p}_i)} \left|I_\mathrm{L}\left(\boldsymbol{p}\right) - I_\mathrm{R}\left(\boldsymbol{p} - D\cdot\boldsymbol{e}_x\right)\right| \ , \tag{4.29}$$

where $\boldsymbol{e}_x$ is the unity vector in x-direction. Conversely, the SAD for the estimation from right to left is

$$\mathrm{SAD}_{\mathrm{R}\rightarrow\mathrm{L}}\left(\boldsymbol{p}_i, D\right) = \sum_{\boldsymbol{p}\in B(\boldsymbol{p}_i)} \left|I_\mathrm{R}\left(\boldsymbol{p}\right) - I_\mathrm{L}\left(\boldsymbol{p} + D\cdot\boldsymbol{e}_x\right)\right| \ . \tag{4.30}$$

Finally, the argument D that minimizes the SAD is the estimated disparity

$$\hat{D}_{\mathrm{L}|\mathrm{R}}\left(\boldsymbol{p}_i\right) = \underset{D}{\operatorname{argmin}}\left(\mathrm{SAD}_{\mathrm{L}\rightarrow\mathrm{R}|\mathrm{R}\rightarrow\mathrm{L}}\left(\boldsymbol{p}_i, D\right)\right) \ . \tag{4.31}$$

This calculation is carried out for every point $\boldsymbol{p}_i$ inside the left or right square ROI. The size of the square ROI is $2r$ by $2r$, where r is the radius of the circular ROI defined in the next section. In the case of a top-and-bottom S3D format, which is commonly used in conjunction with passive 3D TVs, the S3D views are vertically subsampled. Hence, the vertical extend of the square ROI can be halved, making it rectangular.

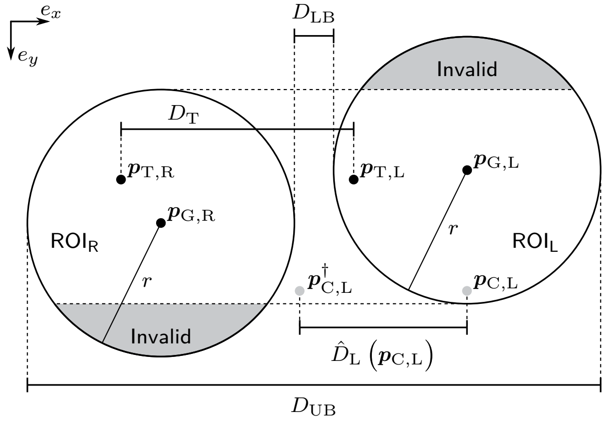

e_x
e_y
D_{LB}
Invalid
D_{T}
$\boldsymbol{p}_{\mathrm{T,R}}$
$\boldsymbol{p}_{\mathrm{T,L}}$
$\boldsymbol{p}_{\mathrm{G,L}}$
$\boldsymbol{p}_{\mathrm{G,R}}$
ROI$_{\mathrm{R}}$
r
r
ROI$_{\mathrm{L}}$
$\boldsymbol{p}^{\dagger}_{\mathrm{C,L}}$
$\boldsymbol{p}_{\mathrm{C,L}}$
Invalid
$\hat{D}_{\mathrm{L}}\left(\boldsymbol{p}_{\mathrm{C,L}}\right)$
D_{UB}

Figure 4.4: Left and right region of interest with true gaze positions $\boldsymbol{p}_{\mathrm{T,L|R}}$ and candidate rejection criteria: vertical rejection area in gray, valid disparity range $[D_{\mathrm{LB}}, D_{\mathrm{UB}}]$, a rejected candidate due to disparity mapping criterion in gray.

4.4.1.3 Candidate Collection and Rejection

As mentioned in **section 4.4.1.1**, the tracking noise is assumed to be normal. However, since extreme outliers will not contribute favorably to the disparity estimation and will be rejected anyway, a maximum spread radius r is introduced. It defines a circular area around the true gaze position $\boldsymbol{p}_{\mathrm{T,L|R}}$ in which all gaze samples, excluding outliers, can be located. Inverting this connection, any true gaze position must be located inside the maximum spread radius r around a valid (non-outlier) gaze position $\boldsymbol{p}_{\mathrm{G,L|R}}$. This is the ROI displayed in **figure 4.4**. Any left or right view candidate $\boldsymbol{p}_{\mathrm{C,L|R}}$ has to satisfy the ROI equation:

$$r_{\mathrm{C,L|R}} = \|\boldsymbol{p}_{\mathrm{C,L|R}} - \boldsymbol{p}_{\mathrm{G,L|R}}\| \leq r \ . \tag{4.32}$$

The maximum spread radius is dependent on eye tracker performance and the individual properties of the eyes.

Invalid candidates in the ROI can be quickly identified by four restrictions. Assuming rectified S3D views, the true gaze positions for the left and right eye must exhibit the same vertical position $y_{\mathrm{T,L}} = y_{\mathrm{T,R}}$. Areas where the

ROIs do not overlap vertically do not fulfill this restriction and the contained candidates are therefore invalid. These areas are marked gray in **figure 4.4**. Mathematically, a candidate $\boldsymbol{p}_{\mathrm{C,L|R}}$ is rejected, if its vertical position $y_{\mathrm{C,L|R}}$ does not satisfy

Restriction 1:

$$y_{\mathrm{C,L|R}} \in \left[\max\left(y_{\mathrm{G,L}}, y_{\mathrm{G,R}}\right) - r \ , \ \min\left(y_{\mathrm{G,L}}, y_{\mathrm{G,R}}\right) + r\right] \ , \tag{4.33}$$

where $y_{\mathrm{G,L}}$ and $y_{\mathrm{G,R}}$ are the left and right vertical gaze coordinates in pixels.

The second restriction is related to disparity. Any combination of candidates $\boldsymbol{p}_{\mathrm{C,L}}$ and $\boldsymbol{p}_{\mathrm{C,R}}$ in the left and right ROI yields a certain disparity via **equation (3.1)**. Since the positions of the candidates are restricted, so is their respective disparity. Candidates at the outermost positions of equally high ROIs yield an upper bound to the valid disparity

$$D_{\mathrm{UB}} = \left(x_{\mathrm{G,L}} + r\right) - \left(x_{\mathrm{G,R}} - r\right) = x_{\mathrm{G,L}} - x_{\mathrm{G,R}} + 2 \cdot r \ , \tag{4.34}$$

where $x_{\mathrm{G,L}}$ and $x_{\mathrm{G,R}}$ are the left and right horizontal noisy gaze coordinates in pixels as delivered by the eye tracker. Conversely, candidates at the innermost positions yield a lower bound

$$D_{\mathrm{LB}} = x_{\mathrm{G,L}} - x_{\mathrm{G,R}} - 2 \cdot r \ . \tag{4.35}$$

Any candidate $\boldsymbol{p}_{\mathrm{C,L|R}}$ that does not satisfy

Restriction 2:

$$\hat{D}_{\mathrm{L|R}}\left(\boldsymbol{p}_{\mathrm{C,L|R}}\right) \in \left[D_{\mathrm{LB}}, D_{\mathrm{UB}}\right] \ , \tag{4.36}$$

i.e., whose associated disparity is not contained in this disparity range, is rejected.

For the third and fourth restriction, every remaining candidate is mapped into the other S3D view using **equation (3.1)**:

$$\boldsymbol{p}^{\dagger}_{\mathrm{C,L}} = \boldsymbol{p}_{\mathrm{C,L}} - \hat{D}_{\mathrm{L}}\left(\boldsymbol{p}_{\mathrm{C,L}}\right) \cdot \boldsymbol{e}_x \tag{4.37}$$

$$\boldsymbol{p}^{\dagger}_{\mathrm{C,R}} = \boldsymbol{p}_{\mathrm{C,R}} + \hat{D}_{\mathrm{R}}\left(\boldsymbol{p}_{\mathrm{C,R}}\right) \cdot \boldsymbol{e}_x \ . \tag{4.38}$$

Consider the case that a given $\boldsymbol{p}^{\dagger}_{\mathrm{C,L}}$ is not contained in the right ROI, as illustrated in **figure 4.4**. This means that either $\hat{D}_{\mathrm{L}}\left(\boldsymbol{p}_{\mathrm{C,L}}\right)$ is erroneous due to the estimation process or that the depth of that candidate is not in the same range as the true gaze depth, e.g., when a foreground object is fixated and the candidate is located on the background far behind it. Because of that, the respective $\boldsymbol{p}_{\mathrm{C,L}}$ is rejected in restriction 3. Conversely, $\boldsymbol{p}_{\mathrm{C,R}}$ is rejected if $\boldsymbol{p}^{\dagger}_{\mathrm{C,R}}$ is not contained in the left ROI. Mathematically,

Restriction 3:

$$r^{\dagger}_{\mathrm{C,L}} = \|\boldsymbol{p}^{\dagger}_{\mathrm{C,L}} - \boldsymbol{p}_{\mathrm{G,R}}\| \leq r \tag{4.39}$$

$$r^{\dagger}_{\mathrm{C,R}} = \|\boldsymbol{p}^{\dagger}_{\mathrm{C,R}} - \boldsymbol{p}_{\mathrm{G,L}}\| \leq r \ . \tag{4.40}$$

This restriction actually also implements restrictions 1 and 2, albeit more elaborate so that the first two restrictions are still used to efficiently reject invalid candidates.

Finally, the disparity of the mapped candidate is compared to that of the original candidate. In an ideal case, the difference is

$$\Delta D_{\mathrm{C,L}} = \left|\hat{D}_{\mathrm{L}}\left(\boldsymbol{p}_{\mathrm{C,L}}\right) - \hat{D}_{\mathrm{R}}\left(\boldsymbol{p}^{\dagger}_{\mathrm{C,L}}\right)\right| = 0 \tag{4.41}$$

$$\Delta D_{\mathrm{C,R}} = \left|\hat{D}_{\mathrm{R}}\left(\boldsymbol{p}_{\mathrm{C,R}}\right) - \hat{D}_{\mathrm{L}}\left(\boldsymbol{p}^{\dagger}_{\mathrm{C,R}}\right)\right| = 0 \ . \tag{4.42}$$

However, these equations may not hold due to occlusion or erroneous disparity estimation. Candidates with erroneous disparities should obviously be rejected. Since observers tend to not look at occlusion areas due to the induced retinal rivalries, as mentioned in **section 3.2.3.4**, occlusion candidates are not likely to be the true gaze position. Furthermore, they exhibit low disparity estimation accuracy. Hence, candidates not satisfying **equation (4.41)** or **equation (4.42)**, respectively, are rejected. In some applications, a certain amount of error in disparity might be tolerable, so that those equations are altered to

Restriction 4:

$$\Delta D_{\mathrm{C,L|R}} \leq \Delta D_{\mathrm{max}} \ . \tag{4.43}$$

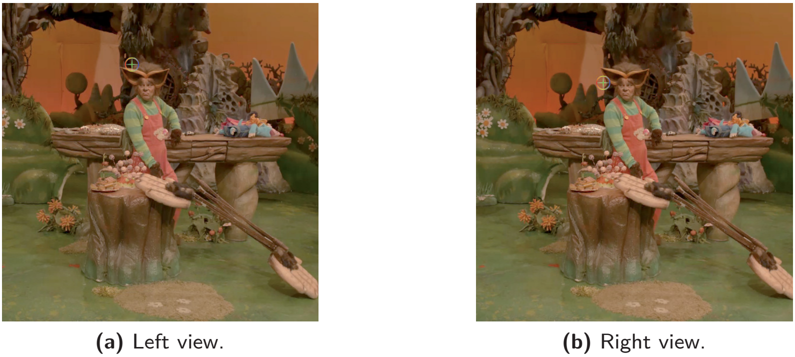
Due to restrictions 3 and 4, most candidates in the left and right view are identical with respect to stereoscopic correspondence. This is because mapped as well as original coordinates must be located inside the respective ROI and the mapping procedure solely depends on the estimated candidate disparity, which must not differ significantly between corresponding stereoscopic points. Therefore, it suffices to analyze the candidates in one view, the left view in the remainder of this work.

To illustrate the processing steps of the PVFDE up to this point, a stereoscopic image pair is displayed in **figure 4.5**. A subject fixated on the ear of the displayed actor and the gaze coordinates are highlighted in each view. The results of the ROI disparity map estimation are displayed in **figure 4.6**. The disparity maps are centered on the gaze coordinates. Please note that there is some vertical offset between the ROIs because of that. This vertical offset must be accounted for when evaluating restriction 4 based on these ROI disparity maps. As can be seen, they contain some errors due to the relatively basic estimation algorithm. The valid candidates of the left ROI are finally displayed in **figure 4.7**. Erroneous estimations are correctly rejected, as are occlusion areas, e.g., behind the ear. The valid candidates approximately assume the form of an eye, which is due to the shape of the ROI and restriction 3, which states that each mapped candidate must be contained in the opposite ROI.

4.4.1.4 Candidate Weights

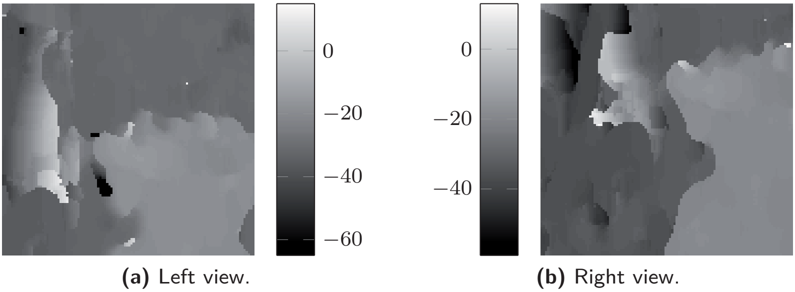
In this step, weights are computed, that quantify the confidence of identifying the correct disparity. A weight of zero is assigned to the previously identified invalid candidates, whereas the remaining ones are processed as follows.

Under the hypothesis, that $\boldsymbol{p}_{\mathrm{C,L}} = \boldsymbol{p}_{\mathrm{T,L}}$, the distance $r_{\mathrm{C,L}}$ is equivalent to the magnitude of the noise term in **equation (4.27)**, i.e., $r_{\mathrm{C,L}} = \left|\boldsymbol{p}_{\mathrm{N,L|R}}\right|$. According to the previous approximation, this means that this distance follows


(a) Left view.

(b) Right view.

Figure 4.5: A stereoscopic image pair with highlighted gaze coordinates.



(a) Left view.

(b) Right view.

Figure 4.6: Estimated square ROI disparity maps with disparity values in pixels.

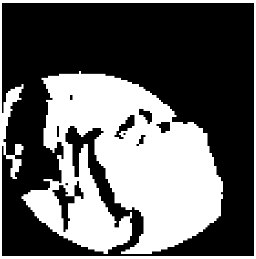


Figure 4.7: Valid candidates (white) for left view disparity map.

a one-sided normal distribution. Hence, a weight can be computed by

$$\omega_{r,\mathrm{C}} = \exp\left(-\frac{{r_{\mathrm{C,L}}}^2}{2 \cdot \sigma_r^2}\right) \ , \tag{4.44}$$

where the normalization of the gaussian distribution is omitted and σ_r^2 is the variance of that distribution. The same reasoning holds for the mapped distance $r^\dagger_{\mathrm{C,L}}$ so that a second weight is given by

$$\omega_{r^\dagger,\mathrm{C}} = \exp\left(-\frac{{r^\dagger_{\mathrm{C,L}}}^2}{2 \cdot \sigma_r^2}\right) \ . \tag{4.45}$$

Finally, **equation (4.43)** allowed for some error in $\Delta D_{\mathrm{C,L}}$. Candidates exhibiting lower errors should be assigned a higher weight. So again, a non-normalized one-sided gaussian distribution is used to compute a third weight:

$$\omega_{\Delta D,\mathrm{C}} = \exp\left(-\frac{{\Delta D_{\mathrm{C,L}}}^2}{2 \cdot \sigma_{\Delta D}^2}\right) \ . \tag{4.46}$$

These three weights are combined to one final weight for each candidate

$$\omega_{\mathrm{C}} = \omega_{r,\mathrm{C}} \cdot \omega_{r^\dagger,\mathrm{C}} \cdot \omega_{\Delta D,\mathrm{C}} \ . \tag{4.47}$$

Revisiting the previous example in **figures 4.5 to 4.7**, the resulting weights based on the left ROI are displayed in **figure 4.8**. The distance weights $\omega_{r,\mathrm{C}}$ and $\omega_{r^\dagger,\mathrm{C}}$ dominate the image, but the effect of the disparity difference weight $\omega_{\Delta D,\mathrm{C}}$ can also be seen in the form of small cracks, that are darker than their surroundings.

4.4.1.5 Weighted Histogram Filter

In order to finally retrieve the estimate $\hat{D}_{\mathrm{T}}$, the calculated weights need to be processed. The whole approach is based on the following two assumptions.

Firstly, any object is looked at in such a way that the fovea of the user contains only the image of that object, i.e., no background or other object images. This assumption concurs with the previous statement that occlusions, which

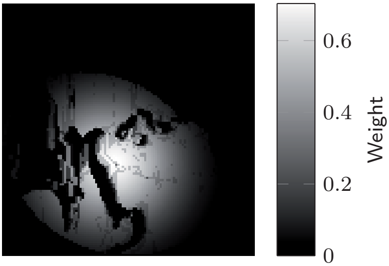


Figure 4.8: Weights for left view disparity map.

only occur at object borders, are not likely to be looked at. The avoidance of occlusions is already enforced by restriction 4. This means that the majority of valid candidates should belong to the object being looked at.

Secondly, the object does not exhibit strong depth variations inside the ROI. Considering the typical depth distribution of S3D images, this assumption is fairly reasonable.

Combining both assumptions, most candidates in the ROI are expected to belong to the targeted object and exhibit very similar disparities. Hence, the candidate confidence weights are used to form a weighted histogram, which is used to find the biggest (and most reliable) disparity cluster. This works as follows: the weights of each disparity value are accumulated and the disparity corresponding to the maximum accumulated weight is the resulting estimate $\hat{D}_{\mathrm{T}}$.

Slight violations of the first assumption can be accounted for by a modification of this step. Consider again the example in **figures 4.5 to 4.7**, where big portions of the ROIs are occupied by background disparities. The corresponding weights in **figure 4.8** yield the weighted histogram depicted in **figure 4.9**. As can be seen, the background disparities induce a big peak at −33 px, which would be wrongfully chosen as $\hat{D}_{\mathrm{T}}$. Since humans are more likely to look at foreground objects rather than background [HT11b], the foremost, i.e., rightmost histogram peak above a certain threshold is used instead as $\hat{D}_{\mathrm{T}}$ instead. For this example, the threshold has been heuristically set to 66 % of the maximum accumulated weight. Finally, $\hat{D}_{\mathrm{T}} = -19$ px is correctly chosen. The disadvantage of this modification is that the unlikely case of true gaze

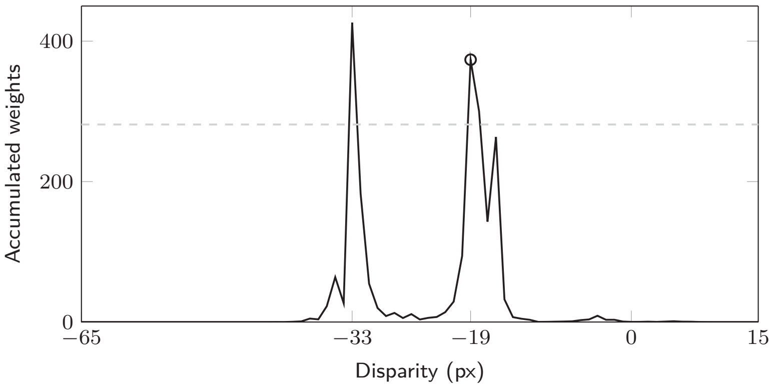


Figure 4.9: Weighted histogram of left view disparities with the 66 %-peak-threshold (dashed, gray) and the chosen disparity (circle).

positions being located on background objects in the vicinity of foreground objects cannot be processed reliably.

4.4.1.6 Parametrization

The size of the ROI is specified by the system noise and oculomotor noise. An underestimation of this size is harmful because the true gaze coordinates would not be included in the ROI. Hence, a generous ROI radius $r = 2°$ of visual angle has been chosen. Using a Full HD display with square pixels at a viewing distance $d = 3.1H$, this value corresponds to $r = 117$ px.

The outer candidates in this ROI do not affect the end result much because of the weight $\omega_{r,\mathrm{C}}$ in **equation (4.44)**. This weight is a function of σ_r, which has heuristically been set to $\sigma_r = 1° = 58.5$ px.

The choice of the maximum disparity divergence $\Delta D_{\max}$ from restriction 4 is dependent on the application. For this work, almost pixel-precise disparity is required, so that this parameter was set to $\Delta D_{\max} = 3$ px. The weight of the disparity divergence as of **equation (4.46)** has been heuristically parametrized using $\sigma_{\Delta D} = 1$ px.

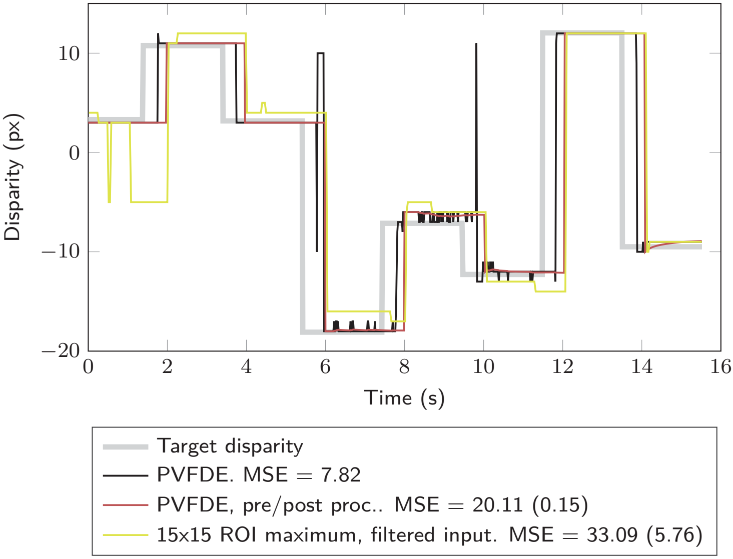


Figure 4.10: Exemplary results of the PVFDE approach proposed in this work and the maximum filter approach [Han14] for comparison. All MSE values in px^2.

4.4.1.7 Pre- and Post-Processing

Returning to the disparity graphs from the beginning of this section, the results of the PVFDE, with the parameters described above, are displayed in **figure 4.10**. The approach yields the best MSE so far, but the graph still exhibits some problems. Firstly, the disparity shortly after saccades is highly erroneous. Secondly, there are some discretization issues, where the algorithm yields disparities repeatedly switching between two neighboring integers, e.g., between 8 s and 10 s.

As a first step, the gaze data input is analyzed in a pre-processing step using the Kalman filter developed in **section 4.3.1**. In this way, outliers can be rejected so that some erroneous disparity estimations using the PVFDE are prevented. Furthermore, the filter analyzes the input for saccades. The output of the Kalman filter is not used any further because it alters the noise distribution of the gaze data. It would be non-normal, while a normal distribution is required in the candidate weight computation, see **section 4.4.1.4**.

Since the disparity of a fixated object might change continuously over time, the Kalman filter with a constant-velocity model is also a good fit for stabilizing the disparity output of the PVFDE. Furthermore, the rejection of input outliers leads to missing disparity samples, which the Kalman filter can naturally extrapolate. In this work, the filter has been heuristically parametrized with a process noise variance of $\sigma_a^2 = 0.01\,\mathrm{px}^2/\mathrm{s}^4$ and a measurement noise variance of $\sigma_n^2 = 10\,\mathrm{px}^2$. Based on the assumption that the end points of saccades are located at different depths when watching S3D content, the disparity Kalman filter is always reinitialized whenever a saccade has been detected by the input analyzer. The saccade detection can also be used in the disparity filter in order to detect when the filter does not perform appropriately and reset it in that case. However, a comparatively short analysis window is used here, in order to reduce the duration of erroneous filter output. In this work, the violation identification threshold is $e_{k,\beta}^2 = 2.1\,\mathrm{px}^2$ and the threshold for the number of connected violations is $n_{\mathrm{th}} = 3$. The results of the PVFDE with pre- and post-processing are displayed in **figure 4.10** as well. Due to the saccade detections, a certain delay is introduced, just like in the filtered examples in **figures 4.1 and 4.2**. This delay increases the MSE again, but the delay-adjusted MSE is reduced a lot.

4.4.1.8 Results

The PVFDE with real-time ROI disparity map estimation as well as pre- and post-processing has been extensively evaluated by Wermers in a supervised student master's thesis [Wer16]. The general experimental setup was the same as described in **section 5.3.1.2**. There were 13 S3D test sequences featuring natural content with varying disparity budgets and a stereoscopic pointer that the subjects were asked to track with their eyes. The pointer locations of the sequences were designed in a way that fixations, saccades, smooth pursuit and vergence movements were induced. A professional, research-grade eye tracker was used [Tob14] to record gaze data. The gaze datasets for each sequence and each subject were inspected manually for errors and (partially) rejected if necessary, resulting in 18 to 26 gaze and ground truth datasets per sequence.

Table 4.1: Depth estimation error of PVFDE compared to the raw disparity computation approach and the optimal calibration thereof. Raw disparity values that represent eye divergence have been excluded from the evaluation.

Approach	MSE (cm^2)	MAE (cm)
PVFDE, pre-/post-proc.	37.41	1.49
Raw, filtered input	691638.01	206.59
Raw, filtered input, calibrated	1785.62	25.47

The ground truth datasets were constructed with consideration of the detected tracking errors and possible delays at saccades. The resulting MSE values of depth over all sequences and subjects are summarized in **table 4.1**. The MSE of the PVFDE was improved by a factor of almost 50 compared to the optimally calibrated conventional approach. Again, this calibration represents a theoretical best-case scenario, which could only be achieved using a head rest [Hol11], if at all. Furthermore, about 5 % of the raw disparity values represented eye divergence so that no depth value could be computed using **equation (3.7)**. These values were excluded from the evaluation, whereas, normally, they would be mapped to infinity, yielding an infinite MSE.

The MAE is also included in the table for comparison to other publications that only report this metric instead of the usual MSE. Please note that the viewing distance $d \approx 180\,\text{cm}$ in this experiment was very high compared to experiments conducted by most other groups and that the eye tracking error in the display plane scales with viewing distance. This should be taken into consideration when comparing the values in **table 4.1** to other publications.

The algorithm was also tested with the cheap $100 “The Eye Tribe Tracker”. While no elaborative subjective evaluation was carried out with this eye tracker, quick tests showed no significant performance degradation compared to the professional eye tracker.

Wermers also extended the PVFDE in his master’s thesis to utilize motion vector fields of the scenes as well [Wer16]. However, the gain of this extension was comparatively small, especially when weighing it against the big computational load of the motion estimation. Because of that, it has not been integrated into the final prototype described in **section 5.2.2**.

4.5 Conclusion

The output of eye trackers is degraded by technological system noise and oculomotor noise. The latter comprises the unconscious eye movements. While 2D gaze filters can improve precision, i.e., reduce the spread of gaze samples, an improvement of accuracy is not necessarily achievable. As a consequence, the choice of a gaze filter is mostly determined by its features. A Kalman filter is useful for eye tracking because of its ability to naturally extrapolate missing samples or rejected outliers. Outliers can be detected reliably by checking the validity of the underlying eye motion model. For the Kalman filter derived in this work, only the slow eye movements were modeled. Saccades are handled by reinitializing the filter upon detecting the saccade. A new saccade detection algorithm was proposed. Since saccades are not incorporated in the model, any saccade will induce a series of model violations. Hence, the number of connected violations is counted and compared with a decision threshold. The approach treats missing or invalid samples as “Don’t Cares”, which is again very useful in eye tracking.

The noise sources described above are, for the most part, uncorrelated between the eyes. This is problematic in the case of 3D visual focus estimation since the induced errors are accumulated, rendering the common approaches very inaccurate. The proposed **probabilistic visual focus disparity estimation (PVFDE)** estimates the 3D scene structure, i.e., disparity of a region of interest in real-time and uses that information in conjunction with computed per-pixel confidence weights to improve the 3D visual focus estimation. This new contribution yields an improvement of the mean squared error of the depth coordinate by a factor of almost 50 compared to the results of a relatively unrealistic, ideal, linear 3D calibration. Without such a calibration, there is a more than 10 000-fold improvement.

5 Dynamic Horizontal Image Translation

This chapter represents the main matter of this work. The mathematical fundamentals of **horizontal image translation (HIT)** and **dynamic horizontal image translation (DHIT)** have already been described in **section 3.1.1 and section 3.2.3.2.1**, respectively. This chapter starts with a review of DHIT-related literature in **section 5.1**. Two major contributions of this work are enhancements of the DHIT called **distortion-free dynamic horizontal image translation (DHIT+)** and **gaze adaptive convergence in stereo 3D applications (GACS3D)**, which are described in **section 5.2**. Finally, the results of four experiments conducted for this work are presented in **section 5.3**. The aim of these experiments was to analyze the basic perceptual properties of the DHIT and the proposed enhancements of it.

This chapter and, predominantly, the description of all experiments are based on previous publications by the author [Eic13c, Eic15, Eic16].

5.1 Review and Motivation

The common approach to design the DHIT is a heuristic one: The stereographer simply checks if it looks right. However, this is problematic because there are individual differences in the perception of the DHIT, see results of experiments (I) and (II) in **sections 5.3.2.5 and 5.3.3.3**. It might just happen that a stereographer is comparatively insensitive towards the DHIT and tunes it wrongly for the general audience. Chamaret et al. have done some quick tests to find out how much of a DHIT shift may be applied between two successive frames without being noticed by an observer [Cha10]. Aside from this, variations of disparity have only been investigated in the context of camera baseline variation, for example by Ware [War95], with the result that small disparity variations are not perceivable. The lacking research was the motivation for the first experiment of this work with the aim to find shift speed thresholds for annoyance and perception. It is described in **section 5.3.2**.

The results of that experiment are especially important in the context of automated DHIT approaches. There are many applications where automation is useful or even mandatory. To start with the HIT, it can be automated simply by estimating the disparity range of a given scene and shift it behind the display [Xu12] or, even better, into the **zone of comfort (ZOC)**. Instead of controlling the disparity range, Kim et al. have proposed to apply the HIT in such a way that the fusion time is minimized, which yielded a reduction of visual discomfort in their subjective experiments [Kim13]. Returning to disparity range control, an automated concept related to DHIT has been described by Zhang et al. [Zha13]: After shot cut detection, the disparity ranges of the frames right before and after the cut are estimated and a DHIT is deployed to mitigate any depth discontinuities. Thereby, an automatic active depth cut is implemented. Another approach is to analyze the visual saliency of a given scene and have the DHIT be set in such a way that salient areas are near or in the convergence plane [Cha10, Han14]. The topic of 3d visual attention has been investigated for example by Huynh-Thu et al. [HT11b, HT11a], and the computations usually require a per-pixel disparity map. Since temporal correlations are very important in visual saliency, the elaborate computation typically also involves evaluation of multiple frames. This renders saliency based approaches rather inappropriate for real-time applications. Instead of the visual saliency estimation, the actual visual focus can be used to control DHIT by utilizing an eye tracker. A system like that called GACS3D was proposed by the author of this thesis in 2013 [Eic13c]: The **zero parallax setting (ZPS)** is slowly established at the visual focus, thereby reducing the **accommodation vergence discrepancy (AVD)**. Hanhart et al. have compared a similar approach to a visual saliency based DHIT design and the unprocessed **stereo 3D (S3D)** views in a subjective experiment [Han14]. Their gaze adaptive approach yielded the best results in picture quality, depth quality and visual discomfort. A similar experiment has been conducted by the author of this thesis, which is described in **section 5.3.5**. Bernhard et al. have evaluated visual fatigue by linking it to the measured stereoscopic fusion times with and without gaze adaptive DHIT [Ber14]. The validity of this link has not been discussed, however. Long fusion time and small fusional limits have previously only been used to identify subjects who are prone to visual fatigue [Kim11], but not to measure visual fatigue objectively. It is also known that fusion time is dependent on disparity [Kim11], which is obviously reduced in the case of gaze adaptive DHIT. However, Bernhard et al. also carried out a subjective evaluation that yielded a slight improvement of gaze adaptive DHIT over the unprocessed stereo views in depth quality and visual discomfort [Ber14]. The differences between GACS3D and these two similar gaze adaptive approaches are described in **section 5.2.2.6**.

Independent of the way the DHIT is controlled, there is still the problem that black borders can appear in the S3D views, which has already been described in the context of HIT in **section 3.1.1**. The black borders are created by shifting content out of the display area so that no image information is available on the other side of the respective S3D view. Since the black borders are generated on opposing sides of the S3D views, they have a certain disparity so that this effect is actually perceived as the commonly used stereoscopic floating window, see **section 3.2.3.5.1**. In the case of the DHIT, the floating window disparity changes over time, making it a dynamic floating window. Broberg argues that the black borders of the HIT create false depth cues and can distract viewers [Bro11]. However, for the DHIT, the author of this thesis argues that this effect can actually be seen as a feature rather than an artifact because it helps to avoid new window violations under certain conditions. This is explained in more detail in **section 5.2.2.4**, where an automated floating window approach is proposed, which is generally applicable to DHIT, but has been developed for GACS3D.

5.2 Enhancements of DHIT

Two enhancements of DHIT have been developed in this work, which are described in this section.

5.2.1 Distortion-Free DHIT

The temporally dynamic distortion of the depth budget due to DHIT has already been described in **section 3.1.1 and section 3.2.3.2.1**. In this section, a new approach for **distortion-free dynamic horizontal image translation (DHIT+)** is proposed. In **section 5.3.3**, it is experimentally compared to the regular DHIT.

Suppose a DHIT is to be performed over a couple of frames on a given convergence disparity range $[D_{\text{conv,min}}, D_{\text{conv,max}}]$. The alteration of the depth budget due to the DHIT can be mitigated by scaling the camera baseline b_{c}. It can be easily shown that this scaling operation by a factor $\alpha \in \Re^+$ implicitly also applies to disparity. Recalling that shifted disparities denote disparities after an HIT with a certain convergence disparity D_{conv} has been applied, the shifted disparity range of the scaled camera baseline views is given by

$$\begin{aligned}\left[\tilde{D}^*_{\text{min}}, \tilde{D}^*_{\text{max}}\right] &= \alpha \cdot \left[\tilde{D}_{\text{min}}, \tilde{D}_{\text{max}}\right] \\ &= \alpha \cdot \left[D_{\text{min}} - D_{\text{conv}}, D_{\text{max}} - D_{\text{conv}}\right] \ , \end{aligned} \tag{5.1}$$

which yields a depth budget of

$$Z = \frac{d \cdot b_{\text{e}}}{b_{\text{e}} + \alpha \cdot \rho \cdot \tilde{D}_{\text{min}}} - \frac{d \cdot b_{\text{e}}}{b_{\text{e}} + \alpha \cdot \rho \cdot \tilde{D}_{\text{max}}} \overset{!}{=} Z_{\text{targ}} \ . \tag{5.2}$$

This depth budget is supposed to be constant, i.e., equal to a chosen target depth budget Z_{targ} for all defined convergence disparities D_{conv}. Rearranging **equation (5.2)** yields the quadratic equation

$$\alpha^2 + \alpha \cdot \underbrace{b_{\text{e}} \frac{\tilde{D}_{\text{max}} + \tilde{D}_{\text{min}} - D_{\text{B}} \cdot d / Z_{\text{targ}}}{\tilde{D}_{\text{max}} \cdot \tilde{D}_{\text{min}}}}_{g} + \underbrace{\frac{{b_{\text{e}}}^2}{\tilde{D}_{\text{max}} \cdot \tilde{D}_{\text{min}}}}_{h} = 0 \ , \tag{5.3}$$

with two singularities, i.e., $\tilde{D}_{\text{max}} \neq 0$ and $\tilde{D}_{\text{min}} \neq 0$. It can be rearranged to find

$$\alpha = -\frac{g}{2} - \operatorname{sign}\left(\tilde{D}_{\text{max}}\right) \cdot \operatorname{sign}\left(\tilde{D}_{\text{min}}\right) \cdot \sqrt{\frac{g^2}{4} - h} \ , \tag{5.4}$$

with the sign-function

$$\operatorname{sign}(x) = \frac{x}{|x|} \quad \text{for} \quad \{x \in \Re | x \neq 0\} \ . \tag{5.5}$$

The two singularities in **equation (5.3)** can be resolved by going back to **equation (5.2)**, substituting either $\tilde{D}_{\text{max}}$ or $\tilde{D}_{\text{min}}$ with zero, and isolating α. In the case of $\tilde{D}_{\text{max}} = 0$, this yields

$$Z_{\text{targ}} = \frac{d \cdot b_{\text{e}}}{b_{\text{e}} + \alpha \cdot \rho \cdot \tilde{D}_{\text{min}}} - d \tag{5.6}$$

$$\Leftrightarrow \alpha = -\frac{b_{\text{e}}}{\tilde{D}_{\text{min}} \cdot (1 + d/Z_{\text{targ}})} \ . \tag{5.7}$$

The minimum shifted disparity is handled in the same way so that the final solution for the scaling factor is

$$\alpha = \begin{cases} -\frac{b_{\text{e}}}{\tilde{D}_{\text{min}}\left(1+d/Z_{\text{targ}}\right)} & \text{if } \tilde{D}_{\text{max}} = 0 \\ -\frac{b_{\text{e}}}{\tilde{D}_{\text{max}}\left(1-d/Z_{\text{targ}}\right)} & \text{if } \tilde{D}_{\text{min}} = 0 \\ -\frac{g}{2} - \operatorname{sign}\left(\tilde{D}_{\text{max}}\right) \cdot \operatorname{sign}\left(\tilde{D}_{\text{min}}\right) \cdot \sqrt{\frac{g^2}{4} - h} & \text{otherwise.} \end{cases} \tag{5.8}$$

Now, the basic DHIT+ algorithm constitutes the following steps:

1. Choose a target depth budget Z_{targ}.
2. Given a sequence of per-frame convergence disparities D_{conv}, the scaling factor α for each frame is computed using **equation (5.8)**. It is recommended to use the average human eye baseline here, which is $b_{\text{e}} =$ 6.3 cm [Dod04, War95].
3. The S3D views are generated with the adjusted camera baseline

   $$b_{\text{c}}^{*} = \alpha \cdot b_{\text{c}}. \tag{5.9}$$

4. The HIT is finally applied individually to each frame. Since the scaling factor also applies to disparity, the convergence disparity needs to be scaled as well to

   $$D_{\text{conv}}^{*} = \alpha \cdot D_{\text{conv}}. \tag{5.10}$$

5.2.1.1 Active Depth Cuts using DHIT+

With the DHIT, an active depth cut is implemented simply by slowly shifting the S3D views before and after a shot cut such that objects of interest are located at approximately the same depth, e.g., the convergence plane, when the cut occurs. In this way, visual fatigue inducing depth discontinuities are reduced.

The design of an active depth cut with the DHIT+ is different because shifting a scene closer to the observer should be avoided. This shift direction leads to a camera baseline increase. Considering that the disparity range is often at the limits of the ZOC, a baseline increase is likely going to induce too big disparities protruding the ZOC. Furthermore, if depth-image-based rendering is used to extrapolate the S3D views, big occlusion holes may occur, that have to be filled appropriately[1]. Hence, the DHIT+ should only be performed before or after the shot cut, but not both, as the following best practice example illustrates.

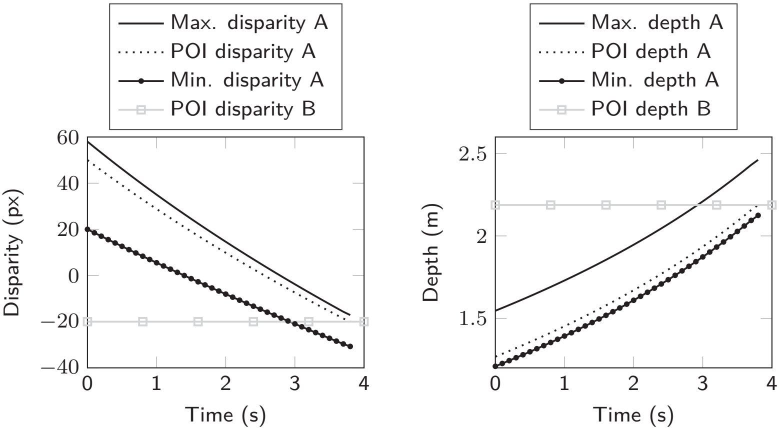
A cut from shot A to B is to be performed. These shots occupy depth budgets Z_{A} and Z_{B}, respectively. A **point of interest (POI)** in shot A exhibits a certain disparity $D_{\mathrm{POI,A}}$, which is supposed to smoothly transition to $D_{\mathrm{POI,B}}$, the disparity of a POI in shot B. In the simple case that $D_{\mathrm{POI,A}} > D_{\mathrm{POI,B}}$, shot A is shifted away from the observer: The convergence disparity of shot A is increased, and the respective scaling factor α is computed with $Z_{\mathrm{targ}} = Z_{\mathrm{A}}$, until the scaled, shifted disparity satisfies

$$(D_{\mathrm{POI,A}} - D_{\mathrm{conv}}) \cdot \alpha = D_{\mathrm{POI,B}} \ . \tag{5.11}$$

This is illustrated in **figure 5.1a**. As can be seen, $D_{\mathrm{POI,A}}$ is slowly shifted towards $D_{\mathrm{POI,B}}$ until **equation (5.11)** is fulfilled at the end of the graph. Then, the POIs in both shots are located at the same depth and the cut can be performed. In **figure 5.1b**, the same operation is displayed in the depth domain, clearly illustrating the constant depth budget.

In the case that $D_{\mathrm{POI,A}} < D_{\mathrm{POI,B}}$, shot A is not shifted closer to the observer for the reasons mentioned above. Instead, the cut to shot B is performed

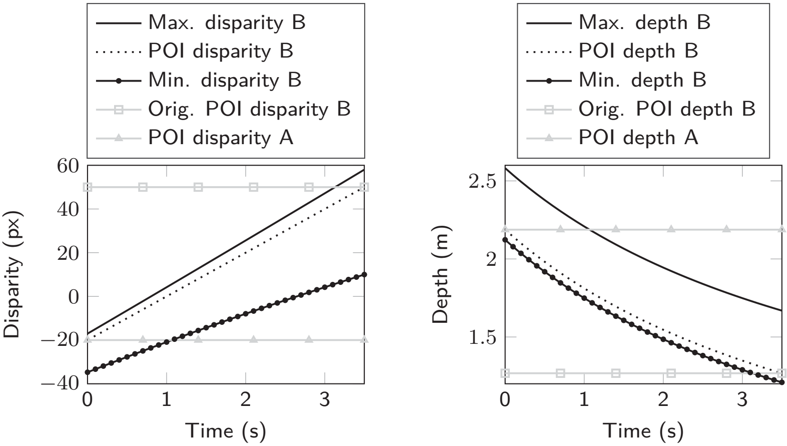
[1] In case of a camera baseline decrease, i.e., S3D view interpolation, the induced occlusion holes are very small and can usually be filled easily.



(a) Scaled disparity domain.

(b) Depth domain.

Figure 5.1: Example for an active depth cut from front (A) to back (B) using DHIT+.



(a) Scaled disparity domain.

(b) Depth domain.

Figure 5.2: Example for an active depth cut from back (A) to front (B) using DHIT+.

without any prior shifting of shot A, while shot B is preconverged so that

$$D_{\mathrm{POI,A}} = (D_{\mathrm{POI,B}} - D_{\mathrm{conv}}) \cdot \alpha \ , \tag{5.12}$$

with $Z_{\mathrm{targ}} = Z_{\mathrm{B}}$. This means that, at the beginning, shot B is actually shown with a decreased baseline that is increased back to its original state again in the course of the DHIT+, as illustrated in **figure 5.2a**. Thereby, S3D view extrapolation is avoided, i.e., big occlusion holes are prevented.

5.2.1.2 Controlling Speed in DHIT+

The convergence disparity D_{conv} is usually adjusted at a fixed speed in the case of the DHIT. However, this fixed speed is not maintained in the DHIT+ because the convergence disparities are multiplied by varying scaling factors α in **equation (5.10)**. Therefore, the speed of the scaled convergence disparity D^{*}_{conv} should be controlled rather than D_{conv}. Since the rate of change of $D^{*}_{\mathrm{conv}} = \alpha \cdot D_{\mathrm{conv}}$ is supposed to be constant, and the computation of α is dependent on D_{conv}, a different sequence of convergence disparities has to be found through numeric optimization approaches, i.e., by trying different convergence disparity values until the desired speed for D^{*}_{conv} is achieved.

A speed controlled scaled convergence disparity is the basis of **figures 5.1a and 5.2a**. However, in contrast to the DHIT, different disparities change at different speeds here. For example, the lower bound of the disparity range in **figure 5.2a** exhibits a certain curvature, whereas the upper bound is almost linear. This is due to the dynamic adjustment of the disparity budget. Instead of controlling speed in the disparity domain, it is also possible to control the speed along the depth axis so that a linear movement is created. This is more natural than what is depicted in **figures 5.1b and 5.2b**. Again, numeric optimization techniques have to be used to find a different sequence of convergence disparities corresponding to the desired depth movement.

5.2.1.3 Further Properties of DHIT+

The solution for α is independent of viewing distance d, which can be seen in **equations (5.3) and (5.7)**: The only occurrence of d is resolved through the

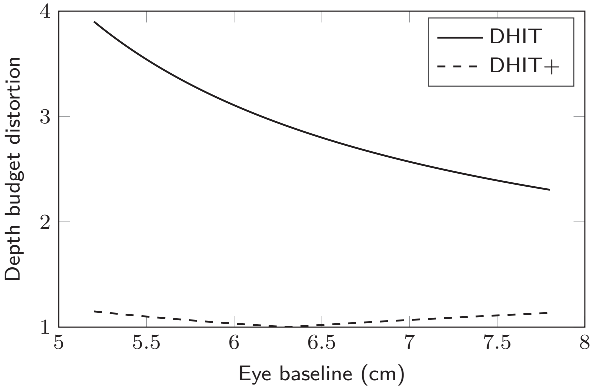


Figure 5.3: Extreme example of depth budget distortion due to DHIT and an eye baseline mismatch in the DHIT+. The DHIT+ has been computed for $b_{\mathrm{e}} = 6.3\,\mathrm{cm}$ and $Z_{\mathrm{targ}} = Z_{\mathrm{min}}$.

division by Z_{targ}, which is linearly dependent on d, see **equation (3.9)**. This means that a constant depth budget is ensured with DHIT+ independently of observer distance, which is important for living room and cinema environments.

The eye baseline b_{e}, on the contrary, does affect α. If the eye baseline of an observer differs from the one used in the computation of α, there will be a certain distortion of the depth budget remaining. However, it is very small compared to the distortion of the regular DHIT. As an extreme example, consider the case that an S3D sequence exhibiting a disparity budget of $D_{\mathrm{B}} = 58\,\mathrm{px}$ is shifted all the way from the front to the back of the display such that the whole ZOC of ±58 px, see **section 3.2.3.1.2**, is occupied at some point in time. This means that the shift budget is $\mathrm{SB} = 58\,\mathrm{px}$, which induces a big depth budget distortion. The minimum depth budget is supposed to be maintained throughout the shift operation. It is given when the scene is at the foremost shift position, i.e., when $D_{\mathrm{conv}} = D_{\mathrm{conv,min}}$. This depth budget can be computed using **equation (3.9)**:

$$Z_{\mathrm{min}} = Z|_{D_{\mathrm{conv}}=D_{\mathrm{conv,min}}} = Z_{\mathrm{targ}} \ . \tag{5.13}$$

Let the depth budget distortion be defined as the ratio of maximum to minimum depth budget. In **figure 5.3**, the depth budget distortion in this example is displayed for varying eye baselines. As expected, the distortion is a lot lower, even for the most extreme outliers. Assuming a normal distribution for the human eye baseline, and using parameters of a large scale anthropometric survey [Gor89], the average depth budget distortion for this extreme example is 1.033 for the DHIT+. In other words, it is neglectable. For the regular DHIT, it is 2.907.

The baseline adjustment in the DHIT+ can be carried out easily for **computer generated imagery (CGI)** by moving the virtual cameras. For natural content on the other hand, this is not as straight-forward because adjusting the baseline dynamically on set is not feasible. In both cases, precise knowledge about the specific active depth cut parameters is necessary to perform the baseline adjustment. As an alternative, a DHIT+ active depth cut can be emulated by appropriate camera movement: Having a constant depth budget move back or forth in the DHIT+ is similar to a camera movement in depth. It is not the same, however. The camera movement differs in a way that image content enters or leaves the vertical field of view depending on the movement direction, whereas this is not the case with the DHIT+. Since this approach is implemented a lot easier than the DHIT+, the camera movement is the recommended procedure whenever the active depth cut parameters are already known while shooting.

However, when that is not the case, applying the DHIT+ in post-production is the only option. Here, depth-image-based rendering techniques, like the ones described by Müller et al. [Mül11], are needed to adjust the camera baseline. A respective approach, which is capable of processing Full HD S3D content in real-time, has been implemented in CUDA C++ for this work. However, due to common view interpolation problems and inaccurate depth maps, its results contain many artifacts, although multiple artifact reduction algorithms have been implemented [Bru15]. In an effort to minimize occlusion artifacts, it is useful to choose $Z_{\mathrm{targ}} = Z_{\mathrm{min}}$ so that the camera baseline is decreased and not increased, as already explained in **section 5.2.1.1**. However, in order to ensure high quality, sophisticated, elaborative, semi-automatic methods from professional 2D to 3D conversion should be used here.

5.2.2 Gaze Adaptive DHIT

The new approach called **gaze adaptive convergence in stereo 3D applications (GACS3D)** is one of the main contributions of this work and has been described in previous publications by the author [Eic13c, Eic16]. Now, recall the distinction between shifted and unshifted disparities, as explained in **section 3.1.1**. The general concept of GACS3D is to estimate the visual focus, retrieve the unshifted disparity at that point and use that as the new target convergence disparity in the DHIT process. Thereby, the ZPS is slowly established at the visual focus, reducing the AVD, which is supposed to reduce visual discomfort and visual fatigue. The approach is basically compatible to the DHIT+. However, as mentioned at the end of **section 5.2.1**, the common real-time depth-image-based rendering approaches contain too many artifacts.

GACS3D comprises five steps, as depicted in **figure 5.4**. The steps are described in more detail in the following sections. Some benchmark results are given afterwards and it is compared to similar approaches. GACS3D is later evaluated in two experiments in **sections 5.3.4 and 5.3.5**.

5.2.2.1 Visual Focus Estimation

A professional, research-grade, remote eye tracker is used [Tob14]. This eye tracker combines both bright and dark pupil tracking and allows for free head movement inside its comparatively big head box, ensuring comfortable viewing conditions. The eye tracker delivers 60 binocular visual focus samples $\left(\tilde{\boldsymbol{p}}_{\mathrm{G,L|R}}\right)_k$ per second in real screen-space pixel coordinates, where k is the index of the time instance. These coordinates are affected by noise due to technological limitations: According to the eye tracker specifications, the accuracy is 0.4° and precision is 0.34°. Accuracy and precision are also degraded by involuntary, unconscious eye movements. Only the conscious eye movements are of interest in human-machine-interaction, as already described in **section 4.3.1.2**. Because of that, the raw gaze data needs to be processed accordingly in the next step.

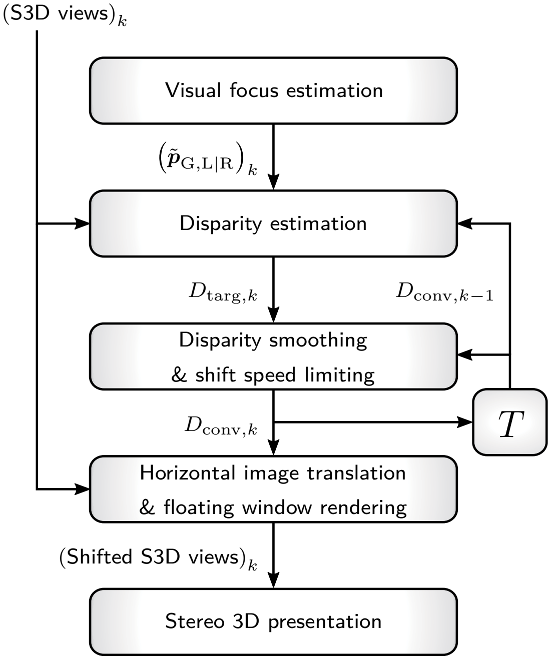


Figure 5.4: Processing steps of GACS3D.

5.2.2.2 Disparity Estimation

In this step, the disparity currently looked at needs to be retrieved, which yields the new unshifted target convergence disparity $D_{\mathrm{targ},k}$. This is done using the **probabilistic visual focus disparity estimation (PVFDE)** described in **section 4.4.1**. The PVFDE has been developed specifically for GACS3D and is capable of retrieving accurate visual focus disparities based on noisy gaze data. It involves a real-time **region of interest (ROI)** disparity map estimation. Due to the DHIT, the displayed views in iteration $k-1$ have been shifted by $\mp D_{\mathrm{conv},k-1}/2$. This means that the gaze coordinates are given in the shifted domain, and applying the PVFDE on these past views would yield shifted domain disparities. However, unshifted disparities are needed for setting the convergence disparity. This could be done by correcting the retrieved shifted target disparity by $D_{\mathrm{conv},k-1}$ using **equation (3.4)**. Instead of that, it is also possible to transform the gaze coordinates to the unshifted domain and apply the PVFDE to the unshifted views of the current iteration k. This is advantageous for two reasons: Firstly, removing a dependency on previous video frames avoids initialization issues, and secondly, the disparity map estimation might be better because there is no reduction of image information due to interpolation in the shift operation. However, there is a certain timing offset since the gaze data is dependent on the presented S3D views from iteration $k-1$, but the disparity is estimated based on the S3D views of iteration k, which are displayed in the future. Compared to the overall system delay, this offset is neglectable, however. The transformation of the gaze data to the unshifted domain can be achieved by inverting the previous HIT:

$$\left(\boldsymbol{p}_{\mathrm{G,L}}\right)_k = \left(\tilde{\boldsymbol{p}}_{\mathrm{G,L}}\right)_k + \frac{1}{2} D_{\mathrm{conv},k-1} \cdot \boldsymbol{e}_x \tag{5.14}$$

$$\left(\boldsymbol{p}_{\mathrm{G,R}}\right)_k = \left(\tilde{\boldsymbol{p}}_{\mathrm{G,R}}\right)_k - \frac{1}{2} D_{\mathrm{conv},k-1} \cdot \boldsymbol{e}_x \ . \tag{5.15}$$

5.2.2.3 Disparity Smoothing and Shift Speed Limiting

In the previous step, the new target convergence disparity $D_{\mathrm{targ},k}$ has been retrieved. The results of experiment (I), which are described in **section 5.3.2**, suggest that the shift speed $v_{\mathrm{s},k}$ needs to be limited to a certain value $v_{\mathrm{s,max}}$. Furthermore, the effect of sudden changes in shift speed was not investigated in that experiment. So, in order to avoid such speed changes, a simple one-tap recursive filter is used to smooth the disparity

$$\overline{D}_{\mathrm{targ},k} = \gamma \cdot D_{\mathrm{targ},k} + (1-\gamma) \cdot D_{\mathrm{conv},k-1} \ , \tag{5.16}$$

where the smoothing factor was heuristically chosen to $\gamma = 0.125$ (for 60 iterations per second). Afterwards, the shift speed is limited. The current unlimited speed is given by

$$v_{\mathrm{s},k} = \left|\overline{D}_{\mathrm{targ},k} - D_{\mathrm{conv},k-1}\right| \ , \tag{5.17}$$

and the shift direction is

$$\mu_k = \operatorname{sign}\left(\overline{D}_{\mathrm{targ},k} - D_{\mathrm{conv},k-1}\right) \ . \tag{5.18}$$

The new speed-limited convergence disparity can be computed by

$$D_{\mathrm{conv},k} = \begin{cases} D_{\mathrm{conv},k-1} + \mu_k \cdot v_{\mathrm{s},k} = \overline{D}_{\mathrm{targ},k} & \text{if } v_{\mathrm{s},k} < v_{\mathrm{s,max}} \\ D_{\mathrm{conv},k-1} + \mu_k \cdot v_{\mathrm{s,max}} & \text{otherwise.} \end{cases} \tag{5.19}$$

5.2.2.4 Horizontal Image Translation and Floating Window Rendering

The S3D views have to be shifted by $\mp D_{\mathrm{conv},k}/2$ to the right. Since the convergence disparity is non-integer, a horizontal cubic interpolation is used to render the shifted views, which can be done very efficiently in parallel computing using CUDA C++.

Afterwards, in order to avoid window violations, the floating window is applied by rendering black borders on the sides of the S3D views, as described in **section 3.2.3.5.1**. There are different, application-dependent strategies for designing the floating window. In this case, an automated approach is needed. Under the reasonable assumption that a floating window is already given in the source material[2], consider now the case that a DHIT is supposed to be added with a certain shift speed v_s and over a shift budget SB. Since the S3D views are shifted in opposite directions, disparity changes with a speed of $2 \cdot v_\mathrm{s}$. The disparity of the border, however, only changes with v_s since only the border in one view is moving, whereas the other border is fixed at the display boundary. This means that no new window violations are induced when the stereoscopic content is shifted further away from the viewer because the content moves back twice as fast as the floating window, which ensures that the floating window remains in front of the stereoscopic content in the border regions. Conversely, if the content is shifted nearer to the viewer, the content moves twice as fast to the front compared to the floating window. Hence, new window violations can only be created in this case. A simple but effective solution to this problem can be achieved by the following steps:

1. Estimate the disparity of the floating window already contained in the source material simply by counting the number of black pixels.

[2] Most S3D content already has a floating window. If not, the disparities at the borders of the S3D views have to be estimated to find the necessary floating window disparity. This is problematic, however, because the border region disparities are error-prone due to lacking pixels for the correspondency estimation.

2. Count the columns of image content that have been shifted out of the display area during a DHIT towards the observer.
3. Increase the width of the black border by that amount simply by rendering black over the respective image content.

These steps ensure that the original floating window disparity is maintained. The floating window moves just as fast towards the viewer as the S3D content. Let $\mathrm{SB}_{\mathrm{front}}$ be defined as the portion of the shift budget where the content is moved closer to the viewer than the source material, i.e., the amount of DHIT where window violations can actually occur. Under the condition that no pixel in the respective border region of width $|\mathrm{SB}_{\mathrm{front}}|$ exhibits a disparity bigger than the floating window disparity, no new window violations are created. This condition is fairly reasonable considering the usual small shift budgets. Furthermore, a violation of this condition will likely only induce mild window violations. As a downside of this approach, the width of the stereoscopic window shrinks much more than without any compensation. In principle, this approach can also be used when the content is shifted away from the viewer to ensure that the depth of the floating window and the content does not differ too much. However, this advantage is traded off with more shrinkage of the stereoscopic window again. As mentioned above, the convergence disparity is non-integer, so that the black border and the image content should be interpolated, e.g., using a simple linear interpolation. A non-integer floating window disparity estimation could also be implemented, which would prove advantageous in the case of dynamic floating windows.

Finally, the shifted views are multiplexed into an S3D frame format and passed to an OpenGL displaying routine.

5.2.2.5 Prototype Benchmark Results

The prototype has been implemented in a combination of MATLAB and CUDA C++ in order to utilize the parallel processing power of a graphics card. The prototype achieves 120 frames/s at Full HD resolution, including video decoding, the OpenGL displaying routine and two disparity map estimations per frame for 201 px by 101 px big regions of interest with a 9 px by 9 px **sum**

of absolute differences (SAD) window size. Without the displaying routine, the prototype achieves 135 frames/s. The test system consists of a 3.4 GHz quad-core with hyper-threading, 16 GB DDR3-1600 and a CUDA compute capability 3.5 graphics card that achieves 5.0 TFLOPS in single precision.

5.2.2.6 Distinction from Similar Approaches

The approach by Bernard et al. does not support natural S3D content [Ber14]. For their experiment, they have solved the problem of inaccurate eye tracking by presenting very simple computer generated stimuli to the subjects with very big targets. This also avoids any issues in the disparity estimation and lookup because this information is readily available in the computer. The approach by Hanhart et al., on the contrary, does support natural S3D content [Han14]. However, it relies on precomputed disparity maps, whereas GACS3D estimates disparity in real-time, making it applicable to all kinds of S3D content without any further pre-processing. Furthermore, Hanhart's group has not considered the huge extent of the system and oculomotor noise, which has already been pointed out at the end of **section 4.4**. For performance reasons, they use a nearest neighbor interpolation to implement the DHIT. This makes the shift speed discretely transition between 0 px/frame and 1 px/frame to achieve the desired 0.5 px/frame on average. This shift speed is comparatively high, as discussed in **section 5.3.5.6**. In fact, in our experiments, many subjects would have been annoyed by such a fast DHIT. The nearest neighbor approach may furthermore lead to asymmetric shifting, with unknown side effects.

5.3 Evaluation

In this section, the results of four experiments are presented. In experiment (I) described in **section 5.3.2**, the basic properties of DHIT were analyzed. The main aim of the experiment was to find out what parametrization leads to a perceivable or annoying DHIT. The proposed DHIT-enhancement DHIT+ was evaluated in a similar manner in experiment (II), which is described in **section 5.3.3**. The results of these experiments were used to parametrize the proposed enhancement GACS3D, which was evaluated in experiments (III) and

(IV), described in **sections 5.3.4 and 5.3.5**.

5.3.1 Commonalities

All four experiments share a common ground, which is described in this section.

5.3.1.1 Rendering

The DHIT, or enhancements of it, is used in all four experiments. In order to avoid discretization issues or blurring, as would be the case with nearest-neighbor or linear interpolation, respectively, the DHIT is rendered in real-time using a horizontal cubic interpolation.

5.3.1.2 Experimental Setup

The following experimental setup as depicted in **figure 5.5** was designed in accordance with the respective ITU recommendations [ITU12b, ITU12a] and has been used in all experiments, except for a few changes that are pointed out in the respective sections. The stimulus display was a 47 inch Full-HD 3D-LCD with passive polarizer glasses that was placed on a long table. All signal processing of the display was disabled in the display settings, and it was driven in Full-HD at a frame rate of 60 frames/s so that any effects of low frame rate can be neglected. The crosstalk of the stimulus display had been measured at nine points equally distributed over the screen plane, which yielded an average crosstalk of 0.88 % according to the definition by Liou, as described by Woods [Woo10]. This level of crosstalk does not induce visual discomfort, as described in **section 3.2.3.3**. The subject sat in an office chair in a distance of $d = 3.1H \approx$ 180 cm away from the screen, with H representing the height of the display. From the subject's point of view, the table was only visible in the extreme periphery. Nevertheless, the table was covered with cloth in all experiments in order to avoid reflections.

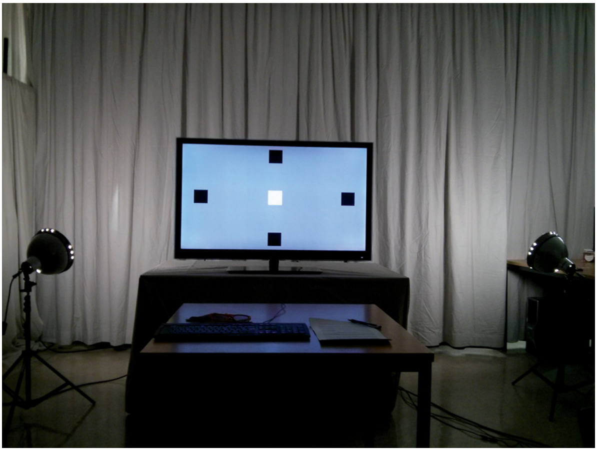


Figure 5.5: The general experimental setup with a keyboard and a questionnaire exemplarily laid out on the table.

5.3.1.3 Subject Screening and Rejection Criteria

As recommended in [ITU12a], subjects were examined regarding their visual performance prior to participation in any experiment in the following way: Snellen charts, the Ishihara color test, the randot butterfly stereogram and the circle test were used to assess binocular visual acuity, color perception, gross stereopsis and fine stereopsis, i.e. stereo acuity, respectively. A good overview of examination methods for subjective S3D evaluations along with rejection criteria is given by Lambooij et al. [Lam09a]. For the utilized examination methods, the authors recommend to reject subjects meeting the following criteria:

- Subjects exhibiting visual acuity worse than 80 %.
- Subjects failing the test for gross stereopsis.
- Subjects exhibiting stereo acuity worse than 60″.

These rejection criteria were adopted in this work except for the one on stereo acuity. In the experiments, subjects exhibiting up to 140″ in stereo acuity were accepted to take the test. The results of these subjects did not significantly differ from the others' results, i.e., they were not classified as outliers by the recommended methods [ITU12b], which legitimizes their inclusion.

5.3.1.4 Statistical Analysis

The experimental results have to be analyzed statistically. In addition to the usual procedure to report mean values along with 95 % confidence intervals, some more sophisticated methods are used throughout this work. While a detailed description of these methods is beyond the scope of this work, their basic concepts shall be described in this section. Further details can be found in the widely available literature, e.g., by Gałecki and Burzykowski [Gał13].

The within-subjects design is a commonly used approach for subjective experiments: Each stimulus is shown to each subject. The advantage of this approach is that a lot of data can be gathered with a minimum number of subjects, in contrast to the between-group design, where, e.g., stimulus A is presented to subject group A and stimulus B to subject group B. The within-subjects design implies that the regular multiple linear regression approach is unsuitable. Instead, the **linear mixed-effect model (LMEM)** can be used [Gał13], which combines fixed and random effects. While the fixed effects are predictors in the model that are constant across subjects, the random effects vary individually [Gel05]. This approach ensures that a portion of the variability in the data is attributed to individual subject properties rather than relating it all to the fixed effects, which would be the case with regular multiple linear regression. The predictors can be continuous or nominal variables. In the latter case, the predictor is actually split up into multiple dummy-predictors, e.g., one for each stimulus in a set of test stimuli. Sometimes, interactions between predictors may also have an effect on the data, which can also be added to the model.

After fitting the LMEM to the data, a multiple correlation coefficient $R^2 \in \Re$ can be computed, which is interpreted in the same way as the Pearson correlation coefficient r^2: These metrics range from 0 % to 100 % and evaluate how much of the variability in the data can be explained by the predictor(s), i.e., the underlying model. The estimated predictor coefficients allow the researcher to analyze the contribution of each predictor to the data-variability. However, a correlation between the data and a predictor might just be due to random chance (null-hypothesis) instead of an actual statistical dependency (alternative hypothesis). There are numerous statistics for calculating or estimating the probability p that a certain outcome is based on random chance, i.e., under the assumption that the null-hypothesis is true. If that probability is lower than a predefined level of significance, e.g., 5 %, 1 % or 0.1 %, the null-hypothesis is rejected in favor of the alternative hypothesis, which states that there is a statistical dependency. In the context of linear models, the null-hypothesis is usually that a specific predictor is null, whereas the alternative hypothesis states that the true coefficient is not null. For LMEMs, the F-statistic is used to approximate the p-values of each predictor. In order to do so, the degrees of freedom of the F-statistic have to be estimated. A conservative estimation method is the one by Satterthwaite [Gał13], which is used throughout this work. Furthermore, confidence intervals for a defined confidence level, i.e., probability can be computed for each estimated predictor coefficient. The true predictor coefficients are located inside these intervals with the defined probability.

5.3.2 Experiment (I): Determination of Perception and Annoyance Thresholds

This section is based on previous publications by the author [Eic15, Eic16]. An explorative experiment has been conducted to find the circumstances under which the DHIT is perceivable or annoying. There are two parameters controlling the DHIT: the shift budget SB, see **equation (3.11)**, and the shift speed v_{s}. As noted in **section 5.1**, v_{s} is usually chosen heuristically by the stereographer and the shift budget is dictated by the content. An application-oriented DHIT design like this prevents proper orthogonalization of the DHIT parameters, so an interactive test under artificial conditions was constructed. In addition to the two DHIT parameters, it was also hypothesized that the disparity budget of the underlying scene affects DHIT perception.

5.3.2.1 Experimental Setup

The experimental setup described in **section 5.3.1.2** was extended by a black keyboard with labeled keys for the subject interaction.

Table 5.1: Parameters of all test images of experiment (I) after applying a static HIT.

		Shifted disparity (px)		
Sequence	Frame #	Budget	Max.	Min.
EBU "Lupo Hands"	945	46	26	-20
RMIT3DV 46	277	62	39	-23
RMIT3DV 29	956	73	28	-45
Industrial Pump	-	92	42	-50

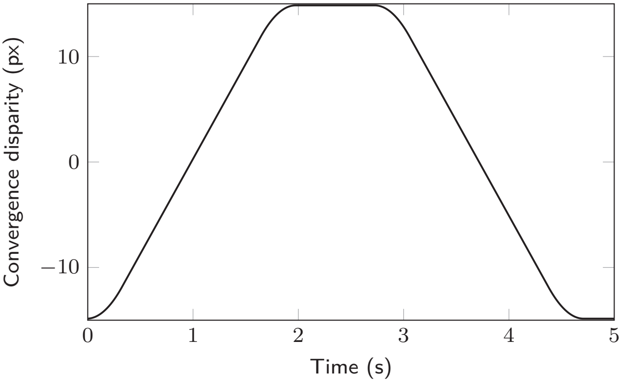


Figure 5.6: Example for the cyclic convergence disparity sequence in experiment (I). Here, the shift speed is $v_{\mathrm{s}} = 0.154\,°/\mathrm{s}$ and the shift budget is $\mathrm{SB} = 30\,\mathrm{px} \triangleq 0.513°$ ($\pm 15\,\mathrm{px}$).

5.3.2.2 Stimuli

For every test stimulus, the S3D views were shifted cyclically back and forth over one of three defined shift budgets SB and at a subject adjustable shift speed v_{s}, which effectively orthogonalizes these two DHIT parameters. The shift budgets were $\mathrm{SB} = \{20\,\mathrm{px}, 30\,\mathrm{px}, 40\,\mathrm{px}\}$, which corresponds to $\mathrm{SB} = \{0.342°, 0.513°, 0.684°\}$ in this experimental setup, see **section 5.3.1.2**. The base material was preconverged by setting a convergence disparity $D_{\mathrm{conv,pre}}$ and the cyclic convergence disparity sequence $D_{\mathrm{conv,cycl}}$ was added. This sequence was essentially a triangle wave with a defined speed and some hold time at peaks and minima, as illustrated in **figure 5.6**. The edges of the cyclic convergence disparity sequences were smoothed a little to avoid abrupt changes in shift speed. The overall convergence disparity was therefore

$$D_{\text{conv}} = D_{\text{conv,pre}} + D_{\text{conv,cycl}} \ . \tag{5.20}$$

The third orthogonalization-parameter was the disparity budget, which is generally determined by the underlying scene and camera setup, specifically the camera baseline. For a true orthogonalization between disparity budget and the other parameters, high resolution multi-baseline material is required, which is rather rare. However, the main reason why such material was not used is that showing only a single test sequence in different conditions risks to bore and tire the subjects. This in turn would reduce the accuracy and reliability of the results. Instead, four different video sequences were used: EBU 3DTV Test sequence "Lupo_Hands" [EBU], RMIT3DV no. 29 and 46 [Che12], and a custom synthetic test image showing an industrial pump. Considering the three shift budget settings, this makes a total of 12 test stimuli. Still images of these videos were shown because they are considered the worst case scenario for DHIT detection since there is nothing distracting the subjects. Some parameters of the used test images are summarized in **table 5.1**. Please note that, while all the preconverged test sequences are contained in the **zone of comfort** of ± 58 px, see **section 3.2.3.1.2**, the addition of the cyclic convergence disparity sequences may lead to protrusion of this limit and therefore possibly create visual discomfort or visual fatigue.

5.3.2.3 Procedure

For each stimulus, the shift speed was initialized to a very annoying setting, which was $v_{\text{s}} = 0.5$ px/frame. This value corresponds to $v_{\text{s}} = 0.513$ °/s in this experimental setup. The subject was asked to lower v_{s} in steps of 0.025 px/frame $\triangleq$ 0.0257 °/s, using the labeled keyboard, until the DHIT was just not deemed annoying anymore. By pressing the respective button, the subject would then confirm that setting, which represents the annoyance threshold. Afterwards, the subject would lower the speed even further until the DHIT was just not perceivable anymore (perception threshold), and move on to the next sequence. The subject was also allowed to increase speed again or correct any vote, if deemed necessary. The 12 stimuli were presented in random order. In the introduction to the test procedure, the subject was asked to explore the screen freely during the test, except for the border regions on the left and right side of the display. These were excluded because the visible frame of the display serves as a reference that makes the DHIT almost always perceivable.

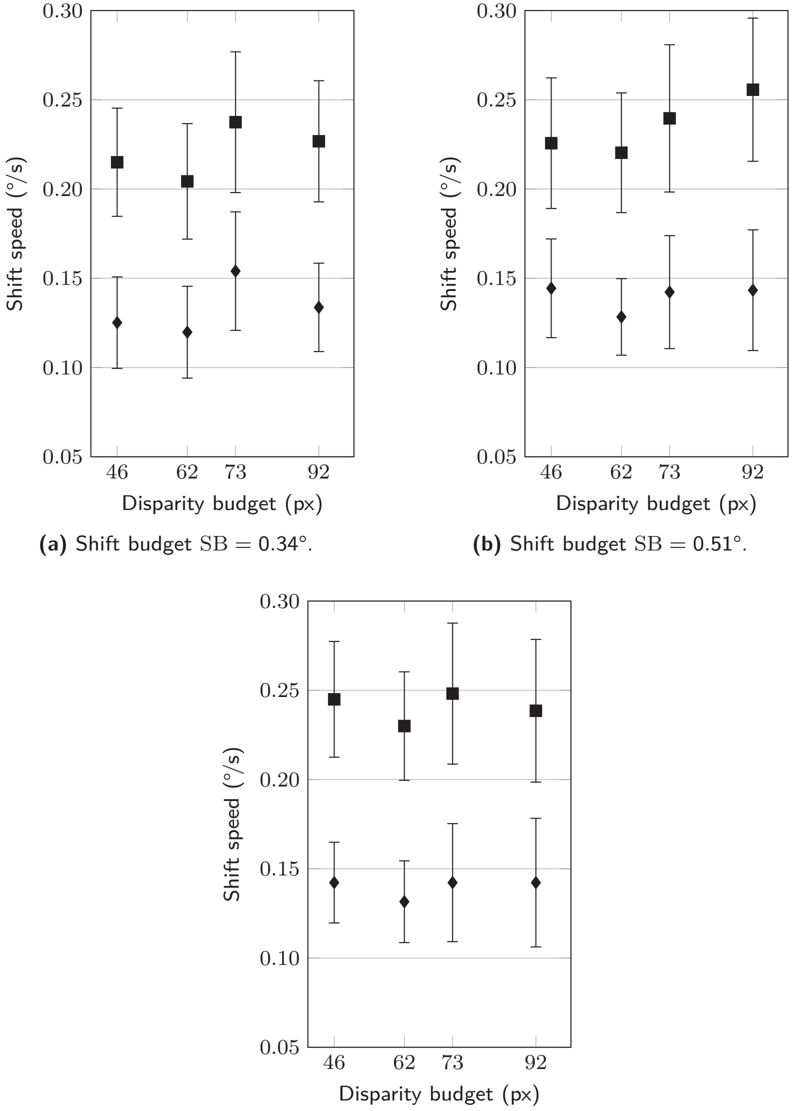
5.3.2.4 Subjects

There were 24 subjects excluding two rejections due to bad examination results in visual acuity or fine stereopsis. Among the accepted subjects, seven had prior experience in the field of subjective experiments, and four were expert viewers. The test group mainly consisted of students and research assistants. The subjects were aged 21 to 31, with an average of 25.04 years. More subject details as well as examination results can be found in **tables A.1 and A.2**.

5.3.2.5 Results

The results are shown in **figure 5.7**. The plots show the perception and annoyance thresholds on shift speed v_s for each shift budget SB as a function of disparity budget D_B. In other words, the abscissas represent the different sequences. The annoyance and perception thresholds were clearly correlated, which is underlined by the high correlation coefficient $r^2 = 68.52\,\%$. The reason for this highly significant correlation is rather simple: One cannot be annoyed by a visual artifact one cannot see.

For further analysis, an LMEM was fitted to the data. Since there was obviously no linear connection between disparity budget D_B and shift speed v_s, see **figure 5.7**, D_B was not used as a continuous predictor for fixed effects. Instead, the sequence name was used as a categorical predictor variable. As for the other fixed effects, the shift budget SB was a continuous predictor, and result type, i.e., annoyance or perception, was another categorical predictor. Furthermore, a random effect of the subjects on the intercept coefficient was added to the model to account for individual differences in this within-subjects experimental design. The fitting operation yielded a very high $R^2 = 79.00\,\%$. The coefficients and their p-values are summarized in **table 5.2**. The annoyance



(a) Shift budget $\mathrm{SB} = 0.34°$.

(b) Shift budget $\mathrm{SB} = 0.51°$.

(c) Shift budget $\mathrm{SB} = 0.68°$.

Figure 5.7: Perception (diamond) and annoyance (square) thresholds of DHIT in experiment (I).

Table 5.2: Estimated linear mixed-effect model parameters with 95 % confidence intervals and p-values in experiment (I) ($R^2 = 79.00\,\%$). The predictors are related to the perception samples of the DHIT for the sequence "Engineering", unless specified differently in the name of the predictor.

Predictor	Coefficient	p-Value
Intercept	$(+0.1232 \pm 0.0301)$ °/s	$2.9 \cdot 10^{-10}$
Seq. (Lupo Hands)	(-0.0071 ± 0.0096) °/s	0.1472
Seq. (RMIT3DV 29)	$(+0.0039 \pm 0.0096)$ °/s	0.4250
Seq. (RMIT3DV 46)	(-0.0176 ± 0.0096) °/s	$3.6 \cdot 10^{-4}$
Shift budget	$(+0.0379 \pm 0.0244)\,\mathrm{s}^{-1}$	0.0024
Result type (Annoyance)	$(+0.0947 \pm 0.0068)$ °/s	$2.2 \cdot 10^{-104}$

and perception thresholds significantly differed with $p < 10^{-103}$. Furthermore, the results for sequence "RMIT3DV 46" significantly differed from the other sequences with $p < 0.001$. This content dependency is observable in **figure 5.7** where each graph exhibits roughly the same value progression. There was also a significant effect of shift budget on the results with $p < 0.01$. Considering the range of shift budgets in this experiment, the effect of this predictor is comparatively low, however. In fact, removing the predictor from the model still yields an $R^2 = 78.64\,\%$ with almost the same coefficients as in **table 5.2**. The mean values of all the data and the 95 % confidence intervals were also calculated. On average, a shift speed of $v_\mathrm{s} = (0.137 \pm 0.008)$ °/s was not perceivable and $v_\mathrm{s} = (0.232 \pm 0.010)$ °/s was not annoying anymore. However, the average annoyance threshold is not really of interest, since in practice it is important to ensure that the DHIT is never deemed annoying by anyone. This makes the lower boundaries of the confidence intervals in **figure 5.7** a lot more interesting, which is as low as $v_\mathrm{s} = 0.171$ °/s for the second sequence "RMIT3DV 46".

A more detailed analysis of the data of each subject revealed a problem, however. The perception and annoyance thresholds were averaged over all sequences for each subject separately and the results are shown in **figure 5.8**[3]. There were obviously strong individual differences in the perception of the DHIT. One subject was annoyed by shift speeds as slow as $v_\mathrm{s} = 0.08$ °/s, but one other extreme example only started to be annoyed by shift speeds as fast as

[3] The graph also shows again the strong correlation between perception and annoyance.

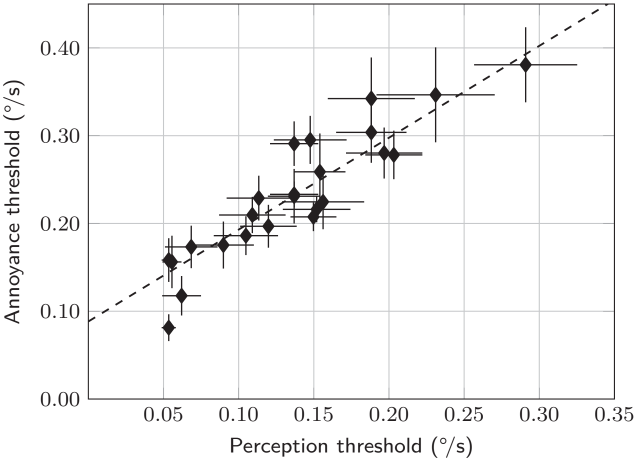


Figure 5.8: Per-subject average annoyance threshold over per-subject average perception threshold in experiment (I) with 95 % confidence intervals in both directions and a linear fit based on raw data to illustrate the strong correlation between annoyance and perception.

$v_\mathrm{s} = 0.38$ °/s. Overall, 76.75 % and 71.76 % of the variability of the perception and annoyance samples, respectively, was due to these individual differences. The identified significant effects on the contrary only explained less than 2 % of the data variability.

5.3.2.6 Discussion

When designing the DHIT, it is important to ensure that nobody in the audience gets annoyed by it. Considering the observed strong individual differences in DHIT sensitivity, this is not as much an easy task as one might assume. Rejecting the extreme outliers in **figure 5.8**, the author of this work recommends to keep the shift speed in the range 0.1 °/s to 0.12 °/s. This recommendation can be applied generally, although a significant effect of shift budget and content was found. However, the impact of these effects on the results was very low compared to the individual differences.

In interviews conducted directly after the test, some subjects said that it was easier to detect DHIT in image regions with big depth discontinuities.

Consequently, the local distribution of depth discontinuities would be a good candidate for further testing. The reason for the heightened sensitivity towards DHIT in these regions could be the unnatural distortion of the depth budget described in **section 3.1.1**. This distortion becomes more distinct the bigger the disparity gradient is. The distortion-free approach DHIT+ is investigated in the next section.

5.3.3 Experiment (II): Comparison of DHIT and DHIT+

In this experiment, the DHIT is compared to the DHIT+, which eliminates the distortion of the depth budget. The experimental setup and procedure is exactly the same as in experiment (I). Annoyance and perception thresholds are measured for both methods rather than just the regular DHIT. CGI is used here because it is such an elaborative undertaking to implement high quality baseline adjustment of natural content, as mentioned in **section 5.2.1.3**. The use of CGI furthermore enables a true orthogonalization of shift budget and disparity budget, in contrast to experiment (I).

5.3.3.1 Stimuli

Once again, the S3D views were shifted cyclically back and forth at a speed that was configurable by the subject. However, this time not only the DHIT but also the DHIT+ was tested. Each test scene should be tested with two shift budgets $\mathrm{SB} = \{15, 30\}\,\mathrm{px} \triangleq \{0.257, 0.513\}\,^{\circ}$ and two disparity budgets $D_{\mathrm{B}} = \{60, 86\}\,\mathrm{px}$. With the two different approaches, this means that subjects would be exposed eight times to each test scene. This risks boring and fatiguing the subjects by showing the same test scenes too often and having the experiment last too long. Instead, only the scene "Computer" was generated with all conditions, while the other test scenes were added for more variety with either one of the disparity budgets and only one shift budget, see **table 5.3**. Hence, all other sequences were only shown twice, once with each approach.

In an effort not to spend too much time on designing visually pleasing 3D test scenes, the S3D capable open source video game "Doom 3" was programmati-

Table 5.3: Parameters of all test sequences in experiment (II).

Sequence name	Disparity budgets (px)	Shift budget (°)
Computer	60, 86	0.257, 0.513
Dungeon	86	0.513
Flying floor	60	0.513
Lava	60	0.513
Workers	86	0.513

cally modified so that still S3D scenes with varying camera baseline could be exported as a video.

5.3.3.1.1 DHIT Stimuli

The DHIT stimuli were designed in the same way as the ones in experiment (I) described in **section 5.3.2.2**. The S3D image pairs were taken from the first frame of the left and right view DHIT+ stimulus video sequences described in the next section. The material was preconverged by setting a convergence disparity $D_{\mathrm{conv,pre}}$ in such a way that the disparity budget was split equally between crossed and uncrossed disparities. This means that any scene with a disparity budget $D_{\mathrm{B}} = 86\,\mathrm{px}$ exhibited shifted disparities of $\tilde{D} = \pm 43\,\mathrm{px}$ in the preconverged state. Again, a cyclic convergence disparity sequence was added. The example in **figure 5.6** shows a cyclic convergence disparity sequence with the maximum shift budget of 30 px (±15 px). For the maximum disparity budget of $D_{\mathrm{B}} = 86\,\mathrm{px}$, this would yield shifted disparities of $\tilde{D} = \pm(43 + 15)\mathrm{px} = \pm 58\,\mathrm{px}$. In this experimental setup, this disparity range corresponds to the zone of comfort as defined by Lambooij et al. [Lam09b], compare **section 3.2.3.1.2**. This was supposed to prevent visual fatigue among the subjects throughout the experiment.

As a final step, each stimulus sequence was inspected for window violations, and appropriate floating windows were manually added to avoid this kind of artifact, compare **section 3.2.3.5.1**. The same floating windows were also used for the DHIT+ stimuli.

5.3.3.1.2 DHIT+ Stimuli

As explained in **section 5.2.1.2**, speed is controlled differently in the DHIT+ in a way that a certain disparity has to be chosen for speed control. Since the aim of both DHIT and DHIT+ is to dynamically adjust convergence in the context of this experiment, it was decided to control the speed of the convergence disparity adjustment. This can be done as follows.

The scaling factors α for the convergence disparity sequence of the DHIT stimuli can be computed using **equation (5.8)**. For the cyclic convergence disparity sequence depicted in **figure 5.6**, this yields something similar to **figure 5.9**. Applying the calculated factors to the convergence disparity from the DHIT stimuli, i.e., **equation (5.20)**, yields

$$D^*_{\mathrm{conv}} = \underbrace{\alpha \cdot D_{\mathrm{conv,pre}}}_{D^*_{\mathrm{conv,pre}}} + \underbrace{\alpha \cdot D_{\mathrm{conv,cycl}}}_{D^*_{\mathrm{conv,cycl}}} \ . \tag{5.21}$$

There is now an additive combination of two temporally dynamic shift sequences. The scaled convergence disparity $D^*_{\mathrm{conv,pre}}$ on its own ensures that objects in the display plane remain in the display plane when adjusting the baseline. In other words, the convergence plane does not move. The addition of $D^*_{\mathrm{conv,cycl}}$ implements the cyclic convergence disparity sequence. For an unbiased comparison between DHIT and DHIT+, the cyclic convergence disparity sequences should be equal. However, this is obviously not the case here, since $D^*_{\mathrm{conv,cycl}} = \alpha \cdot D_{\mathrm{conv,cycl}} \neq D_{\mathrm{conv,cycl}}$. The solution to this problem is to compute new scaling factors for an alternative cyclic convergence disparity sequence $D_{\mathrm{conv,alt}}$. This sequence can be found through numeric optimization methods such that it satisfies $\alpha \cdot D_{\mathrm{conv,alt}} = D_{\mathrm{conv,cycl}}$, i.e., the resulting scaled, cyclic convergence disparity sequence is equal to the original unscaled sequence. **Figure 5.9** depicts the final sequence of scaling factors. This yields the corrected scaled convergence disparity sequence

$$D^*_{\mathrm{conv,corr}} = \alpha \cdot D_{\mathrm{conv,pre}} + \underbrace{\alpha \cdot D_{\mathrm{conv,alt}}}_{D_{\mathrm{conv,cycl}}} \ . \tag{5.22}$$

Every shift speed setting v_{s} yields a different sequence of scaling factors

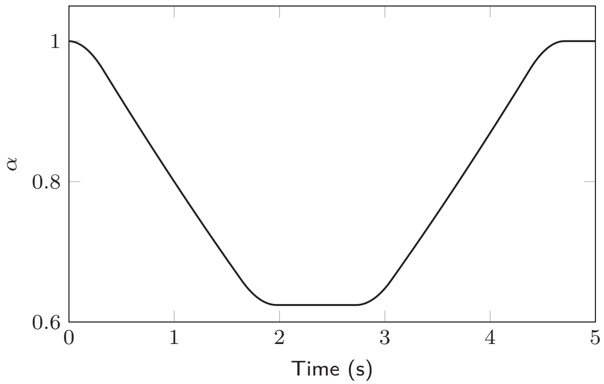
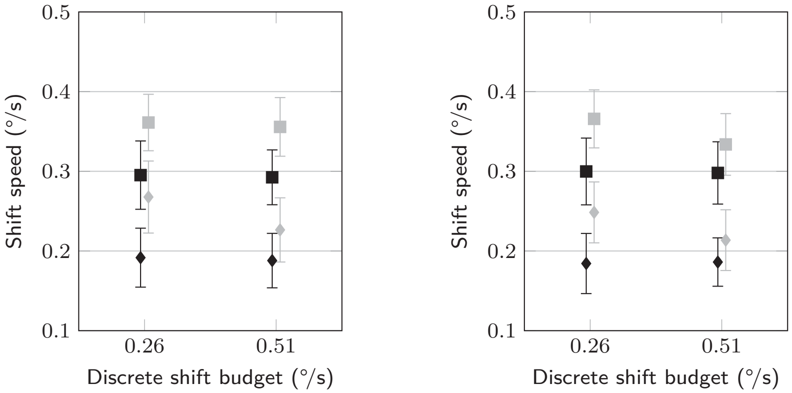


Figure 5.9: Example for a final sequence of linear scaling factors for the DHIT+ in experiment (II).

and corrected scaled convergence disparities $D^*_{\mathrm{conv,corr}}$. Hence, for each v_s, the scaling factors were applied to the camera baseline of a test scene and exported as a video sequence along with the respective convergence disparities $D^*_{\mathrm{conv,corr}}$. The scaling factors were computed using $Z_\mathrm{targ} = Z_\mathrm{min}$ and the average human eye baseline of $b_\mathrm{e} = 6.3\,\mathrm{cm}$ [Dod04, War95]. The eye baseline varies individually, which affects the perceived depth. However, as already mentioned in **section 5.2.1.3**, the depth distortion due to an eye baseline mismatch is neglectable compared to that of the regular DHIT.

During a trial, the videos had to be played repeatedly until the subject was done with the task. The hold time at the end of each cycle, see **figures 5.6 and 5.9**, was used to slowly blend the last video frame into the first. Thereby, sudden (dis-)appearance of some very subtle coding artifacts was avoided. Aside from these coding artifacts, the material generated by "Doom 3" also exhibited some subtle aliasing at vertical and horizontal object borders, a common problem of CGI. This aliasing was temporally dynamic due to the baseline adjustment, and its visibility varied from scene to scene. There were also some software-internal issues that lead to some objects appearing to jitter slightly in depth during the baseline adjustment.



(a) Disparity budget $D_{\mathrm{B}} = 60\,\mathrm{px}$.

(b) Disparity budget $D_{\mathrm{B}} = 86\,\mathrm{px}$.

Figure 5.10: Annoyance (square) and perception thresholds (diamond) for DHIT (black) and DHIT+ (gray) with 95 % confidence intervals for orthogonalized sequence "Computer" in experiment (II).

5.3.3.2 Subjects

Of the 31 subjects, 28 were accepted to take the test. The accepted subjects were aged 22 to 31 with an average of 24.89 years. Six of them were expert viewers, and 13 had prior experience in subjective evaluations. The test group mainly consisted of students and some research assistants. More subject details as well as examination results can be found in **tables A.1 and A.2**.

5.3.3.3 Results

The results for the sequence "Computer" are displayed in **figure 5.10**. This test scene was the only one which was presented in two different disparity budgets and two different shift budgets, i.e., where a true orthogonalization between these two parameters was given. This was not possible in the previous experiment. Once again, an LMEM was fitted to the data. Since disparity budget D_{B} and shift budget SB were both continuous variables this time, they were used as continuous predictors for fixed effects. The method, i.e., DHIT or DHIT+, and the result type, i.e., annoyance or perception, were

Table 5.4: Estimated linear mixed-effect model parameters with 95 % confidence intervals and p-values for sequence "Computer" in experiment (II) ($R^2 =$ 76.34 %). All predictors are related to the perception samples for DHIT, unless specified differently in the name of the predictor.

Predictor	Coefficient	p-Value
Intercept	$(+0.1883 \pm 0.0383)$ °/s	$5.6 \cdot 10^{-15}$
Shift budget	$(-0.0001 \pm 0.0010)\,\mathrm{s}^{-1}$	0.8335
Method (DHIT+)	$(+0.0949 \pm 0.0335)$ °/s	$4.7 \cdot 10^{-8}$
Result type (Annoyance)	$(+0.1120 \pm 0.0106)$ °/s	$2.2 \cdot 10^{-66}$
Shift budget × Method (DHIT+)	$(-0.0018 \pm 0.0014)\,\mathrm{s}^{-1}$	0.0133

categorical predictors for fixed effects. The results for DHIT+ seem to be correlated with SB, which can be modeled as a two-way interaction between SB and the method predictor. As a first explorative step, all other possible two-way interactions were also added to the model. Additionally, a random effect of the subjects on the intercept coefficient was modeled to account for individual differences. This yields a high $R^2 = 76.60\,\%$ and no significant contribution of D_{B} or SB. The annoyance and perception thresholds obviously differed significantly, as did the DHIT from the DHIT+. The only significant interaction was the one between SB and DHIT+. So, a reduced model containing only the significant predictors[4] was fitted to the data, yielding again a high $R^2 = 76.34\,\%$. The coefficients and p-values are summarized in **table 5.4**. The DHIT+ significantly reduced annoyance and perception with $p < 10^{-7}$. The results for the DHIT were not significantly affected by SB. For the DHIT+ on the other hand, there was a significant effect of SB with $p < 0.05$. Despite this significant effect, the mean values and their associated 95 % confidence intervals over all conditions of the sequence "Computer" were calculated, which are summarized in **table 5.5**. This table clearly shows how DHIT+ reduces perception and annoyance.

Figure 5.11 depicts the results of all stimuli with a shift budget of 0.513° for both disparity budgets. While the DHIT still exhibited little variation in the perception and annoyance thresholds for all sequences and disparity budgets,

[4] The predictors are: method, result type and the interaction between SB and method. The insignificant predictor SB also had to be added to the model because it is used in the interaction.

Table 5.5: Mean perception and annoyance thresholds with 95 % confidence intervals for sequence "Computer" in experiment (II).

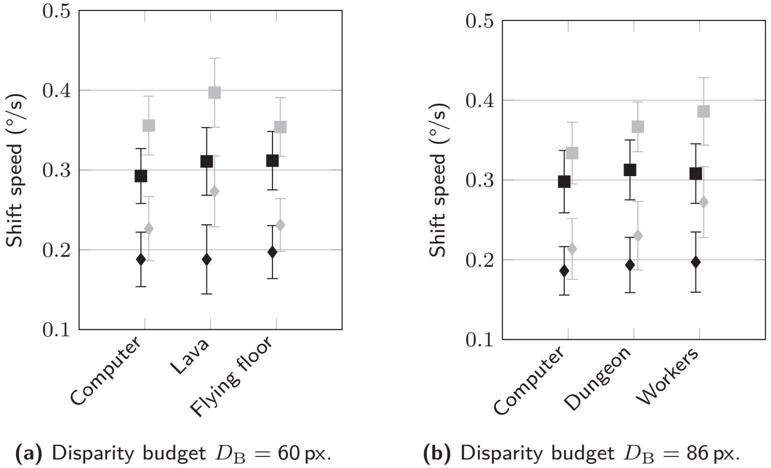
Method	Perception (°/s)	Annoyance (°/s)
DHIT	0.187 ± 0.017	0.296 ± 0.019
DHIT+	0.239 ± 0.020	0.354 ± 0.018

Table 5.6: Estimated linear mixed-effect model parameters with 95 % confidence intervals and p-values for all sequences exhibiting a shift budget of 0.513° in experiment (II) ($R^2 = 77.47\,\%$). The intercept predictor represents the perception samples of the sequence "Computer" at disparity budget $D_\mathrm{B} = 60$ px for the DHIT. All other predictors represent the difference to the intercept under different conditions.

Predictor	Coefficient (°/s)	p-Value
Intercept	$+0.1806 \pm 0.0339$	$7.4 \cdot 10^{-14}$
Seq. (Lava)	$+0.0092 \pm 0.0210$	0.3926
Seq. (Flying floor)	$+0.0142 \pm 0.0210$	0.1853
Seq. (Computer D86)	$+0.0018 \pm 0.0210$	0.8642
Seq. (Dungeon)	$+0.0128 \pm 0.0210$	0.2315
Seq. (Workers)	$+0.0124 \pm 0.0210$	0.2486
Method (DHIT+)	$+0.0509 \pm 0.0210$	$2.5 \cdot 10^{-6}$
Result type (Annoyance)	$+0.1192 \pm 0.0086$	$2.7 \cdot 10^{-109}$
Seq. (Lava) × Method (DHIT+)	$+0.0348 \pm 0.0298$	0.0218
Seq. (Flying floor) × Method (DHIT+)	-0.0128 ± 0.0298	0.3974
Seq. (Computer D86) × Method (DHIT+)	-0.0193 ± 0.0298	0.2044
Seq. (Dungeon) × Method (DHIT+)	-0.0055 ± 0.0298	0.7168
Seq. (Workers) × Method (DHIT+)	$+0.0257 \pm 0.0298$	0.0908

Table 5.7: Mean perception and annoyance thresholds with 95 % confidence intervals over all sequences in experiment (II). Additionally, the results of experiment (I) are included.

Method	Perception (°/s)	Annoyance (°/s)
DHIT	0.191 ± 0.014	0.304 ± 0.013
DHIT+	0.245 ± 0.015	0.365 ± 0.012
DHIT (Exp. I)	0.137 ± 0.008	0.232 ± 0.010



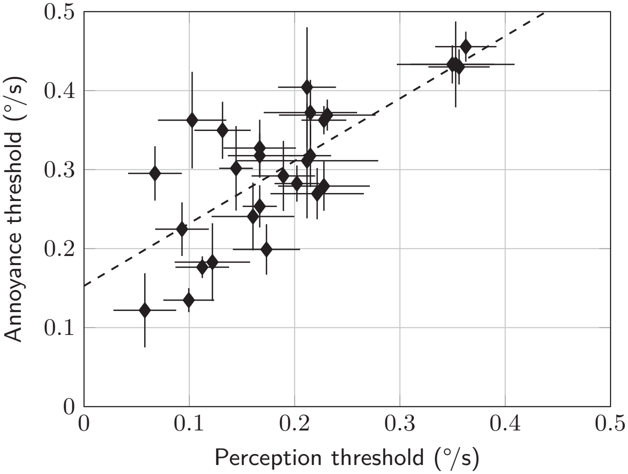
(a) Disparity budget $D_{\mathrm{B}} = 60\,\mathrm{px}$.

(b) Disparity budget $D_{\mathrm{B}} = 86\,\mathrm{px}$.

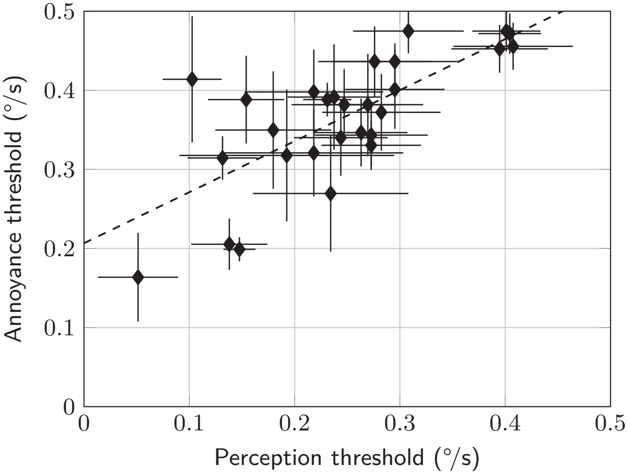
Figure 5.11: Annoyance (square) and perception thresholds (diamond) for DHIT (black) and DHIT+ (gray) with 95 % confidence intervals for all sequences exhibiting a shift budget of 0.513° in experiment (II).

there was some variation in the results for the DHIT+. An LMEM with the categorical predictors sequence, method and result type as well as all possible two-way interactions was fitted to the data. The model also included a random effect of the subjects on the intercept coefficient to account for individual differences and yielded $R^2 = 77.57\,\%$. The interactions between sequence and result type as well as method and result type were insignificant, which lead to a reduced model with only one interaction between sequence and method, yielding $R^2 = 77.47\,\%$. The coefficients and p-values are summarized in **table 5.6**. Again, the DHIT+ significantly reduced perception and annoyance with $p < 10^{-5}$. As expected from **figure 5.11**, the effect of the sequences on the DHIT results is insignificant. For the DHIT+, however, the sequence "Lava" differed significantly with $p < 0.05$. Despite this significant effect, the perception and annoyance thresholds were also averaged over all available data, see **table 5.7**. This shows again the improvement of the DHIT+ over the regular DHIT.

Most of the variability in the data originates once again from individual differences, as can be seen in **figure 5.12**. In this figure, the average annoyance



(a) DHIT results.



(b) DHIT+ results.

Figure 5.12: Per-subject average annoyance threshold over per-subject average perception threshold in experiment (II) with 95 % confidence intervals in both directions and a linear fit based on raw data to illustrate the strong correlation between annoyance and perception.

thresholds for each subject are plotted over their average perception thresholds along with 95 % confidence intervals for both parameters. The sensitivity varied strongly between individuals. Within individuals however, there was a strong correlation between DHIT and DHIT+ with $r^2 = 89.23\,\%$ and $r^2 = 86.81\,\%$ for perception and annoyance data, respectively. Furthermore, a significant correlation between perception and annoyance was again observed. The correlation coefficients were $r^2 = 53.34\,\%$ and $r^2 = 49.43\,\%$ for DHIT and DHIT+, respectively.

5.3.3.4 Discussion

For comparison, **table 5.7** also includes the results of experiment (I). The previously measured perception and annoyance thresholds for the DHIT were considerably lower than those of the current experiment. This might be due to the observed strongly varying individual differences in the perception of DHIT or DHIT+: the subjects having participated in the last experiment might just be more sensitive to DHIT-like processing. It is also possible that some content characteristic is responsible for this difference because natural video content was used in the last experiment, whereas CGI from a video game was used here. Since the true origin of this mismatch in the results is unknown, a conservative recommendation is to keep the shift speed for the DHIT in the range 0.10 °/s to 0.12 °/s, as of experiment (I). With the DHIT+, the convergence disparity can be set about 50 % faster, i.e., the shift speed can be increased by 0.05 °/s due to its significantly reduced annoyance-potential.

In this experiment, disparity budget and shift budget could be truly orthogonalized and no significant effect of disparity budget was found. Surprisingly, there was no significant effect of shift budget (or content) for the DHIT either, in contrast to experiment (I). The magnitude of these effects in experiment (I) was comparatively low, which might be the reason for their vanishing. For the DHIT+, however, both shift budget and content contributed significantly to the results. Since the effect of disparity budget was found to be insignificant, the content dependency was likely due to some other content characteristic. The aliasing and depth-jitter issues described in **section 5.3.3.1**, whose visibility varies from sequence to sequence, might be the originators here. The existence of these artifacts in the test material unfortunately prevents any conclusions about the distortion of depth affecting DHIT perception. However, the important conclusion is that the new approach DHIT+ allows for faster convergence disparity adjustment, i.e., it significantly reduces annoyance and perception even though it was tested with such erroneous test material. The artifacts could also be the explanation for the dependency on shift budget since a longer shift allows more artifacts to show up.

5.3.4 Experiment (III): Effect of Gaze Adaptivity

The scope of this experiment was twofold. Firstly, the effect of GACS3D on the perception of DHIT should be investigated. Secondly, while the DHIT was designed completely independent of the underlying scene structure in the first two experiments, it was designed in a more application-oriented way here: Different interesting objects were slowly brought into the ZPS. Thereby, practical limits for the shift speed should be found. This section is based on previous publications by the author [Eic15, Eic16].

5.3.4.1 Experimental Setup

The experimental setup was the same as described in **section 5.3.1.2**. Additionally, a questionnaire was placed on the table for the subject to fill out.

5.3.4.2 Stimuli

The aim of this experiment was to evaluate how different shift speeds v_s are perceived in a gaze adaptive DHIT design compared to the regular design. In order to truly isolate the effect of gaze adaptivity, the presented DHIT sequence must be the same in both conditions. Furthermore, the DHIT sequence must be predetermined for each stimulus, so that results are comparable between subjects. These restrictions prevented the actual usage of GACS3D and called for an emulation of gaze adaptivity. Again, still images were used as the base material for the same reason as in the first experiments: It is considered the worst case scenario for DHIT detection. Every test image was presented under three different conditions:

1. “GACS3D”

   A stereoscopic pointer was shown at a deterministic series of different interesting locations for 0.5 s to 1.5 s each and wherever the pointer was, the ZPS was slowly established through DHIT at a defined maximum speed $v_{\mathrm{s,max}}$. The subject was asked to always fixate on that pointer, which effectively emulates GACS3D by replacing the eye tracker with deterministic pointer locations. The pointer strongly attracts attention, because it is the only moving object on those still images. In order to truly synchronize eye movements and DHIT, the human saccadic reaction time has to be compensated for. As mentioned in **section 2.1**, it is 240 ms on average [Joo03] but can be a lot faster [Gez97]. Because of that, the DHIT was delayed by 150 ms, which was judged most natural by a small test group.

   The DHIT sequence generated by this test condition was used in all other test conditions as well. The only difference is the presentation of the pointer.

2. “Movie”

   Here the pointer was simply hidden and the subject was allowed to explore the screen freely, just like when watching a movie. The results of this condition serve as a reference because the condition is similar to experiments (I) and (II).

3. “Control”

   The effect of gaze adaptivity can be analyzed by comparing the results of the first two conditions. However, the visible pointer might pose a distraction that alters the results. So, in order to check the validity of that comparison, a “Control” condition was introduced: A pointer was shown at a different deterministic series of locations that did not actually converge. Hence, the DHIT was not gaze adaptive just like in the “Movie” condition, while a pointer was shown. If the “Control” condition yields approximately the same results as the “Movie” condition, it is safe to say that the effect of the visible pointer is neglectable and any differences in the results of the "Movie" and "GACS3D" conditions can be attributed to gaze adaptivity.

There were 26 still images accompanied by estimated disparity maps exhibiting disparity budgets ranging from 14 px to 92 px. The images were taken from the EBU [EBU] and RMIT3DV [Che12] S3D test sequence libraries in addition to the industrial pump test image, which has already been used in experiment (I). Each test image was assigned one of seven maximum DHIT speed values equally distributed on the interval 0.125 px/frame to 0.5 px/frame, which is equivalent to 0.128 °/s to 0.513 °/s in this experimental setup. The mean shift speed is dependent on the disparity budget of a scene, the pointer locations and the maximum shift speed setting $v_{\mathrm{s,max}}$. The series of pointer locations and $v_{\mathrm{s,max}}$ were chosen in such a way that maximum and mean shift speed were approximately of the same order of magnitude. This constraint and the limited available test material resulted in unequal numbers of stimuli per maximum shift speed setting and, furthermore, a slight correlation between shift speed an disparity budget. The details about the stimuli are summarized in **table A.3**.

5.3.4.3 Procedure

The test images were rated under the three conditions independently in random order in a non-interactive single stimulus impairment scale test [ITU12a], where the subject was instructed to rate their perception of the DHIT. A discrete scale was used with labels translated to German: "5: imperceptible", "4: perceptible, but not annoying", "3: slightly annoying", "2: annoying", "1: very annoying". Again, the subject was instructed not to rate a sequence while watching at the border of the display, because the DHIT is always perceivable there. The first five ratings were neglected as recommended in [ITU12b]. In order to familiarize the subject with the procedure and the DHIT, some anchor sequences were shown before the test.

5.3.4.4 Subjects

There were 17 subjects allowed to participate in this experiment. They were aged 23 to 30 years, 24.65 years on average. All of the subjects were students,

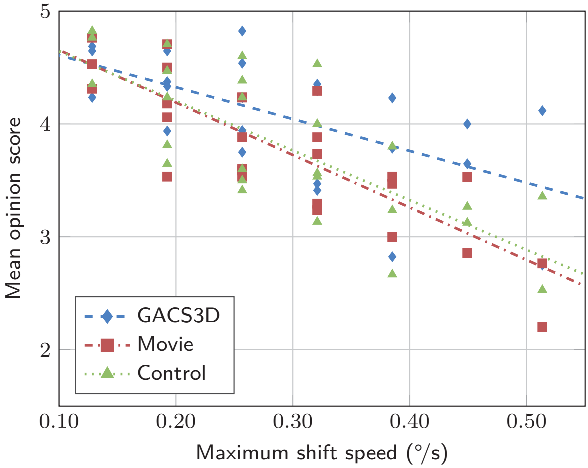


Figure 5.13: Mean opinion scores for each sequence over maximum shift speed in all three conditions in experiment (III) along with the LMEM-fit.

but two of them are to be considered expert viewers, who also had prior experience in subjective experiments. There were two rejections due to bad examination results in visual acuity or fine stereopsis. Further details can be found in **tables A.1 and A.2**.

5.3.4.5 Results

As mentioned before, one of seven maximum shift speed settings was assigned to each test sequence so that there are multiple sequences per maximum shift speed. This speed and the GACS3D pointer location determined the mean shift speed of the sequence. Although the **mean opinion scores (MOSs)** of the sequences correlated a little bit better with mean shift speed, the MOSs are plotted over maximum shift speed in **figure 5.13** because the mean shift speed is of limited interpretability and applicability to a stereographer[5]. For a more detailed analysis, an LMEM was fitted to the data. The model consisted of the continuous predictor $v_{\mathrm{s,max}}$ and the categorical predictor condition, i.e., “GACS3D”, “Movie” or “Control”. The interaction of these predictors

[5] This leads to the discrete and multivalent distribution of samples in **figure 5.13**.

Table 5.8: Estimated linear mixed-effect model parameters with 95 % confidence intervals and p-values in experiment (III) ($R^2 = 44.31\,\%$). All predictors are related to the "Control" condition, unless specified differently in the name of the predictor.

Predictor	Coefficient	p-Value
Intercept	$+5.086 \pm 0.305$	$1.5 \cdot 10^{-20}$
Condition (GACS3D)	-0.195 ± 0.303	0.2054
Condition (Movie)	$+0.034 \pm 0.318$	0.8308
Max. speed	$(-4.402 \pm 1.140)\,\mathrm{s/°}$	$1.3 \cdot 10^{-7}$
Condition (GACS3D) × Max. speed	$(+1.579 \pm 1.047)\,\mathrm{s/°}$	0.0036
Condition (Movie) × Max. speed	$(-0.248 \pm 1.216)\,\mathrm{s/°}$	0.6794

Table 5.9: Annoyance and perception thresholds in experiment (III).

Condition	Perception (°/s)	Annoyance (°/s)
GACS3D	0.138	0.493
Movie	0.133	0.348
Control	0.133	0.360

was also added to the model in order to find out whether the perception of "GACS3D" relates differently to $v_{\mathrm{s,max}}$ compared to the other two conditions. The results of the first two experiments revealed strong individual differences in the relation between shift speed and perception. Because of that, a random effect of subjects on the whole fixed effect portion of the model was added, which finally yielded $R^2 = 44.31\,\%$. The model parameters and p-values are summarized in **table 5.8**.

As expected, there was a significant effect of $v_{\mathrm{s,max}}$ with $p < 10^{-6}$. The linear functions of this model are plotted in **figure 5.13** for each condition. The lines for "Movie" and "Control" are almost the same. This also shows in **table 5.8** where any prediction with "Condition (Movie)" was highly insignificant with $p > 0.65$. Hence, these two conditions were declared as approximately equal, which rendered the comparison of the conditions "GACS3D" and "Movie" legitimate. The line of "GACS3D" is a lot more flat-angled. This can also be seen in **table 5.8** in the form of a significant interaction with $v_{\mathrm{s,max}}$ with $p < 0.01$.

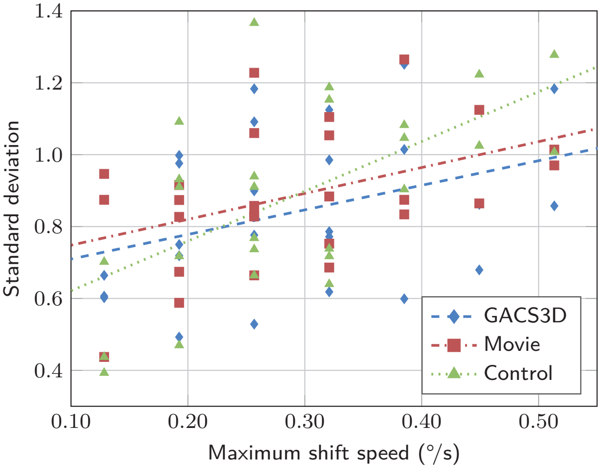


Figure 5.14: Standard deviation of scores for each sequence over maximum shift speed in all three conditions in experiment (III) along with linear fits for each condition.

It is possible to derive annoyance and perception thresholds from the LMEM. The transition from "imperceptible" to "perceptible, but not annoying" at score 4.5 represents the perception threshold and the next transition to "slightly annoying" at score 3.5 represents the annoyance threshold. Rearranging the linear equations of the LMEM to find the critical speeds for these scores yielded the average thresholds summarized in **table 5.9**. While the perception thresholds are almost the same, the annoyance threshold of "GACS3D" is increased a lot compared to the other conditions.

The standard deviation of scores per sequence is plotted in **figure 5.14**. It was slightly correlated to $v_{\mathrm{s,max}}$ with $r^2 = 12.53\,\%$ and $r^2 = 17.99\,\%$ for "GACS3D" and "Movie", respectively.

5.3.4.6 Discussion

The main finding of this experiment was that GACS3D significantly positively interacted with maximum shift speed $v_{\mathrm{s,max}}$. It is concluded that GACS3D reduces annoyance, compare **table 5.9**.

As mentioned before, the “Movie” condition was supposed to serve as a reference for comparison to the previous experiments. The perception threshold in **table 5.9** strongly agreed with the results from experiment (I), but the annoyance threshold was increased, compare **table 5.7**. In fact, it was almost as high as the annoyance threshold of DHIT+. This increase might be due to the different experiment design and the type of DHIT. The DHIT occurred only sporadically and over comparatively short time periods in this experiment.

The multiple correlation coefficient $R^2 = 44.31\,\%$ of the LMEM was rather low compared to the first two experiments. It is believed that this was due to two reasons. One the one hand, the design of the DHIT had previously been very strict and homogeneous, whereas it was rather sporadic in this experiment. Hence, $v_{\mathrm{s,max}}$ might just not be a very good predictor. On the other hand, rating impairments is inherently more difficult than simply adjusting shift speed. Hence, in this experiment, more error variance was given.

The slight correlation between the standard deviation of the per-sequence scores and $v_{\mathrm{s,max}}$ could be due to individual differences in DHIT sensitivity, which was already observed in experiments (I) and (II), see **figures 5.8 and 5.12**. For very slow $v_{\mathrm{s,max}}$ values, the stimuli are likely rated imperceptible by all subjects, whereas for faster stimuli, the perception of the DHIT individually differs.

5.3.5 Experiment (IV): Evaluation of the Full GACS3D Prototype

The proposed approach GACS3D is supposed to reduce visual fatigue. Evaluating visual fatigue directly would require prolonged viewing sessions, which would be a very elaborate and ambitious undertaking. Instead, visual discomfort was evaluated as an indicator for visual fatigue in a pair comparison test, similarly to what Hanhart et al. did [Han14]. This section is based on a previous publication by the author [Eic16].

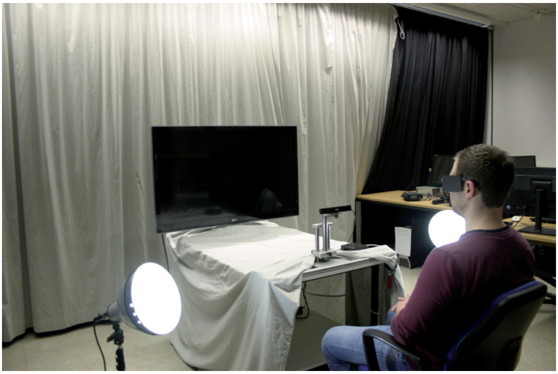


Figure 5.15: Experimental setup in experiment (IV).

5.3.5.1 Experimental Setup

The basic experimental setup was the same as in **section 5.3.1.2**. However, some modifications were necessary because of the utilized eye tracker. The eye tracker was placed on the table on a custom-made stand that enables a precise eye tracker calibration at this comparatively high operating distance of $d = 3.1H \approx 180$ cm. Due to the passive polarized S3D glasses, the eye tracker performance was degraded. Because of that, the subject was illuminated by two spotlights positioned left and right of the table, as can be seen in **figure 5.15**. In order not to blind the subject with those spotlights, the background illumination was increased by enabling the ceiling lighting of the laboratory. Because of that, the light density of the background was approximately 45 cd/m^2 and its color was CIE $(x, y) = (0.39, 0.40)$. Hence, this experimental setup was not compliant with the respective ITU recommendations [ITU12a, ITU12b]. Prior to the test, subjects were asked whether the spotlights blinded them and none affirmed. Furthermore, black blinders were added on both sides of the glasses to prevent visible reflections on the back-facing side of the filter glasses.

5.3.5.2 Stimuli

Since the final prototype should be evaluated in this experiment, moving sequences from the EBU [EBU], NAMA3DS1 [Urv12] and RMIT3DV [Che12]

Table 5.10: Parameters of all test video sequences of experiment (IV).

			Disparity (px)			
Sequence	ID	Frames	Max.	Min.	ZPS	FW
EBU "Lupo Hands"[a]	1	301[a]-600[a]	46	-36	-24	20
NAMA3DS1 "Umbrella"	2	76-325	19	-39	-23	20
RMIT3DV 02	3	1-300	31	0	15	22
RMIT3DV 29	4	951-1250	45	-28	19	23
RMIT3DV 43	5	1-300	0	-69	-30	35
RMIT3DV 46	6	251-550	-9	-72	-40	30

[a] This sequence is actually available in 50 frames/s, but was downsampled to 25 frames/s so that it has the same frame rate as the other sequences.

stereo 3d test sequence libraries were used, rather than still images. The six videos were presented under three different conditions:

1. "Raw"

   No HIT was applied to the stereo views. The scene was presented "as is". Please note that all the material was already pre-converged in some way.

2. "ZPS"

   A static HIT was applied to establish a certain ZPS. The convergence disparity was manually chosen by an expert with the aim to minimize visual discomfort while simultaneously generating a visually pleasing depth sensation.

3. "GACS3D"

   The full prototype as described in **section 5.2.2** was used to realize the gaze adaptive DHIT. The views were shifted at a maximum speed of $v_{\mathrm{s,max}} = 0.12\,°/\mathrm{s}$, which was the recommended shift speed as of experiment (I). The DHIT was updated at a rate of 60 Hz, in accordance with the previous experiments.

The video sequences are available in 25 frames/s. In order to avoid any motion judder, the sequences were played at an increased speed of 30 frames/s, so that every frame was played twice on the 60 Hz display. The resulting test sequences still looked natural at this increased speed and were 10 s long. Details on the video sequences are summarized in **table 5.10**[6].

To avoid window violations, floating windows were added to each stimulus. Instead of the automated floating window approach of GACS3D described in **section 5.2.2.4**, a static floating window was applied in order to avoid distraction of the subjects by temporal changes in floating window disparity. This was achieved by cropping the views on both sides in such a way that no content is shifted out of the display area and manually setting the floating window to a fixed disparity for each sequence. This decreases the width of the stereoscopic window, but completely removes window violations in a static fashion. Furthermore, retinal rivalries due to the visible display frame, as described in **section 3.2.3.5.1**, are avoided in this way. The floating window disparities can be found in **table 5.10**, alongside the disparities for the ZPS of the second condition. The listed floating window disparities can be a lot bigger than the respective maximum disparities, because the floating window must be able to eliminate window violations even for the most extreme convergence disparities, i.e., when the background is looked at with GACS3D so that the whole scene is shifted in front of the display.

5.3.5.3 Procedure

The visual discomfort of the conditions was evaluated using the pair comparison method [ITU12a] in a simple preference judgment ("A is better", "equal", "B is better"). A graded scale was not necessary because the differences between the conditions were hardly perceivable. In an effort to help the subject to see the subtle differences, each condition was repeated once per trial (A-B-A-B). In the case of "GACS3D", this means that the DHIT applied to the sequence was not the same in both playthroughs due to the gaze adaptivity. However, this ensures that it is actually GACS3D that is being rated and not some random DHIT sequence. After stimulus exposure, the subject would simply speak out the rating of the current comparison. All condition combination pairs were tested ("Raw" vs. "ZPS", "Raw" vs. "GACS3D", "ZPS" vs. "GACS3D"), but

[6] Please note that some values may differ from those in **table 5.1** because an HIT was applied there and only a single frame was shown, which might exhibit a different disparity budget than the whole scene.

not in both possible orders to keep the required time acceptable. Instead, the order of each combination was randomized per subject.

Prior to the test, the eye tracker was calibrated individually and some anchor sequences were shown to familiarize the subject with the concept of visual discomfort related to excessive AVD. The Subject was instructed to sit still during a trial, so that proper eye tracker performance was ensured. Between trials, a window showing the position of the eyes was presented to the subject to ensure that the optimal tracking position was maintained throughout the whole test. In order to reduce subject movement to a minimum, the ratings were furthermore collected by an operator, who also triggered the start of the next trial.

5.3.5.4 Subjects

There were 30 accepted subjects in this experiment. Five subjects wearing optical aids had to be rejected due to bad eye tracking performance. Because of that, there was a certain preference for people without optical aids. Additionally, one subject was rejected due to low visual acuity. The accepted subjects were aged 22 to 36, 26.27 years on average. The subjects were mostly students and research assistants. There were two expert viewers and one subject had prior experience in subjective evaluations. Further details on the subjects and some examination results can be found in **tables A.1 and A.2**.

5.3.5.5 Results

The eye tracker data of all sequences involving eye tracking was inspected manually because the eye tracker still occasionally yielded bad results for some subjects. In that case, individual scores of problematic sequences were rejected, which lead to a reduced number of 15 to 25 samples per stimulus comparison where “GACS3D” is presented, as can be seen in **figure 5.16**.

The results of the experiment are displayed in **figure 5.17** in the form of histograms for each comparison. Here, the relative frequencies of specific ratings are plotted for all sequences separately and globally. As can be seen, “equally good” is picked most often for almost all stimulus comparisons. Globally,

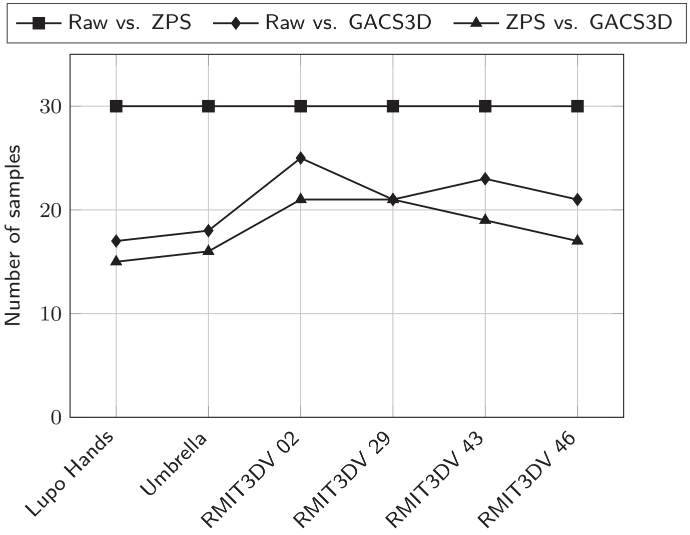
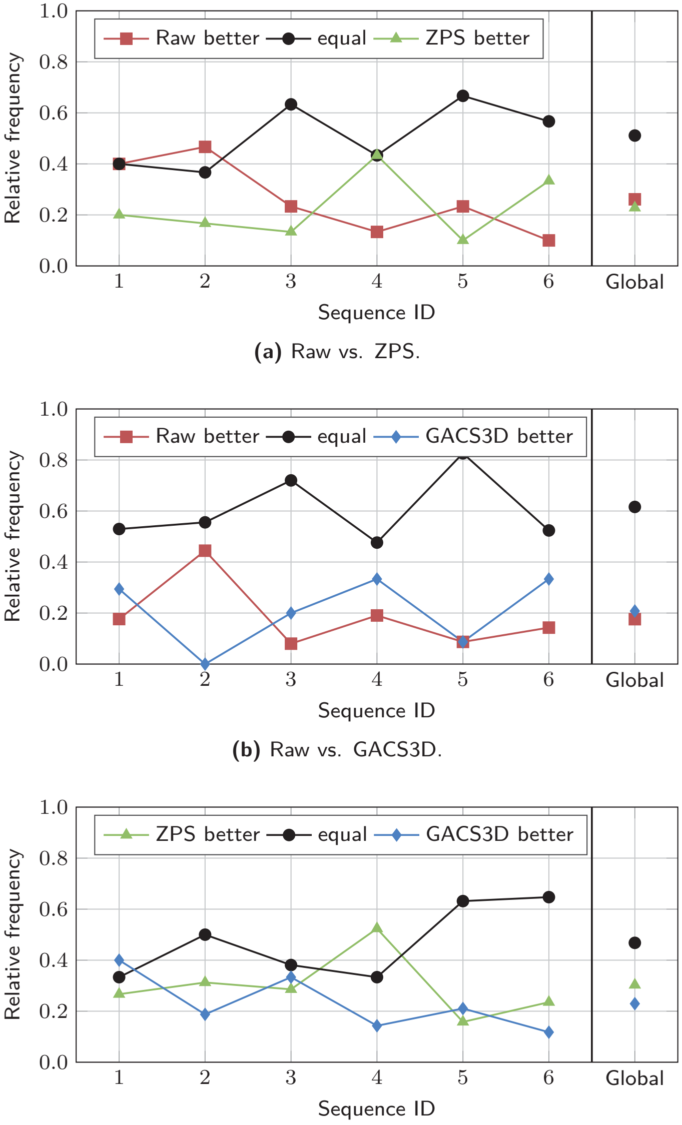


Figure 5.16: Number of samples in each comparison in experiment (IV).

“GACS3D” was rated slightly better than “Raw”, but worse than “ZPS”. However, the differences are neglectable compared to the relative frequency of “equally good”. The standard deviations of the comparison scores are displayed in **figure 5.18**. These values are very high, considering that the comparison ratings correspond to scores of {-1,0,1}. There were also no subjects exhibiting a clear systematic preference for any condition.

5.3.5.6 Discussion

None of the tested conditions yielded significantly differing results in this experiment. This outcome had already been expected because of the subject interviews conducted directly after the test. Most subjects said that it would look all the same or that they tried to concentrate on details because they could not tell the difference between the stimuli. However, all of them affirmed a strong visual discomfort when an anchor sequence with big AVDs was shown prior to the test. The results of the “Raw” vs. “ZPS” comparison should be most reliable, since no individual scores had to be rejected, but this comparison



(a) Raw vs. ZPS.

(b) Raw vs. GACS3D.

(c) ZPS vs. GACS3D.

Figure 5.17: Relative frequencies of ratings in each comparison for all sequences separately and globally in experiment (IV).

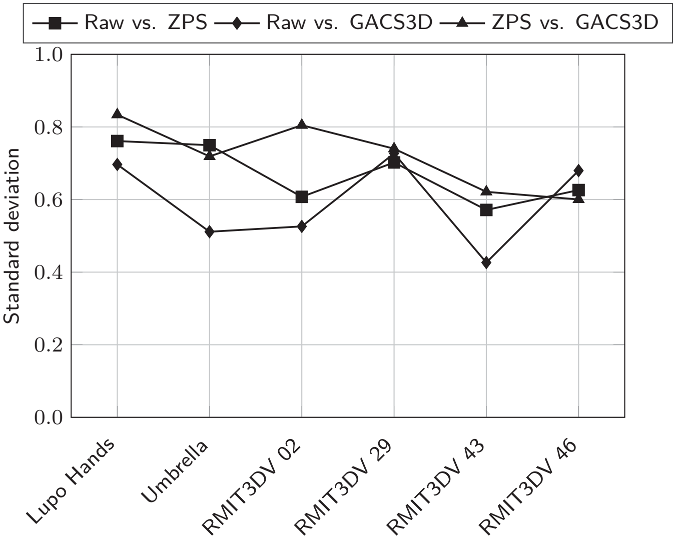
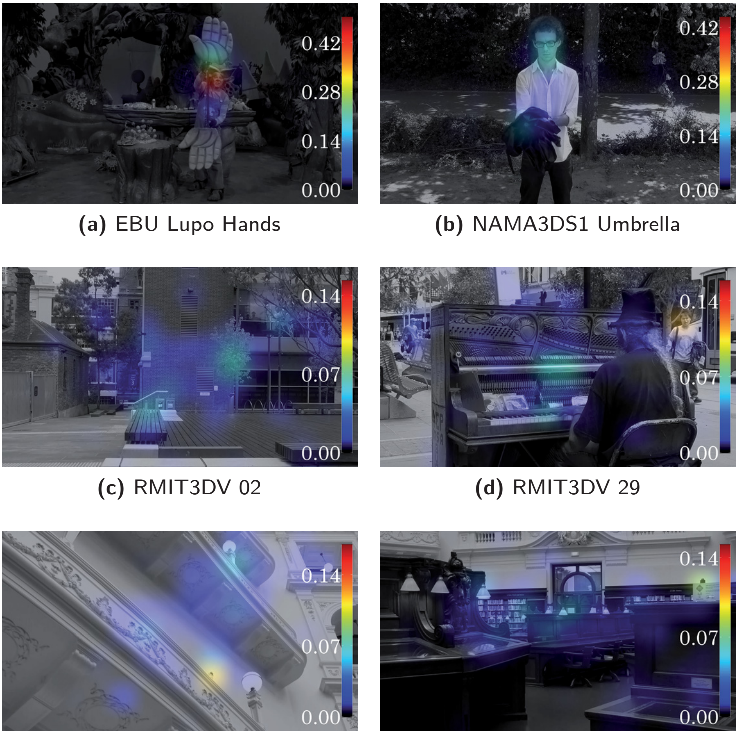


Figure 5.18: Standard deviation of comparison ratings in experiment (IV).

exhibits standard deviations just as high as the other conditions. Considering all the evidence in this section, it is concluded that the perceived visual discomfort was the same in all conditions. In other word, GACS3D does not affect visual discomfort, as long as the resulting disparities don't significantly protrude the zone of comfort, and the shift speed is kept below the annoyance threshold.

The results of this experiment are in contrast to the results by Hanhart and Ebrahimi [Han14], which suggest that gaze adaptive DHIT reduces visual discomfort, although the DHIT was designed in a much more critical way in their experiment. Their maximum shift speed was $v_{\mathrm{s,max}} = 0.5$ px/frame, which corresponds to 0.21 °/s at the frame rate of 25 frames/s. This is almost as high as the average annoyance threshold from the first experiment and a lot higher than some of the individual annoyance thresholds. Furthermore, as already pointed out in **section 5.2.2.6**, the group uses a nearest neighbor interpolation for the DHIT, which might lead to unknown side effects. A possible reason for the differing results could be the usage of more critical test material: Hanhart and Ebrahimi have used sequences with bigger disparity budgets, and some of which have actually protruded the ZOC.



(a) EBU Lupo Hands

(b) NAMA3DS1 Umbrella

(c) RMIT3DV 02

(d) RMIT3DV 29

(e) RMIT3DV 43

(f) RMIT3DV 46

Figure 5.19: Exemplary frames of the test sequences with heat map overlays that show the relative observation frequencies of certain image regions in in experiment (IV).

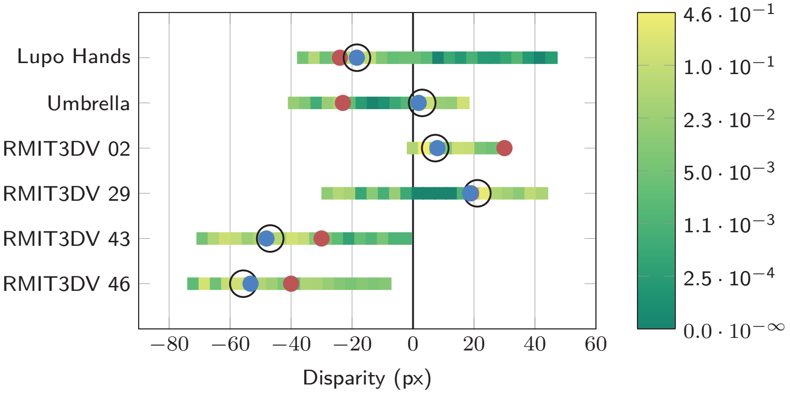


Figure 5.20: Relative frequencies of watched disparities (greenscale) in experiment (IV) for each sequence. The plot also includes the "Raw"-Setting (black line), the "ZPS" (red dot), the average watched disparity (blue dot) and the average convergence disparity (black circle). This plot is displayed in the unshifted disparity domain. In the "ZPS" condition, the whole disparity range is shifted so that the red dot is placed at zero disparity.

More conclusive data could possibly be gathered by adjusting the experimental design. On the one hand, the overall data quality could be improved by using an autostereoscopic display. This would eliminate the need for S3D-glasses, which would improve eye tracker performance and therefore also reduce rejections. On the other hand, more discriminating material could be used, as in the experiment of Hanhart and Ebrahimi [Han14]. The material should significantly protrude the zone of comfort so that actual discomfort is induced. Nevertheless, rating immediate visual discomfort is a relatively difficult task for a subject. Rating symptoms of visual fatigue is a lot less error-prone, but would involve an elaborate experiment with two or three prolonged viewing sessions of at least one hour to achieve a certain severity of the symptoms. Obviously, proper eye tracker performance would have to be ensured during the whole session, which poses another problem.

Since an eye tracker was used in the experiment, it was also possible to analyze visual attention. In **figure 5.19**, snapshots of the six video sequences are shown along with heat map overlays visualizing the actual areas of visual attention of all subjects. Scene elements like faces or occlusions were highly salient and reduced the exploration of the scene. Conversely, when these attractors were absent, as in RMIT3DV 02, 43 and 46, the scene was explored a lot more freely. This also shows in the normalized histograms of watched disparities in **figure 5.20**, where concentrations can be observed on the disparities of visual attractors. This strongly influenced the behavior of GACS3D because it almost degenerated to a static HIT in presence of visual attractors. These observations fortify the common procedure to place objects of interest near the display plane. However, this is not necessarily easy to do because, as can be seen in **figure 5.20**, the chosen ZPS and the average watched disparity may differ quite a bit due to wrongfully identified visual attractors. In absence of such visual attractors, the whole disparity budget should simply be fitted into the zone of comfort, ideally while considering the scene depth statistics.

A very small group of subjects also hinted that they disliked the DHIT, i.e., GACS3D, while the rest of the subjects did not detect it at all. This shows again that there are individual differences in the DHIT sensitivity, as already mentioned in **sections 5.3.2.5 and 5.3.3.3**.

5.4 Conclusion

There has been little research on the perception of the **dynamic horizontal image translation (DHIT)**. However, knowledge in this field is much-needed, especially for parameterizing automated DHIT approaches. The first experiment revealed that the shift speed is the main determinant of DHIT perception and that the sensitivity towards DHIT varies strongly between subjects. This latter finding is of great importance and led to a rather conservative recommendation to keep the shift speed in the range of 0.10 °/s to 0.12 °/s in order to ensure that nobody in the audience gets annoyed by the DHIT.

In this chapter, two new enhancements of the DHIT were proposed. The **distortion-free dynamic horizontal image translation (DHIT+)** mitigates the distortions of the depth budget due to horizontal image translation by adjusting the disparity budget through depth-image-based rendering techniques. This is perceptually similar to a camera movement along the depth axis. The DHIT+ proved to be significantly less perceivable and annoying compared to the DHIT. The convergence disparity could be altered about 50 % faster without any perceptual side effects.

The proposed approach **gaze adaptive convergence in stereo 3D applications (GACS3D)** utilizes an eye tracker in order to slowly establish the zero parallax setting at the visual focus using the DHIT. Since precise knowledge about the disparity of the visual focus is needed, the highly accurate PVFDE, proposed in **section 4.4.1**, is used to estimate it. In an experiment with emulated eye tracking, GACS3D significantly reduced annoyance compared to a regular DHIT. In the experimental evaluation of the complete prototype, however, no significant effect on visual discomfort was found compared to a static horizontal image translation. In contrast to these results, Hanhart and Ebrahimi were able to show an improvement with an approach similar to GACS3D [Han14]. The reason for this improvement could be the usage of more critical test material with bigger disparity budgets.

6 Conclusion

6.1 Summary

One factor in the lacking popularity of **stereo 3D (S3D)** nowadays is the visual fatigue many people still experience after prolonged exposure to respective content. There are numerous methods to reduce visual fatigue. The **dynamic horizontal image translation (DHIT)** is one of them, and its analysis is the main topic of this thesis.

In order to understand how visual fatigue is generated by S3D content, knowledge about the **human visual system (HVS)** is mandatory. So, in **chapter 2**, the HVS has been described with a special focus on the perception of depth. It has been pointed out that the HVS analyzes and combines multiple depth cues according to their estimated reliability, and that certain cue conflicts are problematic because they induce discomfort and may lead to an unstable perception.

In **chapter 3**, the fundamentals of S3D have been briefly described, in order to finally explain how visual fatigue is created, measured, and prevented. One of the most important sources of visual fatigue is the **accommodation vergence discrepancy (AVD)**: While the true imagery is located at the display distance to which the eyes theoretically have to adjust the focal length to (accommodation), the eyes may actually converge on a stimulus far behind or in front of the S3D display. This means that accommodation and vergence distance differ, which is unnatural. However, the AVD only becomes problematic once the **depth of field (DOF)** is protruded. Hence, the S3D content is placed in a **zone of comfort (ZOC)** inside that DOF. This is achieved, on the one hand, by setting the distance between the two S3D cameras such that the depth budget, i.e., the difference between maximum and minimum scene depth is small enough to fit inside the ZOC. On the other hand, the scene is placed in that ZOC by shifting it along the depth axis. This can be done by a **horizontal image translation (HIT)** of the S3D views in opposite directions in a static or temporally dynamic manner. The latter case is the aforementioned DHIT. Furthermore, it has been mathematically shown that there is a certain distortion of depth when any form of HIT is performed.

The DHIT has been analyzed and enhanced extensively in the later chapters of this work. One of the enhancements utilizes a remote eye tracker, which is why the principles of eye tracking have been described in **chapter 4**. Advanced filtering techniques had to be developed for the prototype, which have been described after a short review of the state of the art. For filtering the 2D gaze data, a Kalman filter has been derived that is tailored to gaze-directed human-machine-interaction. It has been extended with a new outlier and saccade[1] detection algorithm that is capable of handling missing samples, which is a common issue in eye tracking. Another significant contribution of this work is the highly accurate 3D visual focus estimation called **probabilistic visual focus disparity estimation (PVFDE)**. This newly proposed method yields an improvement of the **mean squared error (MSE)** by multiple orders of magnitude compared to the state of the art and even performs well with an inexpensive consumer-grade eye tracker. This improvement stems from the utilization of the estimated 3D structure of the presented scene, which was implemented in CUDA C++ in order to utilize the parallel processing power of a **graphics processing unit (GPU)**.

The DHIT has finally been analyzed in **chapter 5**, starting with a review of related publications. Research on the perception of DHIT is lacking, which was the motivation to conduct experiment (I). The scope of that experiment was to find out what shift speeds render the DHIT perceivable or even annoying. On average, a shift speed of 0.137 °/s per view was not perceivable, and 0.232 °/s was the threshold for annoyance. However, one of the main findings was that there are strong individual differences in the perception of DHIT. This is problematic because the usual heuristic DHIT design approach might yield undesirable results for some members of the audience. In order to ensure that no observer gets annoyed by the DHIT, it is therefore recommended to keep the shift speed in the range 0.10 °/s to 0.12 °/s. A small, but statistically

[1] Volatile eye movement from one visual focus to another.

significant effect of content on the results was also discovered. This effect might be due to the distortion of depth as induced by the HIT.

This problem can be solved by the newly proposed DHIT enhancement called **distortion-free dynamic horizontal image translation (DHIT+)**, one of the main contributions of this work. This approach mitigates the distortion of the depth budget by adjusting the distance between the S3D cameras. The results are similar to a camera movement along the depth axis. DHIT+ functions independently of viewing distance, but a slight dependency on eye baseline is given. The depth distortion due to eye baseline[2] mismatch is a lot smaller than that of the regular DHIT, however. The results of another conducted experiment have been presented, in which the DHIT+ was compared to the regular DHIT. The main finding was that DHIT+ significantly reduces perception and annoyance so that the convergence disparity can be set about 50 % faster. This makes this new technique a useful addition to the tool set of the modern stereographer. Furthermore, a significant effect of shift budget[3] and content on the results was detected for the DHIT+ but surprisingly not for the DHIT. The **computer generated imagery (CGI)** material used for the DHIT+ experiment exhibited some artifacts whose visibility varied between sequences, which might be the origin of those significant effects. Unfortunately, the existence of these artifacts prevented any conclusions regarding the depth distortion hypothesis from experiment (I), as mentioned above.

Finally, another major contribution of this work called **gaze adaptive convergence in stereo 3D applications (GACS3D)** has been described. This approach aims to lessen visual fatigue by reducing the AVD at the visual focus via slowly shifting it into the display plane using the DHIT. The visual focus is estimated using an eye tracker. In this approach, precise knowledge about the fixated depth is necessary, which was the motivation to develop the PVFDE described above. GACS3D is theoretically compatible to DHIT+, which requires a real-time adjustment of the distance between the S3D cameras. A respective depth-image-based rendering approach has been implemented in CUDA C++. However, these fully automated approaches generally exhibit some artifacts, which rendered this combination inappropriate for further testing. One problem

[2] The distance between the eyes.

[3] The total amount of shift applied over a couple of seconds.

with such an automated DHIT design as in GACS3D is that window violations[4] can be created. An easy to implement, yet powerful automated floating window[5] algorithm has been proposed to prevent that. While it has been described in the context of GACS3D, it is generally applicable to all automated DHIT design approaches.

The effect of a gaze adaptive DHIT design was investigated in two experiments, i.e., experiments (III) and (IV). In experiment (III), gaze adaptivity was only emulated using a pointer that subjects were asked to track. This was done so that all subjects were exposed to the exact same stimuli. The results showed a significant decrease of annoyance in the case of gaze adaptivity compared to the regular DHIT. In experiment (IV), the complete prototype was used and compared with stationary HIT test sequences. However, the results were rather inconclusive because no clear tendency could be observed. It was concluded that the perceived visual discomfort was the same in all conditions. A similar approach by Hanhart and Ebrahimi has yielded an improvement of visual discomfort in the experiments of that group [Han14], despite having used a more aggressive implementation and parametrization of the DHIT. The reason for this difference could be their use of more critical test material with bigger disparity budgets, partially protruding the ZOC. A later analysis of visual attention showed that GACS3D almost degenerates to a static HIT in presence of strong visual attractors. Still, GACS3D is a useful automated tool for real-time visualization of unprocessed S3D content.

6.2 Outlook

The conducted experiments raised new interesting questions for future research. The results of experiment (I) exhibited a very small, but statistically significant content dependency, which could possibly be attributed to the depth budget distortion of HIT. Experiment (II) showed how the DHIT+ significantly reduces annoyance compared to DHIT, even though the DHIT+ test stimuli contained some artifacts. Furthermore, a content dependency was found for DHIT+, but

[4] A visually unpleasant conflict between occlusion and disparity, that occurs when an object in front of the screen is cut off by the left or right screen border.

[5] The 3D content is seen through a stereoscopic window, which can actually be moved in 3D space, e.g., in front of a window violation in order to eliminate it.

not for DHIT, which is in contrast to the first experiment. While experiment (II) was a success and the test stimuli served their purpose, the presence of these artifacts prevented any additional conclusions about the content dependencies of either approach. The artifacts were only present in the DHIT+ stimuli, resulting in a comparatively strong content dependency that would dominate the results over the content dependency of the regular DHIT, which would in turn simply disappear. Hence, a redesign of the test stimuli using professional studio software for CGI and repeating the exact same procedure might yield an answer.

Furthermore, the DHIT yielded slightly better results in experiment (II) than in experiment (I). Considering the discovered strong individual differences, the group of subjects in experiment (II) might just have been less sensitive towards the DHIT. Another explanation could be the different nature of the test material, i.e., CGI vs. natural content. These hypotheses could be tested by repeating the first experiment with both sets of stimuli. If the stimuli still yield different results, the reason for that would definitely be some content characteristic because the same group of subjects evaluated the stimuli.

Finally, GACS3D is supposed to reduce visual fatigue, whereas only immediate visual discomfort was rated in experiment (IV), and no significant effect was found for any tested approach. Since almost all test stimuli were contained in the ZOC, no strong discomfort was induced. Hence, using more critical material might be an option to identify significant effects. However, the theoretically most reliable solution would be a direct evaluation of visual fatigue, which is a very elaborative undertaking exceeding the scope of this work. Visual fatigue is accumulated over prolonged viewing sessions and some of its symptoms take some time to recede. So, in order to evaluate visual fatigue, an experiment would involve exposing each subject to GACS3D-processed content and regular content (in random order) for at least one hour each, separated by a break of about 24 hours. Respective evaluation questionnaires would have to be answered before and after each exposure so that different daily conditions of each subject can be accounted for. Proper eye tracker performance is mandatory during the whole experiment. Hence, data quality needs to be improved. This could be achieved by using an autostereoscopic display to eliminate the polarization glasses.

List of Abbreviations

AVD	Accommodation vergence discrepancy.
CGI	Computer generated imagery.
DHIT+	Distortion-free dynamic horizontal image translation.
DHIT	Dynamic horizontal image translation.
DOF	Depth of field.
GACS3D	Gaze adaptive convergence in stereo 3D applications.
GPU	Graphics processing unit.
HIT	Horizontal image translation.
HVS	Human visual system.
LMEM	Linear mixed-effect model.
MAE	Mean absolute error.
MOS	Mean opinion score.
MSE	Mean squared error.
POI	Point of interest.
PVFDE	Probabilistic visual focus disparity estimation.
ROI	Region of interest.
S3D	Stereo 3D.
SAD	Sum of absolute differences.
SFR	Stereoscopic fusion range.
ZOC	Zone of comfort.
ZPS	Zero parallax setting.

List of Symbols

$\mathcal{A}$	State transition matrix of a Kalman filter.
a_k	One-dimensional acceleration of a particle in a Kalman filter at time instance k.
α	In the DHIT+, the disparity budget is altered by this factor to maintain a certain depth budget.
b_{c}^{*}	Adjusted camera baseline in the DHIT+.
b_{c}	Camera baseline, i.e., the distance the stereo cameras.
b_{e}	Eye baseline, i.e., the distance between the eyes.
β	Confidence level for the model violation identification of a Kalman filter.
$B\left(\boldsymbol{p}_i\right)$	A block of pixels in 2D image space centered on $\boldsymbol{p}_i$.
c_j	Polynomial 3D eye tracker calibration coefficients.
d	Distance from the observer's eyes to the display.
D_{B}	Disparity budget of a given scene, i.e., the difference between $D_{\max}$ and $D_{\min}$.
D_{conv}	Convergence disparity.
D_{conv}^{*}	Scaled convergence disparity in the DHIT+.
$D_{\mathrm{conv,alt}}$	Alternative cyclic convergence disparity in the DHIT+ experiment.
$D_{\mathrm{conv,corr}}^{*}$	Corrected scaled convergence disparity in the DHIT+ experiment.
$D_{\mathrm{conv,cycl}}$	Cyclic convergence disparity in the DHIT and DHIT+ experiment.
$D_{\mathrm{conv,cycl}}^{*}$	Scaled cyclic convergence disparity in the DHIT+ experiment.

$D_{\mathrm{conv},k}$	Convergence disparity at time instance k in GACS3D.
$D_{\mathrm{conv},\max}$	Maximum convergence disparity in a DHIT sequence.
$D_{\mathrm{conv},\min}$	Minimum convergence disparity in a DHIT sequence.
$D_{\mathrm{conv,pre}}$	Pre-convergence disparity in the DHIT and DHIT$+$ experiment.
$D^{*}_{\mathrm{conv,pre}}$	Scaled pre-convergence disparity in the DHIT$+$ experiment.
$\Delta D_{\mathrm{C,L\|R}}$	Difference between disparity and mapped disparity.
$\Delta D_{\max}$	Maximum allowed disparity difference.
ΔZ	The depth distortion due to any form of horizontal image translation is defined as the ratio of maximum to minimum depth budget.
D°_{i}	Retinal disparity of point $\boldsymbol{P}_i$.
$\tilde{D}^{*}_{i}$	Shifted and scaled disparity after applying the DHIT$+$.
$\tilde{D}_{i}$	Shifted disparity after applying the HIT.
D_{i}	Planar disparity that is generated by a 3D point $\boldsymbol{P}_i$.
D_{LB}	Lower bound of the valid disparity range in the PVFDE.
$D_{\mathrm{L\|R}}\left(.\right)$	Left or right view true disparity map.
$\hat{D}_{\mathrm{L\|R}}\left(.\right)$	Left or right view estimated disparity map.
$\tilde{D}^{*}_{\max}$	Maximum shifted and scaled disparity after applying the DHIT$+$.
$D_{\max}$	Maximum disparity in a given recorded scene.
$\tilde{D}_{\max}$	Maximum shifted disparity after applying the HIT.
$\tilde{D}^{*}_{\min}$	Minimum shifted and scaled disparity after applying the DHIT$+$.
$D_{\min}$	Minimum disparity in a given recorded scene.
$\tilde{D}_{\min}$	Minimum shifted disparity after applying the HIT.
D_{POI}	Disparity of a point of interest.
$\hat{D}_{\mathrm{T}}$	Estimate of the true watched disparity.
D_{T}	True watched disparity.

$\overline{D}_{\mathrm{targ},k}$	Target disparity at time instance k in GACS3D.
$D_{\mathrm{targ},k}$	Target disparity at time instance k in GACS3D.
D_{UB}	Upper bound of the valid disparity range in the PVFDE.
$\mathrm{E}\left\{.\right\}$	Expectancy value operator.
e_k^2	Normalized error of a Kalman filter in iteration k.
$e_{k,\beta}^2$	Normalized error threshold for a confidence level β of a Kalman filter in iteration k.
$\boldsymbol{e}_x$	The unity vector in x-direction.
$\boldsymbol{F}$	Fixation point in 3D space.
g	Help variable in the solution for α.
γ	Disparity smoothing coefficient in GACS3D.
H	Height of a display.
h	Help variable in the solution for α.
$\mathcal{H}$	Observation model of a Kalman filter.
$\mathcal{I}$	Unity matrix.
$\boldsymbol{i}_k$	Innovation of a Kalman filter in iteration k.
$I_{\mathrm{L}\vert\mathrm{R}}\left(.\right)$	Left or right view image.
k	Time instance, e.g., iteration number.
$\mathcal{K}_k$	Kalman gain minimizing a posteriori MSE in iteration k.
L	Luminance.
$\mathcal{M}$	Data matrix for 3D eye tracker calibration.
$\boldsymbol{m}_k$	Measurement vector in the Kalman filter model.
μ_k	Shift direction at time instance k in GACS3D.
N	Number of data points during 3D eye tracker calibration.
n_{inv}	Number of invalid samples in a connected series of model violations in the saccade detection process.
$\boldsymbol{n}_k$	Measurement noise of a Kalman filter in iteration k.
n_{th}	Connected violation threshold in the saccade detection process.

$\hat{n}_{\mathrm{V}}$	Estimated number of connected violations in the saccade detection process.
n_{V}	Number of connected violations in the saccade detection process.
$n_{\mathrm{V,LB}}$	Lower bound of the number of connected violations in the saccade detection process.
$n_{\mathrm{V,UB}}$	Upper bound of the number of connected violations in the saccade detection process.
ω_{C}	Total weight of a candidate.
$\omega_{\Delta D,\mathrm{C}}$	Weight of disparity difference in the disparity estimation process.
$\omega_{r^{\dagger},\mathrm{C}}$	Weight of mapped distance in the disparity estimation process.
$\omega_{r,\mathrm{C}}$	Weight of distance in the disparity estimation process.
$\boldsymbol{P}_i$	A point, number i, in 3D space.
p	Probability of a certain result based on the assumption that the null-hypothesis is true.
$\boldsymbol{p}_{\mathrm{C,L\|R}}$	Candidates in the left or right region of interest in the disparity estimation process.
$\boldsymbol{p}^{\dagger}_{\mathrm{C,L\|R}}$	Candidates in the left or right region of interest mapped to the right or left view in the disparity estimation process.
$\tilde{\boldsymbol{p}}_{\mathrm{G,L\|R}}$	Shifted domain version of gaze vector $\boldsymbol{p}_{\mathrm{G,L\|R}}$.
$\boldsymbol{p}_{\mathrm{G,L\|R}}$	Two dimensional gaze data in real screen-space pixel coordinates for the left or right eye, as returned from the eye tracker. $\boldsymbol{p}_{\mathrm{G,L\|R}} = \left(x_{\mathrm{G,L\|R}}, y_{\mathrm{G,L\|R}}\right)^{\top}$.
$\phi_{\mathrm{L\|R},i}$	Angular distance between an image of point $\boldsymbol{P}_i$ and the fovea on the left or right retina.
$\phi_{P,i}$	Angle between the left and right visual lines through a non-fixated point $\boldsymbol{P}_i$.
ϕ_{V}	Angle between the left and right visual axes, i.e., the vergence angle.

$\boldsymbol{p}_i$ A 2D point, number i, in 2D image space.

p_k One-dimensional position of a particle in a Kalman filter at time instance k.

$\hat{\mathcal{P}}_k$ Estimated a posteriori error covariance matrix of a Kalman filter in iteration k.

$\mathcal{P}_k$ A posteriori error covariance matrix of a Kalman filter in iteration k.

$\hat{\mathcal{P}}_k^-$ Estimated a priori error covariance matrix of a Kalman filter in iteration k.

$\mathcal{P}_k^-$ A priori error covariance matrix of a Kalman filter in iteration k.

$\boldsymbol{p}_{\mathrm{N,L|R}}$ A 2D additive noise term for gaze data of the left or right eye in real screen-space pixel coordinates.

$\boldsymbol{p}_{\mathrm{T,L|R}}$ The true 2D gaze position of the left or right eye in real screen-space pixel coordinates.

$\mathcal{Q}$ Process noise covariance matrix in the Kalman filter model.

r Maximum spread radius of gaze samples around the true gaze position $\boldsymbol{p}_{\mathrm{T,L}}$.

$\mathcal{R}$ Measurement noise covariance matrix in the Kalman filter model.

R^2 Correlation coefficient for multiple linear regression.

r^2 Pearson correlation coefficient.

$r_{\mathrm{C,L|R}}$ Distance between candidate $\boldsymbol{p}_{\mathrm{C,L|R}}$ and gaze data $\boldsymbol{p}_{\mathrm{G,L|R}}$.

$r^{\dagger}_{\mathrm{C,L|R}}$ Distance between candidate $\boldsymbol{p}^{\dagger}_{\mathrm{C,L|R}}$ and gaze data $\boldsymbol{p}_{\mathrm{G,L|R}}$ of the opposite eye.

ρ Pixel pitch, i.e., the distance between two pixels on the display screen in meters.

$\mathrm{SAD}\,(.)$ The sum of absolute differences of a Block.

SB Shift budget in the DHIT.

$\mathrm{SB}_{\mathrm{front}}$ Portion of the shift budget in the DHIT that places the S3D content closer to the observer than the raw source material.

σ_a^2 Process noise variance of acceleration in the constant-velocity Kalman filter model.

$\sigma_{\Delta D}^2$ Weighting parameter for $\Delta D_{\mathrm{C,L|R}}$.

σ_n^2 Measurement noise variance of position in the constant-velocity Kalman filter model.

σ_r^2 Variance of the normally distributed distance between true gaze and noisy gaze data.

$\mathrm{sign}\,(.)$ The sign function returns $+1$ for positive arguments and zero, and -1 for negative arguments.

$\mathcal{S}_k$ Innovation covariance matrix of a Kalman filter in iteration k.

$\hat{\boldsymbol{s}}_k$ A posteriori process state estimate of a Kalman filter in iteration k.

$\hat{\boldsymbol{s}}_k^-$ A priori process state estimate of a Kalman filter in iteration k.

$\boldsymbol{s}_k$ Process state of a Kalman filter in iteration k.

T The time between two measurement instances in a Kalman filter context.

t Time in seconds.

v_k One-dimensional speed of a particle in a Kalman filter at time instance k.

v_{s} Shift speed as applied to each S3D view separately during the DHIT or DHIT$+$.

$v_{\mathrm{s,max}}$ Maximum shift speed as applied to each S3D view separately during the DHIT or DHIT$+$.

W Width of a display.

$\boldsymbol{w}_k$ Process noise of a Kalman filter in iteration k.

$x_{\mathrm{G,L|R}}$ Horizontal, real screen-space pixel coordinate for the gaze of the left or right eye as delivered from the eye tracker, i.e., the horizontal component of $\boldsymbol{p}_{\mathrm{G,L|R}}$.

$x_{\mathrm{L},i}$ Real, horizontal screen-space coordinates of a 3D point $\boldsymbol{P}_i$ in the left view of a stereoscopic image pair.

$x_{\mathrm{R},i}$ Real, horizontal screen-space coordinates of a 3D point $\boldsymbol{P}_i$ in the right view of a stereoscopic image pair.

$y_{\mathrm{C,L|R}}$ Vertical, real screen-space pixel coordinate of a candidate for the left or right eye in the disparity estimation process, i.e., the vertical component of $\boldsymbol{p}_{\mathrm{C,L|R}}$.

$y_{\mathrm{G,L|R}}$ Vertical, real screen-space pixel coordinate for the gaze of the left or right eye as delivered from the eye tracker, i.e., the vertical component of $\boldsymbol{p}_{\mathrm{G,L|R}}$.

$y_{\mathrm{T,L|R}}$ Vertical, real screen-space pixel coordinate for the *true* gaze of the left or right eye, i.e., the vertical component of $\boldsymbol{p}_{\mathrm{T,L|R}}$.

Z Depth budget of a given scene, i.e., the difference between the maximum and minimum depth.

z Depth coordinate.

$\boldsymbol{z}$ Vector of reference depths for each recorded data point during 3D eye tracker calibration.

z_{F} Fixation or vergence distance.

z_{far} Maximum depth of a given scene.

$\hat{z}_i$ Visual focus depth estimated from 2D gaze data point number i.

z_i Depth of a point $\boldsymbol{P}_i$.

Z_{min} Minimum depth budget of a given scene when applying any form of HIT.

z_{near} Minimum depth of a given scene.

Z_{targ} The target depth budget that is supposed to be maintained in the DHIT+.

List of Figures

List of Tables

Glossary

Accommodation	Adjustment of the focal length of the eyes.
Accommodation vergence conflict	See accommodation vergence discrepancy.
Accommodation vergence discrepancy	Accommodation and vergence distance unnaturally differ when watching a stimulus outside the stereoscopic display plane.
Accuracy	The mean absolute error between the true and the measured gaze coordinates.
Active depth cut	A dynamic horizontal image translation is used in order to mitigate depth discontinuities at scene cuts.
Adaption	The human visual system adapts to varying lighting conditions by different receptors and iris diameter.
Angular disparity	See retinal disparity.
Binocular disparity	See retinal disparity.
Binocular rivalry	See retinal rivalry.
Blind spot	The spot on the retina where the optic nerve connects to the eye does not inhabit any light receptors.
Cone	A light sensitive cell type on the retina enabling color vision in bright environments.

Convergence disparity	The disparity in the unshifted domain that is zeroed by the horizontal image translation.
Convergence plane	The sensor-parallel plane of points in the 3D scene space yielding zero disparity. In 3D observer space, the convergence plane is the display plane.
Crosstalk	The left view is partially visible in the right view or vice versa.
CUDA C++	A programming language for parallel computing on graphics cards.
Cyclopean perception	The single 3D view perceived by an observer, which is located right in the middle between the eyes.
Depth budget	The difference between maximum and minimum scene depth.
Depth of field	The range of depths that can be viewed sharply.
Diplopia	The double vision when fusion fails.
Disparity	The onscreen parallax in pixels in a stereo 3D context.
Disparity budget	The difference between maximum and minimum disparity.
Drift	The eye drifts away from a fixation in this unconscious eye movement.
Dynamic horizontal image translation	A temporally dynamic form of the horizontal image translation
Eye tracker	A device that tracks the position and the gaze direction of the eyes.
Fixation	The eyes foveate an object of interest.

Floating window Moving the sterescopic window in front of the display plane, in order to avoid window violations.

Fovea The central part of the retina, which exhibits the highest density of light sensitive cells.

Fusion The process of the human visual system that fuses two views to a single 3D view.

Glissade Small correction saccades during or after a saccade.

Horizontal image translation The process of translating stereoscopic views horizontally in opposite directions such that the content is shifted further behind or in front of the display.

Horopter The locus of points exhibiting zero retinal disparity.

Innovation The difference between a priori prediction and measurement of a Kalman filter.

Iris The iris functions as an aperture, that controls how much light enters the eye.

Kinetic depth effect The analogon to motion parallax, but related to object motion.

Micro saccade After a drift occured, the position of the eye is corrected again to the center of the fixation.

Motion parallax The images of static objects at different depths move at different speeds across the retina when the observer moves.

Motoric fusion The fusion process supported by vergence movements.

Occlusion Occlusion is a strong depth cue, that enables the HVS to perform depth ordering.

Oculomotor An adjective relating objects or functions to the system of muscles of the eyeball.

Oculomotor noise The unconscious eye movements are deemed to be noise in the context of eye tracking in many applications.

Panum's fusional area The small area around the horopter that can be fused so that no diplopia occurs.

Parallax See disparity.

Precision The spread of gaze samples, mostly given as standard deviation.

Reconvergence See horizontal image translation.

Retina The light sensitive area of the eye.

Retinal disparity The horizontal angular parallax on the retinas of an observer's eyes induced by a 3D stimulus. There is also vertical retinal disparity.

Retinal rivalry Differences in the images on the retinas, that may be unnatural and cause visual discomfort.

Rod	Light sensitive cell type on the retina enabling gray scale vision in dark environments.
Saccade	An eye movement that switches rapidloy from one stimulus to another.
Shift budget	The absolute difference between maximum and minimum convergence disparity in the DHIT.
Smooth pursuit	A moving target stimulus is instinctively tracked by the eyes.
Stereopsis	See fusion.
Stereoscopic fusion range	The range of disparities that can be fused on a stereoscopic display with appropriate vergence movement.
Stereoscopic window	The stereoscopic display and its borders can be seen as a window through which a 3D world can be observed.
Stereoscopy	Presentation of horizontally offset views of a scene to the eyes of an observer so that a 3D scene is perceived.
System noise	The eye tracker inherent noise.
Texture gradient	A monocular pictorial depth cue, which especially enables estimation of the slant of a surface.
Tremor	An unconscious eye movement in the form of small amplitude jitter.
Vergence	See vergence movement.
Vergence angle	The angle between the visual axes of the eyes.

Vergence movement The adjustment of the vergence angle of the eyes.

Vertical retinal disparity The vertical angular parallax of images on the retina due to the vergence angle of the eyes.

Vieth-Müller circle See horopter.

Visual axis The Line connecting fovea, pupil and target stimulus.

Visual discomfort The feeling of discomfort induced by stereo 3D content.

Visual fatigue The objectively measurable exhaustion of the human visual system induced by stereo 3D content.

Visual focus The point gazed upon.

Visual line The line connecting the nodal point of an eye and a non-fixated point.

Window violation The conflict between occlusion and disparity induced by an object in front of the display plane that is occluded by the display border.

Zero parallax setting See convergence plane.

Zone of clear single binocular vision See stereoscopic fusion range.

Zone of comfort The depth or disparity range that can be viewed comfortably on a stereoscopic display.

Bibliography

[AA02] Abd-Almageed, W.; Fadali, M. & Bebis, G., *A Non-intrusive Kalman Filter-Based Tracker for Pursuit Eye Movement*. In: *Proceedings of the 2002 American Control Conference*, vol. 2, pp. 1443–1447, IEEE (2002).

[All04] Allison, R. S., *The Camera Convergence Problem Revisited*. In: Woods, A. J.; Merritt, J. O.; Benton, S. A. & Bolas, M. T. (editors), *Proceedings of SPIE*, vol. 5291, pp. 167–178, SPIE (2004).

[Alt14] Alt, F.; Schneegass, S.; Auda, J.; Rzayev, R. & Broy, N., *Using Eye-Tracking to Support Interaction with Layered 3D Interfaces on Stereoscopic Displays*. In: *Proceedings of the 19th International Conference on Intelligent User Interfaces*, pp. 267–272, ACM Press, New York, USA (2014).

[AN10] Al-Nahlaoui, M. Y., *Methoden und Konzepte der Blickrichtungserkennung*. Ph.D. thesis, TU Dortmund University (2010).

[Ban12] Banks, M. S.; Read, J. C. A.; Allison, R. S. & Watt, S. J., *Stereoscopy and the Human Visual System*. In: *SMPTE Motion Imaging Journal*, vol. 121(4), pp. 24–43 (2012).

[Bar11] Barkowsky, M.; Tourancheau, S.; Brunnström, K.; Wang, K. & Andrén, B., *55.3: Crosstalk Measurements of Shutter Glasses 3D Displays*. In: *SID Symposium Digest of Technical Papers*, vol. 42(1), pp. 812–815 (2011).

[Ber10] Berezin, O., *Digital cinema in Russia: Status of 2D/3D DC rollout*. Tech. rep., 3D Stereo Media 2010 Summit, Liège, Belgium (2010).

[Ber14] Bernhard, M.; Dell'mour, C.; Hecher, M.; Stavrakis, E. & Wimmer, M., *The Effects of Fast Disparity Adjustment in Gaze-Controlled*

Stereoscopic Applications. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 111–118, ACM Press, New York, New York, USA (2014).

[Bro11] Broberg, D. K., *Guidance for horizontal image translation (HIT) of high definition stereoscopic video production*. In: Woods, A. J.; Holliman, N. S. & Dodgson, N. A. (editors), *Proc. SPIE 7863, Stereoscopic Displays and Applications XXII*, pp. 78632F/1–78632F/11, SPIE, San Francisco (2011).

[BS04] Bar-Shalom, Y.; Li, X. R. & Kirubarajan, T., *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. John Wiley & Sons, 1 ed. (2004).

[Cam57] Campbell, F., *The Depth of Field of the Human Eye*. In: *Optica Acta: International Journal of Optics*, vol. 4(4), pp. 157–164 (1957).

[Cha77] Charman, W. & Whitefoot, H., *Pupil Diameter and the Depth-of-field of the Human Eye as Measured by Laser Speckle*. In: *Optica Acta: International Journal of Optics*, vol. 24(12), pp. 1211–1216 (1977).

[Cha10] Chamaret, C.; Godeffroy, S.; Lopez, P. & Le Meur, O., *Adaptive 3D Rendering based on Region-of-Interest*. In: Woods, A. J.; Holliman, N. S. & Dodgson, N. A. (editors), *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics (2010).

[Che12] Cheng, E.; Burton, P.; Burton, J.; Joseski, A. & Burnett, I., *RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database*. In: *2012 Fourth International Workshop on Quality of Multimedia Experience*, pp. 212–217, IEEE (2012).

[Cro73] Crone, R. A. & Leuridan, O. M. A., *Tolerance for Aniseikonia: I. Diplopia Thresholds in the Vertical and Horizontal Meridians of the Visual Field*. In: *Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*, vol. 188(1), pp. 1–16 (1973).

[Dod04] Dodgson, N. A., *Variation and extrema of human interpupillary distance*. In: *Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems XI*, pp. 36–46, San Jose, CA (2004).

[Dol] Dolby, *Dolby 3D*.
Online: `http://www.dolby.com/us/en/technologies/dolby-3d.html#4` (Date accessed: 19.09.2016).

[Duc11] Duchowski, A. T.; Pelfrey, B.; House, D. H. & Wang, R., *Measuring Gaze Depth with an Eye Tracker During Stereoscopic Display*. In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pp. 15–22, ACM Press, New York, USA (2011).

[Duc14] Duchowski, A. T.; House, D. H.; Gestring, J.; Congdon, R.; Świrski, L.; Dodgson, N. A.; Krejtz, K. & Krejtz, I., *Comparing Estimated Gaze Depth in Virtual and Physical Environments*. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 103–110, ACM Press, New York, USA (2014).

[Dve10] Dvernay, F. & Beardsley, P., *Stereoscopic Cinema*. In: Ronfard, R. & Taubin, G. (editors), *Image and Geometry Processing for 3-D Cinematography*, pp. 11–52, Springer Verlag, Heidelberg, 5 ed. (2010).

[EBU] EBU, *3DTV Test Sequences*.
Online: `https://tech.ebu.ch/testsequences/3dtv_test` (Date accessed: 29.08.2016).

[Ee03] van Ee, R.; Adams, W. & Mamassian, P., *Bayesian modeling of cue interaction: bistability in stereoscopic slant perception*. In: *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 20(7), pp. 1398–1406 (2003).

[Emo05] Emoto, M.; Niida, T. & Okano, F., *Repeated Vergence Adaptation Causes the Decline of Visual Functions in Watching Stereoscopic Television*. In: *Journal of Display Technology*, vol. 1(2), pp. 328–340 (2005).

[Ess06] Essig, K.; Pomplun, M. & Ritter, H., *A neural network for 3D gaze recording with binocular eye trackers*. In: *International Journal of Parallel, Emergent and Distributed Systems*, vol. 21(2), pp. 79–95 (2006).

[Gał13] Gałecki, A. & Burzykowski, T., *Linear Mixed-Effects Models Using*

R. Springer Texts in Statistics, Springer New York, New York, USA (2013).

[Gar11] Gardner, B. R., *The Dynamic Floating Window - a new creative tool for 3D movies*. In: Woods, A. J.; Holliman, N. S. & Dodgson, N. A. (editors), *Stereoscopic Displays and Applications XXII*, vol. 7863, pp. 78631A/1–78631A/12 (2011).

[Gel05] Gelman, A., *Analysis of variance - why it is more important than ever*. In: *The Annals of Statistics*, vol. 33(1), pp. 1–53 (2005).

[Gez97] Gezeck, S.; Fischer, B. & Timmer, J., *Saccadic Reaction Times: a Statistical Analysis of Multimodal Distributions*. In: *Vision Research*, vol. 37(15), pp. 2119–2131 (1997).

[Gol02] Goldstein, E. B., *Sensation and Perception*. Wadsworth, Belmont, 6 ed. (2002).

[Gor89] Gordon, C. C.; Walker, R. A.; Tebbetts, I.; McConville, J. T.; Bradtmiller, B.; Clauser, C. E. & Churchill, T., *1988 Anthropometric Survey of US Army Personnel-Methods and Summary Statistics. Final Report*. Tech. rep., US Army Natick Research (1989).

[Hak11] Hakala, J.; Nuutinen, M. & Oittinen, P., *Interestingness of stereoscopic images*. In: *Stereoscopic Displays and Applications XXII*, San Francisco, USA (2011).

[Ham83] Hampton, D. R. & Kertesz, A. E., *The Extent of Panum's Area and the Human Cortical Magnification Factor*. In: *Perception*, vol. 12(2), pp. 161–165 (1983).

[Ham08] Hammoud, R. I., *Passive Eye Monitoring*. Signals and Commmunication Technologies, Springer Berlin Heidelberg, Berlin, Heidelberg (2008).

[Han14] Hanhart, P. & Ebrahimi, T., *Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience*. In: Woods, A. J.; Holliman, N. S. & Favalora, G. E. (editors), *IS&T/SPIE Electronic Imaging*, pp. 90110D/1–90110D/11, International Society for Optics and Photonics (2014).

[Hof08] Hoffman, D. M.; Girshick, A. R.; Akeley, K. & Banks, M. S., *Vergence-accommodation conflicts hinder visual performance and cause visual fatigue*. In: *Journal of Vision*, vol. 8(3), pp. 33/1–33/30 (2008).

[Hof11] Hoffman, D. M.; Karasev, V. I. & Banks, M. S., *Temporal presentation protocols in stereoscopic displays: Flicker visibility, perceived motion, and perceived depth.* In: *Journal of the Society for Information Display*, vol. 19(3), pp. 271–297 (2011).

[Hol11] Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H. & de Weijer, J., *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, New York, 1st ed. (2011).

[How95] Howard, I. P. & Rogers, B. J., *Binocular Vision and Stereopsis*. Oxford University Press, Oxford (1995).

[HT11a] Huynh-Thu, Q.; Barkowsky, M. & Le Callet, P., *The Importance of Visual Attention in Improving the 3D-TV Viewing Experience: Overview and New Perspectives*. In: *IEEE Transactions on Broadcasting*, vol. 57(2), pp. 421–431 (2011).

[HT11b] Huynh-Thu, Q. & Schiatti, L., *Examination of 3D visual attention in stereoscopic video content*. In: *Proceedings of SPIE Electronic Imaging*, vol. 7865, pp. 78650J/1–78650J/15 (2011).

[Iat14] Iatsun, I.; Larabi, M.-C. & Fernandez-Maloigne, C., *On the comparison of visual discomfort generated by S3D and 2D content based on eye-tracking features*. In: Woods, A. J.; Holliman, N. S. & Favalora, G. E. (editors), *IS&T/SPIE Electronic Imaging*, pp. 901124/1–901124/14, International Society for Optics and Photonics (2014).

[Isr11] Israel, S. G., *3D-Verfahren im Verglich: "Shuttered Glasses" (Aktiv) versus "Film Patterned Retarder" (Passiv)* (2011).
Online: `http://www.lgblog.de/2011/03/30/3d-verfahren-im-vergleich-shuttered-glasses-aktiv-versus-film-patterned-retarder-passiv/` (Date accessed: 2016-09-19).

[ITU12a] ITU, *ITU-R BT.2021: Subjective methods for the assessment of stereoscopic 3DTV systems.* Tech. rep., ITU (2012).

[ITU12b] ITU, *ITU-R BT.500: Methodology for the subjective assessment of the quality of television pictures*. Tech. rep., ITU (2012).

[Joo03] Joos, M.; Rötting, M. & Velichkowsky, B. M., *Spezielle Verfahren I: Bewegungen des menschlichen Auges : Fakten, Methoden und innovative Anwendungen*. In: *Psycholinguistik. Psycholinguistics: Ein internationales Handbuch. An International Handbook*, 1, chap. 2-10, pp. 142–167, Walter de Gruyter GmbH & Co. KG, Berlin (2003).

[Jul60] Julesz, B., *Binocular Depth Perception of Computer-Generated Patterns*. In: *Bell System Technical Journal*, vol. 39(5), pp. 1125–1162 (1960).

[Kal60] Kalman, R. E., *A New Approach to Linear Filtering and Prediction Problems*. In: *Transactions of the ASME-Journal of Basic Engineering*, vol. 82(Series D), pp. 35–45 (1960).

[Kap07] Kaptein, R. & Heynderickx, I., *Effect of Crosstalk in Multi-View Autostereoscopic 3D Displays on Perceived Image Quality*. In: *SID Symposium Digest of Technical Papers*, vol. 38(1), pp. 1220–1223 (2007).

[Ki07] Ki, J.; Kwon, Y.-M. & Sohn, K., *3D Gaze Tracking and Analysis for Attentive Human Computer Interaction*. In: *2007 Frontiers in the Convergence of Bioscience and Information Technologies*, pp. 617–621, IEEE (2007).

[Kim11] Kim, D.; Choi, S.; Park, S. & Sohn, K., *Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks*. In: *2011 17th International Conference on Digital Signal Processing (DSP)*, pp. 1–6, IEEE (2011).

[Kim13] Kim, D.; Choi, S. & Sohn, K., *Visual Comfort Enhancement for Stereoscopic Video Based on Binocular Fusion Characteristics*. In: *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23(3), pp. 482–487 (2013).

[Kim14] Kim, J.; Kane, D. & Banks, M. S., *The rate of change of vergence-accommodation conflict affects visual discomfort*. In: *Vision Research*, vol. 105, pp. 159–165 (2014).

[Koh09a] Koh, D. H.; Gowda, S. M.; Komogortsev, O. V.; Munikrishne Gowda, S. A. & Komogortsev, O. V., *Input Evaluation of an Eye-Gaze-Guided Interface: Kalman Filter vs. Velocity Threshold Eye Movement Identification*. In: *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pp. 197–202, ACM, New York, NY, USA (2009).

[Koh09b] Kohlbecher, S. & Schneider, E., *On-Line Classification and Prediction of Eye Movements by Multiple-Model Kalman Filtering*. In: *Annals of the New York Academy of Sciences*, vol. 1164(1), pp. 400–402 (2009).

[Kom07] Komogortsev, O. V. & Khan, J. I., *Kalman Filtering in the Design of Eye-Gaze-Guided Computer Interfaces*. In: Jacko, J. (editor), *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, vol. 4552, pp. 679–689, Springer Berlin Heidelberg (2007).

[Kor78] Kornhuber, H. H., *Blickmotorik*. In: Gauer; Kramer & Jung (editors), *Physiologie des Menschen*, vol. 13, pp. 357–426, Urban und Schwarzenberg, München (1978).

[Lam07] Lambooij, M. T. M.; IJsselsteijn, W. A. & Heynderickx, I., *Visual Discomfort in Stereoscopic Displays: A Review*. In: Woods, A. J.; Dodgson, N. A.; Merritt, J. O.; Bolas, M. T. & McDowall, I. E. (editors), *Proceedings of SPIE*, vol. 6490, pp. 64900I/1–64900I/13, SPIE (2007).

[Lam09a] Lambooij, M.; Fortuin, M.; Ijsselsteijn, W. a. & Heynderickx, I., *Measuring Visual Discomfort associated with 3D Displays*. In: Woods, A. J.; Holliman, N. S. & Merritt, J. O. (editors), *Proceedings of SPIE-IS&T Electronic Imaging*, vol. 7237, pp. 72370K/1–72370K/12 (2009).

[Lam09b] Lambooij, M.; IJsselsteijn, W.; Fortuin, M. & Heynderickx, I., *Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review*. In: *Journal of Imaging Science and Technology*, vol. 53(3), pp. 030201/1–030201/14 (2009).

[Lee12] Lee, J. W.; Cho, C. W.; Shin, K. Y.; Lee, E. C. & Park, K. R., *3D gaze tracking method using Purkinje images on eye optical model and pupil*. In: *Optics and Lasers in Engineering*, vol. 50(5), pp. 736–751 (2012).

[Mar99] Marcos, S.; Moreno, E. & Navarro, R., *The depth-of-field of the human eye from objective and subjective measurements*. In: *Vision Research*, vol. 39(12), pp. 2039–2049 (1999).

[Men09] Mendiburu, B., *3D Movie Making - Stereoscopic Digital Cinema from Script to Screen*. Elsevier, Burlington (2009).

[Men11] Mendiburu, B., *3D TV and 3D Cinema*. Focal Press, Waltham, USA (2011).

[Mül11] Müller, K.; Merkle, P. & Wiegand, T., *3-D Video Representation Using Depth Maps*. In: *Proceedings of the IEEE*, vol. 99(4), pp. 643–656 (2011).

[Ogl64] Ogle, K. N., *Researches in binocular vision.* Hafner, New York (1964).

[Pal61] Palmer, D., *Measurement of the Horizontal Extent of Panum's Area by a Method of Constant Stimuli*. In: *Optica Acta: International Journal of Optics*, vol. 8(2), pp. 151–159 (1961).

[Pan58] Panum, P. L., *Physiologische Untersuchungen über das Sehen mit zwei Augen*. Schwers, Kiel (1858).

[Pas97] Pastoor, S. & Wöpking, M., *3-D displays: A review of current technologies*. In: *Displays*, vol. 17(2), pp. 100–110 (1997).

[Pfe08] Pfeiffer, T.; Latoschik, M. E. & Wachsmuth, I., *Evaluation of Binocular Eye Trackers and Algorithms for 3D Gaze Interaction in Virtual Reality Environments*. In: *Journal of Virtual Reality and Broadcasting*, vol. 5(16) (2008).

[Rea] RealD, *RealD XL*.
Online: `http://www.reald.com/#/xl` (Date accessed: 2016-09-19).

[Rea15] Read, J., *If 3D Movies Make You Feel Sick, It's Likely All In Your Mind* (2015).

Online: `http://www.gizmodo.com.au/2015/07/if-3d-movies-make-you-feel-sick-its-likely-all-in-your-mind/` (Date accessed: 29.08.2016).

[Rei10] Reichelt, S.; Häussler, R.; Fütterer, G. & Leister, N., *Depth cues in human visual perception and their realization in 3D displays*. In: Javidi, B.; Son, J.-Y.; Thomas, J. T. & Desjardins, D. D. (editors), *Proceedings of SPIE*, vol. 7690, pp. 76900B/1–76900B/12 (2010).

[Ric70] Richards, W., *Stereopsis and Stereoblindness*. In: *Experimental Brain Research*, vol. 10(4), pp. 380–388 (1970).

[Rol09] Rolfs, M., *Microsaccades: Small steps on a long way*. In: *Vision Research*, vol. 49(20), pp. 2415–2441 (2009).

[Rub97] Rubin, G. & West, S., *A Comprehensive Assessment of Visual Impairment in a Population of Older Americans*. In: *Investigative Ophthalmology & Visual Science*, vol. 38(3), pp. 557–568 (1997).

[Sam15] Samsung, *Alles in 3D - Gehen Sie auf Entdeckungsreise* (2015). Online: `http://www.samsung.com/de/entdecken/entertainment/alles-in-3d-gehen-sie-auf-entdeckungsreise/` (Date accessed: 19.09.2016).

[San12a] Sandrew, B. B., *3D Movies and the Primal Brain* (2012). Online: `http://bsandrew.blogspot.de/2012/02/3d-movies-and-primal-brain.html` (Date accessed: 29.08.2016).

[San12b] Sandrew, B. B., *Engaged in 2D and Immersed in 3D* (2012). Online: `http://bsandrew.blogspot.de/2012/01/passive-voyeurs-look-at-2d-and-3d-movie.html` (Date accessed: 29.08.2016).

[Sch84] Schor, C.; Wood, I. & Ogawa, J., *Binocular sensory fusion is limited by spatial resolution*. In: *Vision Research*, vol. 24(7), pp. 661–665 (1984).

[Sch05] Schreer, O.; Kauff, P. & Sikora, T., *3D Videocommunication*. John Wiley & Sons, Ltd (2005).

[She34] Sheard, C., *The prescription of prisms*. In: *American Journal of Optometry*, vol. 11(10), pp. 364–378 (1934).

[She03] Sheedy, J. E.; Hayes, J. & Engle, J., *Is all Asthenopia the Same?* In: *Optometry & Vision Science*, vol. 80(11), pp. 732–739 (2003).

[Shi11] Shibata, T.; Kim, J.; Hoffman, D. M. & Banks, M. S., *The zone of comfort: Predicting visual discomfort with stereo displays*. In: *Journal of Vision*, vol. 11(8) (2011).

[Smi12] Smith, M. D. & Collar, B. T., *Perception of size and shape in stereoscopic 3D imagery*. In: Woods, A. J.; Holliman, N. S. & Favalora, G. E. (editors), *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics*, pp. 82881O/1–82881O/31, International Society for Optics and Photonics (2012).

[Son] Sony, *LKRL-A502 PACK*.
Online: `http://www.sony.de/pro/product/digital-cinema-3d-projection/lkrl-a502-pack/overview/` (Date accessed: 19.09.2016).

[Špa12] Špakov, O., *Comparison of eye movement filters used in HCI*. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 281–284, ACM Press, New York, USA (2012).

[Sta97] State, A.; Ackerman, J.; Hirota, G.; Lee, J. & Fuchs, H., *Dynamic Virtual Convergence for Video See-through Head-mounted Displays: Maintaining Maximum Stereo Overlap throughout a Close-range Work Space*. In: *Proceedings of IEEE and ACM International Symposium on Augmented Reality*, pp. 137–146 (1997).

[Tau] Tauer, H., *3D-Calc - stereoscopic calculators for computer, tablets and mobile devices*.
Online: `http://www.stereo-3d-info.de/3d-calculator.html` (Date accessed: 08.11.2016).

[Tob11] Tobii Technology, *Accuracy and precision test method for remote eye trackers* (2011).
Online: `http://www.tobiipro.com/learn-and-support/learn/how-do-tobii-eye-trackers-work/what-affects-the-accuracy-and-precision-of-an-eye-tracker/` (Date accessed: 29.08.2016).

[Tob14] Tobii Technology, *Tobii X2-60 Eye Tracker Product Description* (2014).

Online: http://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-x2-product-description.pdf/?v=1.0 (Date accessed: 19.09.2016).

[Tod04] Todd, J., *The visual perception of 3D shape*. In: *Trends in Cognitive Sciences*, vol. 8(3), pp. 115–121 (2004).

[Toy14] Toyama, T.; Sonntag, D.; Orlosky, J. & Kiyokawa, K., *A Natural Interface for Multi-focal Plane Head Mounted Displays Using 3D gaze*. In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pp. 25–32, ACM Press, New York, USA (2014).

[Uji11] Ujike, H. & Watanabe, H., *Effects of stereoscopic presentation on visually induced motion sickness*. In: *Stereoscopic Displays and Applications XXII*, pp. 786314/1–786314–6 (2011).

[Urv12] Urvoy, M.; Barkowsky, M.; Cousseau, R.; Koudota, Y.; Ricordel, V.; Le Callet, P.; Gutierrez, J. & Garcia, N., *NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences*. In: *Fourth International Workshop on Quality of Multimedia Experience*, pp. 1–6, Yarra Valley, Australia (2012).

[Wan14] Wang, R. I.; Pelfrey, B.; Duchowski, A. T. & House, D. H., *Online 3D Gaze Localization on Stereoscopic Displays*. In: *ACM Transactions on Applied Perception*, vol. 11(1), pp. 1–21 (2014).

[War95] Ware, C., *Dynamic stereo displays*. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 310–316, ACM Press, New York, USA (1995).

[Wat05] Watt, S.; Akeley, K.; Ernst, M. & Banks, M., *Focus cues affect perceived depth*. In: *Journal of Vision*, vol. 5(10), pp. 834–862 (2005).

[Wel06] Welch, G. & Bishop, G., *An Introduction to the Kalman Filter*. Tech. rep., Universitiy of North Carolina at Chapel Hill, Department of Computer Science, Chapel Hill (2006).

[Whe38] Wheatstone, C., *Contributions to the Physiology of Vision.–Part the*

First. On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. In: *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394 (1838).

[Wib14] Wibirama, S. & Hamamoto, K., *3D Gaze Tracking on Stereoscopic Display Using Optimized Geometric Method*. In: *IEEJ Transactions on Electronics, Information and Systems*, vol. 134(3), pp. 345–352 (2014).

[Wis10] Wismeijer, D. A.; Erkelens, C. J.; van Ee, R. & Wexler, M., *Depth cue combination in spontaneous eye movements*. In: *Journal of Vision*, vol. 10(6) (2010).

[Woo10] Woods, A., *Understanding Crosstalk in Stereoscopic Displays*. In: *Three-Dimensional Systems and Applications Conference*, Tokyo, Japan (2010).

[Woo12] Woods, A. J., *Crosstalk in stereoscopic displays: a review*. In: *Journal of Electronic Imaging*, vol. 21(4), pp. 040902/1–040902/21 (2012).

[XPAa] XPAND, *XPAND Active Shutter 3D*.
Online: `http://www.xpand.me/technology/xpand-3d/` (Date accessed: 19.09.2016).

[XPAb] XPAND, *XPAND Passive 3D Polarization Modulator Gen2*.
Online: `http://www.xpand.me/products/passive-gen2.1.html` (Date accessed: 19.09.2016).

[Xu12] Xu, D.; Coria, L. E. & Nasiopoulos, P., *Quality of Experience for the Horizontal Pixel Parallax Adjustment of Stereoscopic 3D Videos*. In: *2012 IEEE International Conference on Consumer Electronics*, pp. 394–395, IEEE (2012).

[Yan04] Yano, S.; Emoto, M. & Mitsuhashi, T., *Two factors in visual fatigue caused by stereoscopic HDTV images*. In: *Displays*, vol. 25(4), pp. 141–150 (2004).

[Yeh90] Yeh, Y.-Y. & Silverstein, L. D., *Limits of Fusion and Depth Judgment in Stereoscopic Color Displays*. In: *Human Factors: The Journal of* *the Human Factors and Ergonomics Society*, vol. 32(1), pp. 45–60 (1990).

[Zha13] Zhang, T.; He, S. & Zou, D., *Method of presenting three-dimensional content with disparity adjustments*. U.S. Patent 0162641 A1 (2013).

Publications by the Author

[Eic13a] Eickelberg, S., *Blickpunkt-adaptive Verschiebung der Stereo-3D-Konvergenzebene*. In: *15. ITG-Fachtagung für Elektronische Medien*, TU Dortmund University, Dortmund (2013).

[Eic13b] Eickelberg, S., *Reduktion von Ermüdungserscheinungen bei Stereo-3D*. In: *FKT*, pp. 382–386 (2013).

[Eic13c] Eickelberg, S. & Kays, R., *Gaze adaptive convergence in stereo 3D applications*. In: *2013 IEEE Third International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, pp. 1–5, IEEE (2013).

[Eic15] Eickelberg, S., *Perception of Dynamic Horizontal Image Translation in Stereo 3D Content*. In: *2015 IEEE Third International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, IEEE, Berlin (2015).

[Eic16] Eickelberg, S., *Evaluation of the Perception of Dynamic Horizontal Image Translation and a Gaze Adaptive Approach*. In: *Stereoscopic Displays and Applications XXVII*, IS&T, San Francisco (2016).

Student Works Supervised by the Author

[Alb14] Albers, F., *Entwicklung eines Verfahrens für die Erzeugung von Floating Stereoscopic Windows zur Vermeidung von Window Violations bei gleichzeitiger Minimierung retinaler Rivalitäten*. Bachelor thesis, TU Dortmund University (2014).

[Bec15] Beckmann, R., *Aufbereitung von fehlerbehafteten Disparitätskarten auf einem Parallelprozessor*. Bachelor thesis, TU Dortmund University (2015).

[Bru15] Bruchhaus, J., *Verbesserung der Bildqualität synthetisierter Stereo-3D-Zwischenansichten auf einem Parallelprozessor*. Bachelor thesis, TU Dortmund University (2015).

[Lau14] Laukamp, D., *Erzeugung von Zusatzansichten für die Realisierung von Bewegungsparallaxe auf einem Stereo-3D-Display*. Master thesis, TU Dortmund University (2014).

[Özt15] Öztabakci, E., *Entwicklung eines Versuchsaufbaus für die subjektive Bewertung der Bildqualität*. Bachelor thesis, TU Dortmund University (2015).

[Pat14] Patzke, S., *Einfluss der Hintergrundbeleuchtung auf die Wahrnehmung von Stereo-3D-Inhalten*. Bachelor thesis, TU Dortmund University (2014).

[Sch13] Schneider, C., *Implementierung einer künstlichen Schärfentiefenreduktion für Stereo-3D-Sequenzen auf einem Parallelprozessor*. Diploma thesis, TU Dortmund University (2013).

[Tel12] Telders, P., *Implementierung eines Verfahrens zur Tiefenschätzung in Bilddaten durch Analyse von Texturgradienten auf einem Parallelprozessor*. Bachelor thesis, TU Dortmund University (2012).

[Wer16] Wermers, J., *Echtzeit-Verfahren zur 3D-Blickpunktbestimmung in Stereo-3D-Anwendungen*. Master thesis, TU Dortmund University (2016).

[Xu14] Xu, J., *Implementierung einer künstlichen Schärfentiefenreduktion mit Bokeh auf einem Parallelprozessor*. Bachelor thesis, TU Dortmund University (2014).

Additional Tables for the Experiments

Table A.1: Number, classification, and age of subjects in all experiments conducted for this work.

General statistic	Exp. (I) (Sec. 5.3.2)	Exp. (II) (Sec. 5.3.3)	Exp. (III) (Sec. 5.3.4)	Exp. (IV) (Sec. 5.3.5)
Total	26	31	19	36
Rejected	2	3	2	6
Accepted	24	28	17	30
Statistic of accepted subjects				
Females	3	5	3	4
Test-experienced[a]	7	13	2	1
Experts	4	6	2	2
Research Assistants	7	3	0	9
Students	16	25	17	17
Other occupation	1	0	0	4
Minimum Age	21	22	23	22
Maximum Age	31	31	30	36
Average Age	25.04	24.89	24.65	26.27
Standard Deviation	2.84	1.75	1.80	3.18

[a] Subjects with prior experience in subjective image quality evaluation.

Table A.2: Visual performance of accepted subjects in all experiments and reasons for rejections in gray. All values represent numbers of subjects.

Visual acuity	Exp. (I) (Sec. 5.3.2)	Exp. (II) (Sec. 5.3.3)	Exp. (III) (Sec. 5.3.4)	Exp. (IV) (Sec. 5.3.5)
100%	10	13	6	19
100%, corrected[a]	10	13	10	6
80%	3	1	1	3
80%, corrected[a]	1	1	0	2
$<$ 80% (reject)	1	3	1	1
Stereo acuity[b]				
40”	20	27	9	21
50”	1	0	1	4
60”	1	1	2	1
80”	0	0	2	3
140”	2	0	3	1
$>$ 140” (reject)	1	0	1	0
Color perception				
Unimpaired	21	27	17	28
Mild Deuteranopia	3	1	0	2
Other rejections				
Bad eye tracking	-	-	-	5

[a] Subjects wearing either glasses or contact lenses.
[b] Values are seconds in angle of stereopsis.

Table A.3: Sequence overview of experiment (III). All sequences except Engineering were taken from RMIT3DV [Che12] and EBU [EBU] S3D test sequence librarires.

	Shift speed (°/s)		Disparity (px)			Convergence disparity (px)			
Sequence	Mean	Max.	Budget	Min.	Max.	Min.	Max.	Mean	Std.
RMIT3DV 21	0.026	0.128	14	-45	-31	-21.59	-16.38	-19.74	1.53
RMIT 01	0.032	0.193	15	-39	-24	-19.00	-12.00	-15.64	2.33
RMIT 27	0.041	0.193	15	-3	12	-1.50	5.81	2.30	2.30
RMIT 33	0.044	0.128	35	-11	24	-5.03	6.99	0.75	3.87
RMIT 13	0.046	0.257	24	5	29	3.48	13.75	7.40	2.67
RMIT 43	0.048	0.257	29	-69	-40	-33.59	-22.29	-27.29	3.22
RMIT 16	0.048	0.321	29	7	36	3.82	13.47	6.52	2.71
RMIT 28	0.054	0.193	32	23	55	12.63	24.99	17.70	3.08
EBU “Fountain Dancer 2 F474”	0.055	0.321	24	-26	-2	-11.44	-1.50	-6.20	3.13
RMIT 19	0.058	0.257	30	-12	18	-6.00	6.00	-0.51	3.92
RMIT 26	0.064	0.385	29	-6	23	-2.47	9.41	3.08	3.57
RMIT 32	0.068	0.321	32	-35	-3	-17.00	-4.94	-11.87	3.81
RMIT 30 F499	0.071	0.128	70	-29	41	-14.49	1.51	-9.14	4.28
EBU “Dancer Normal”	0.072	0.513	28	-28	0	-13.96	-3.19	-7.38	4.17
EBU “Dancer Low”	0.078	0.449	28	-29	-1	-14.50	-2.66	-8.71	4.05
EBU “Lupo Confetti”	0.079	0.257	33	-28	5	-13.53	2.09	-6.22	5.05
RMIT 38	0.085	0.385	37	-2	35	1.06	16.94	7.11	5.38
RMIT 31	0.088	0.193	40	-6	34	-2.03	15.31	5.67	5.47
Engineering	0.098	0.257	92	23	116	15.73	44.87	33.54	8.82
RMIT 29	0.100	0.257	73	-28	45	-2.77	19.09	11.72	5.90
RMIT 03	0.101	0.321	34	-85	-51	-41.51	-26.94	-33.19	5.62
RMIT 30 F251	0.106	0.321	73	-27	46	-13.64	9.78	-4.71	6.03
EBU “Fountain Dancer 2 F573”	0.110	0.449	42	-26	16	-11.53	7.55	-1.68	5.73
RMIT 46	0.114	0.193	62	-70	-8	-34.41	-11.38	-25.40	6.93
EBU “Lupo Hands”	0.131	0.513	46	-37	9	-16.03	3.43	-9.24	6.05
RMIT 14	0.142	0.385	62	-21	41	-9.60	14.99	5.75	6.91

Curriculum Vitae

Personal Information

Surname	Eickelberg
First name	Stefan
Date of birth	12.07.1986
Place of birth	Herdecke, Germany

Education

09/2011 - 04/2017	PhD degree (Dr.-Ing.) in Electrical Engineering and Information Technology, Faculty of Electrical Engineering & Information Technology, TU Dortmund University, Germany
10/2006 - 07/2011	Diploma degree (Dipl.-Ing.) in Information Technology, TU Dortmund University, Germany
06/2006	University-entrance diploma (Abitur), Gymnasium an der Schweizer Allee, Dortmund, Germany

Professional Experience

Since 05/2017	Software engineer, ISRA Surface Vision GmbH, Herten, Germany
09/2011 - 12/2016	Research assistant, Communication Technology Institute, TU Dortmund University, Germany
2009 - 2010	Intern (15 weeks), Fujitsu Deutschland, Augsburg, Germany
08/2008 - 07/2009	Student assistant, Circuits and Systems Lab, TU Dortmund University, Germany