Interactive Stereoscopic Content for Multiple Co-Located Users

Stefan M. Niederhumer

MASTERARBEIT

eingereicht am
Fachhochschul-Masterstudiengang

Interactive Media

in Hagenberg

im September 2016

Declaration

I hereby declare and confirm that this thesis is entirely the result of my own original work. Where other sources of information have been used, they have been indicated as such and properly acknowledged. I further declare that this or similar work has not been submitted for credit elsewhere.

Hagenberg, September 26, 2016

Stefan M. Niederhumer

Contents

Preface

Before beginning with the actual topic, I would like to give thanks to the *Playful Interactive Environments* research group, in particular to the supervisor of this thesis, Jeremiah Diephuis MA, for providing the necessary environment to perform development and tests of the systems created alongside this publication. The support given from the initial concept phases to the final written pages was very much appreciated, as was the opportunity to be able to use such an installation and have access to it for testing the solutions that were devised in this thesis.

Thank you.

Kurzfassung

Systeme, welche die Illusion von Tiefe durch die Darstellung von zwei separaten Bildern vermitteln, verwenden eine Technik namens Stereoskopie. Stereoskopische Systeme sind weit verbreitet und bereits seit Beginn des 20. Jahrhunderts kommerziell erhältlich. Seit der Entstehung von stereoskopischen Systemen wurden mehrere Wege entwickelt diese zu erzielen, jedoch bieten die Meisten dieser Systeme keine Interaktionsmöglichkeiten, oder schränken diese auf einen einzelnen Benutzer ein.

Diese Arbeit wird bereits existierende Systeme, welche interaktive stereoskopische Erlebnisse für mehrere am selben Ort befindliche Benutzer bereitstellen, vorstellen, deren Stärken und Schwächen aufzeigen und gemeinsame Herausforderungen, welche von solch einem System behandelt werden müssen, aufzeigen. Basierend auf diesen Problemen wird eine Lösung zur Optimierung eines solchen interaktiven Systems vorgestellt und implementiert. Das fertige System wird dann durch eine heuristische Auswertung von drei Experten begutachtet um mögliche Probleme zu erkennen. Die Ergebnisse dieser Begutachtung werden bereitgestellt und dienen als Grundlage, um weitere Verbesserungen am System vorzuschlagen. Das Hauptziel dieser Arbeit ist der Versuch ein funktionierendes System zu erstellen, welches als Grundlage für die Entwicklung von interaktiven, stereoskopischen Anwendungen dienen kann und wichtige Richtlinien für solch ein System etabliert.

Abstract

Systems to create the illusion of depth via two separate images utilize a technique called stereoscopy. Stereoscopic systems are by no means new, as patented systems for commercial use can already be found at the turn of the 20th century. Since the beginning of stereoscopy, multiple ways of achieving the illusion of depth have been proposed, but most of them either do not provide any interactivity or are only controlled by a single user.

This thesis will introduce existing systems capable of delivering stereoscopic experiences to multiple co-located users, identify their strengths and weaknesses and extract common problems that have to be tackled by such systems. Based on these issues, a solution to optimize an interactive stereoscopic experience is proposed and implemented. The implemented system is then made subject to a heuristic evaluation by three experts with experience in creating co-located systems to find possible usability issues. The results of said evaluation are then made available and serve as a basis to propose further enhancements of the devised system. The overall goal of this thesis is to try and provide a working system which can be used to design interactive experiences for multiple co-located users and provide guidance about which aspects need to be considered when designing such a system.

Chapter 1

Introduction

1.1 Motivation

Stereoscopy has been available for a long period time in different formats. Early versions used red and cyan colored images on paper in conjunction with glasses with lenses tinted in the same color to transport a different image into each eye of a user, allowing them to perceive depth in those simple images. Nowadays, stereoscopy is used in cinemas around the world to enhance the experience while watching a feature film, most televisions are able to display three-dimensional images and even hand-held devices like the *Nintendo 3DS*[1] or the *LG Optimus 3D*[2] are capable of delivering 3-D content. In addition, larger public installations like the *Ars Electronica Deep Space* or the *CAVE* are able to provide immersive experiences utilizing stereoscopy.

With all the advances and reach of this technology, it still has one common drawback. Most systems available today are only designed to accommodate and provide interactive content to a single user or, if they are capable of catering to a larger number of users, do not provide an interactive experience for each user. Cinemas for example only provide one fixed perspective and do not provide interactivity of any kind. The *CAVE* system on the other hand is able to provide an interactive experience, but it is only controlled by a single user, leaving everyone else to play the role of an observer.

Some research has already been performed in regards to providing interactive stereoscopic content to multiple users, but has mostly dealt with a fixed number of users that could not easily be changed. This thesis is motivated by the lack of systems capable of dealing with larger, dynamic groups of users and will try to establish a system capable of accommodating such a group.

[1] See http://www.nintendo.com/3ds.
[2] See http://www.lg.com/uk/mobile-phones/lg-P920-optimus-3d.

1.2 Projected Goals

The overall goal of this thesis will be to provide a system capable of serving a variable number of co-located users with interactive stereoscopic content. To achieve this goal, the first step will be to identify the problems that need to be solved when dealing with multiple users inside such a system. This will be done by building on existing research and observing users inside existing stereoscopic systems which are capable of housing multiple users, even though they do not provide an interactive perspective to said content.

In compiling a list of challenges that need to be dealt with when designing such a system, it will then be possible to tackle each issue and provide solutions that fit well enough within the boundaries of available technology, hence optimizing the experience for multiple users. As the solution to some of the found problems will undoubtedly have an impact on the other issues, balancing each issue will be a crucial step in providing a pleasant overall experience for all users of the system.

To be able to perform tests in a realistic environment, some form of use case will then be devised that will implement the proposed solutions. The resulting application will then be made subject to an expert evaluation, inspecting each issue by checking against a list of defined heuristics which are designed to find any shortcomings of the system in regards to its ability to accommodate multiple co-located users. The results of the evaluation and the system's implementation will then be made available with this thesis.

1.3 Structure

Chapter 2 describes the term stereoscopy in more detail, showing different methods of achieving said effect and detailing their strengths and weaknesses. Furthermore, systems that provide stereoscopic content for multiple users are introduced. This introduction includes the underlying methods used for displaying stereoscopic images, how many users are accommodated by the system and what its drawbacks are.

Following this introduction, chapter 3 puts forth issues that arise once multiple users are introduced to a stereoscopic system. These include user tracking, assigning each user to a specific group, how each group's viewport is displayed, how a comfortable experience can be ensured and how a projection needs to adapt, in order to provide a correct perspective, depending on each group's position.

After each issue has been introduced and explained, chapter 4 discusses possible solutions for each problem. Applying the same structure as chapter 3, the proposed solutions for each issue are then laid out and any implications on other areas of the system detailed, weighing each side effect against the benefit that the associated solution has. In addition to a viable multi-user system, section 4.6 of this chapter also introduces a use case built around an implementation of the described system.

Chapter 5 elaborates on the procedure that has been used to evaluate the created system. After a short introduction to expert evaluation, the performed steps are detailed in the following section, including a brief overview over the installation used for assessing the viability of the proposed solution. The final section of this chapter begins by detailing how the evaluation's results were processed, followed by a list of issues that were compiled from the result of the performed heuristic evaluation.

The final chapter 6 summarizes the findings acquired during the writing of this thesis and concludes by providing a retrospective of the evaluation results and proposing additional enhancements which could be made to the system, but were omitted due to time constraints or limitations of the available hardware.

Chapter 2

State of The Art

The following section will give an overview over the different types of stereoscopic systems. Furthermore, a list of systems which try to serve stereoscopic content to multiple users will be presented and their capabilities detailed.

2.1 Stereoscopy

Systems to create the illusion of depth via two separate images, one for each eye of a potential viewer, use a technique called stereoscopy. Stereoscopic systems are by no means new. Patented systems for commercial use can already be found at the turn of the 20th century, such as a stereoscopic motion picture device by Hammond [16]. Since the beginning of stereoscopy, multiple ways of achieving the illusion of depth have been proposed, which Urey et al [24] divides into three categories, stereoscopic direct view, head-mounted stereoscopy and autostereoscopic direct view.

2.1.1 Stereoscopic Direct View

Direct view approaches utilize a display which is directly viewed using some sort of special glasses [24], hence the name. Two images, one for each viewer's eye, are combined and projected onto a display, to then be separated and directed into the corresponding eye of a viewer by the previously mentioned glasses [10]. To combine both images, color, polarization and time-multiplexing can be used to combine left and right eye images [9, 10, 20].

Color multiplex uses complimentary colors to encode the left and right eye images, which can then be separated by anaglyph filters. The most common color combinations are red for the left channel and cyan for the right channel. A notable recent example for using color multiplexing is the video game *Fez*[1] which can be played in an anaglyph mode after beaten twice.

[1] https://en.wikipedia.org/wiki/Fez_(video_game)

Figure 2.1: Color, polarization and time multiplexing glasses for stereoscopic viewing of images. Not the bulky frame of the shutter glasses on the right, as they have to house additional electronics which the other two types of glasses do not need.

Polarization multiplex uses light with two different states of polarization for image encoding. In order for this to work, both states have to be positioned orthogonally to one another. Either linear or circular polarization is possible, the latter however allows more freedom in head movement. Most modern cinemas employ this technology to display movies produced with a stereoscopic camera setup.

Time multiplex approaches alternate between left and right eye images on the display and utilize shutter glasses. These glasses are synchronized to the displayed content and thus allow the perception of depth. This system is more prominent in 3-D projectors available for end users, but is also used in some cinemas.

Figure 2.1 depicts different glasses for each of the mentioned multiplexing methods. Notable is that time multiplexing glasses are the only type of multiplexing technology that actually need some form of electronics and therefore a power source integrated into the glasses.

2.1.2 Autostereoscopic Direct View

Similar to stereoscopic direct view, this method also utilizes a directly viewed display, but no special glasses are needed to perceive the 3-D effect. Autostereoscopic views are capable of providing the same stereo image to multiple viewing zones in different positions. Autostereoscopic displays can be further enhanced by employing head tracking. This allows the normally fixed viewing zones to adapt their position according to each viewer [23]. Notable examples of such technology include so called holographic images and the *Nintendo 3DS* seen in figure 2.2, as they are widely available and use autostereoscopy, with the successor to the *Nintendo 3DS* even employing head tracking to increase the effect.

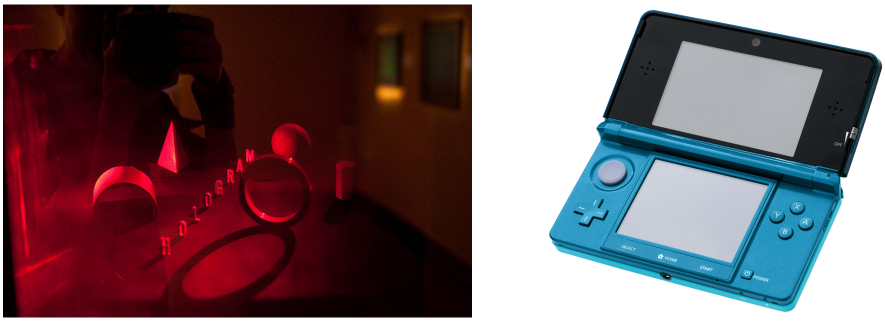


Figure 2.2: A holographic image on the left and the *Nintendo 3DS* on the right. Both employ autostereoscopy to display 3-D images without needing additional equipment.

2.1.3 Head Mounted Display

Instead of having a screen or a projector display two images in a time, color or polarization multiplexed manner, this category of stereoscopic devices employs two screens, or one split screen to render the target image for each eye directly in front of it. This enables a high degree of mobility without breaking stereoscopy, as the device is mounted on a user's head and is therefore in front of the user's eyes at all times.

This approach has some advantages over traditional direct view systems. Due to each image only being displayed directly in front of the desired eye, no special image separation technique is needed to ensure that it is hidden from the other eye. However, due to the close proximity between eyes and display, a high field of view and resolution need to be achieved for satisfying results [24]. Furthermore, prolonged exposure to such a device can cause fatigue and even cybersickness [5].

2.2 Multi-User Stereoscopy

To extend the amount of users of a single stereoscopic system, several different attempts have been made, a selection of which will be presented in the following sections. To evaluate them, apart from the amount of users accommodated, three performance parameters provided by Fröhlich et al in [9] will be used.

Brightness per view measures the delivery of light to each viewers' eye. This affects the overall perceived brightness when viewing stereoscopic content and should be kept at appropriate levels for the viewed content, as to not cause fatigue.

When not only the target eye image is visible, but parts of the other eye or even images dedicated for other users are as well, crosstalk is occurring.

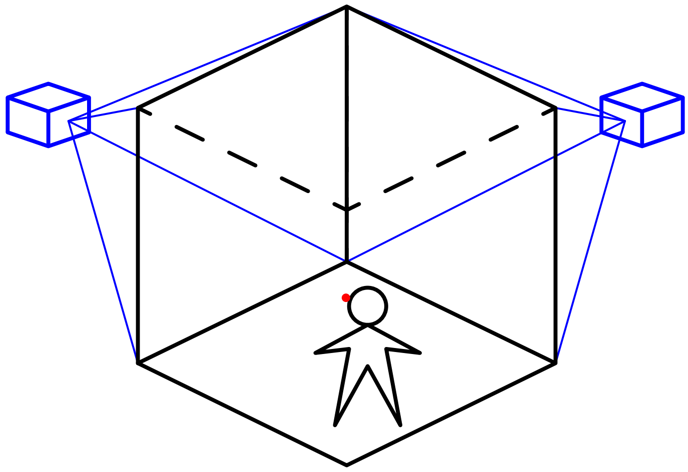


Figure 2.3: The basic setup of a *CAVE* system, as described by [4]. The projection system is depicted in blue, the position tracker attached to the user's glasses in red. To provide a simpler graphic, only two walls are depicted with projected images. A full *CAVE* system would include projected images on each surface inside the cube shaped room.

As this influences the ability to perceive the illusion of depth and can cause eye strain, it should be kept to a minimum, if not avoided completely.

Perceived flicker can be caused by a multitude of reasons. Low video rate, a low shutter frequency for active stereoscopy and low brightness levels can all contribute to this effect, which degrades the experience inside a stereoscopic system as a whole.

2.2.1 CAVE

The *CAVE* automatic virtual environment was design as an interactive virtual reality interface and was first introduced by Cruz-Neira et al in [4]. It consists of a cube-shaped room with stereoscopic images projected onto each wall, as seen in figure 2.3. Users inside the *CAVE* system are required to wear special glasses which enable them to view the stereoscopic images as such. In addition, one pair of glasses is equipped with a position tracking system which allows the projected images to be rendered according to the user's view. This is called view-centered perspective. Since only one user is tracked, all users share the same perspective, regardless of their position inside the system.

Since users inside a *CAVE* share the same view, crosstalk is a non-issue. Brightness and flicker of the viewed image is also not hindered by any design aspect of the system and therefore only dependent on the stereoscopic hardware in use.

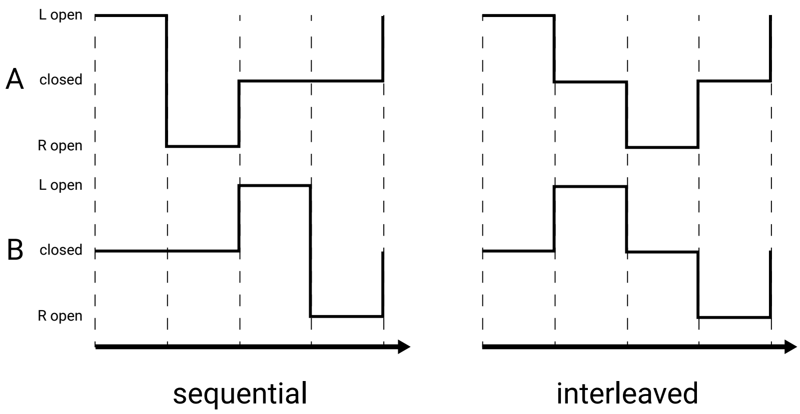


Figure 2.4: Viewer sequential and viewer interleaved image sequencing as described by Agrawala et al in [2]. Sequential alternates eyes first, then viewers. Interleaved alternates first the viewers and then eyes, leading to an improved perceived brightness.

2.2.2 3-D Stereo for Two Users Using a Single Display

De la Re et al [20] propose an inexpensive direct view stereoscopic solution for two users utilizing only one display. Combining time and color-multiplexing by using shutter glasses and anaglyph filters, each viewer is provided with different stereoscopic content. Due to the nature of anaglyph filters, this approach suffers from lost color information and high crosstalk. Since no position tracking was used, fixed positions are assumed for both users.

2.2.3 Two User Responsive Workbench

Utilizing a 120 Hz projector and shutter glasses, Agrawala et al [2] designed an interactive workbench for two simultaneous users. By displaying four images in either viewer sequential or viewer interleaved sequence, as can be seen in figure 2.4, a system was devised to serve two different viewpoints for each user. To stop crosstalk between both views, a custom blanking state was introduced which closes both eyes on the glasses corresponding to the inactive view. However, this results in noticeable flicker and loss of perceived brightness. To allow a view-centered perspective for both users, head tracking is also employed to allow the system to perform the necessary viewport correction.

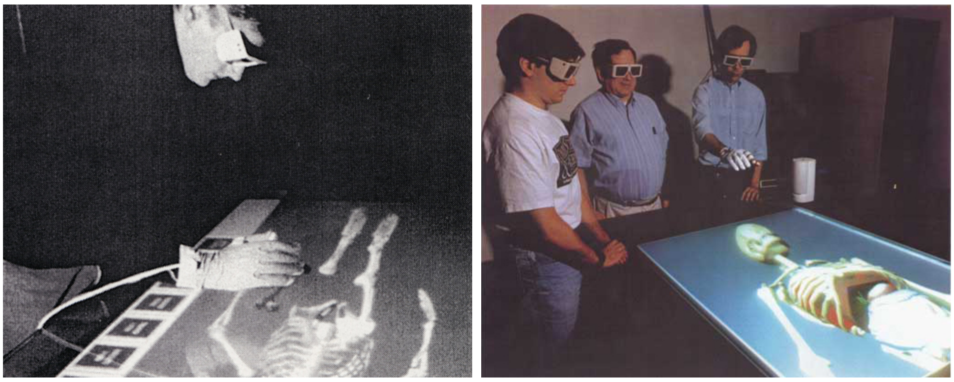


Figure 2.5: The responsive workbench in its original form by Agrawala et al [2] on the left and the improved version by Fröhlich et al [9] on the right.

2.2.4 Stereo for Two to Four Users

Building on the work of Agrawala et al [2], Fröhlich et al [9] extended her approach by adding one polarization-based stereoscopic projector per user and a custom shutter encompassing all projectors. Eye separation is now achieved by polarization and user separation by the custom shutter. By synchronizing the shutter position with the corresponding glasses, every inactive view is set to the aforementioned blanking state, eliminating cross talk. By increasing the frequency of the custom shutter to 320 Hz, the system could sustain four different stereo pairs, as seen in figure 2.5, with only slight flicker and acceptable brightness.

2.2.5 The Illusion Hole

Kitamura et al [11] describes a system using one display and a display mask to accommodate three or more users. By making only a single view visible for each user, as featured in figure 2.6, the need for a blanking state is eliminated. This prevents crosstalk, loss of brightness and flicker. However, the perceived resolution of each view is also decreased, as only parts of the display are used. In addition, head tracking was used to include correct perspective when rendering each view.

In 2006, Kitamura et al [12] extended this system by using polarization, instead of shutter glasses, decreasing the cost and size of the installation.

2.2.6 Laser-based Multi-user System

By using a custom made laser backlit lenticular display, Surman et al [23] managed to create an autostereoscopic display that renders a single stereoscopic image pair to multiple viewers, with minimum loss of resolution and brightness, no flicker and large depth of view. The lenses incorporated in

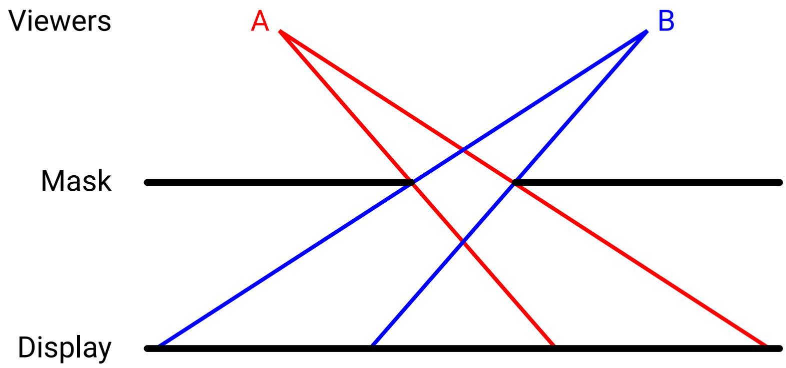


Figure 2.6: The basic working principle of the illusion hole of Kitamura et al [11, 12]. Due to the mask, different viewers observe different areas of the display, which contain a rendering tailored to their position.

this display, coupled with the precise laser light source, are able to form an image region, also called exit pupil, directly at the position of each viewer's eyes. Using head tracking, these exit pupils can be moved in accordance with the position of each individual viewer, but since only one stereo pair is displayed, separate views for each user are not possible.

2.3 Progress

Significant work has been put into building stereoscopic environments for multiple users, as can be seen by the systems introduced in this chapter. Despite these efforts, there is still more ground to cover, as evident by the still ongoing research in this field.

Chapter 3

Issues of Multiple Co-located Users

Looking at the different systems introduced in chapter 2, a common drawback which is shared by most systems can be observed. While some of these systems are capable of delivering a stereoscopic experience to multiple users, most do so only for a fixed number of users. If they allow additional users to be added to the system, it is only by employing additional hardware which has to be set up and configured, therefore not allowing users to drop in or out of the system at will. This section identifies different problems that arise when multiple co-located users should be accommodated inside an interactive stereoscopic system.

3.1 Tracking

In order to accommodate any number of users inside an interactive system of any kind, tracking information has to be collected. This data should at least include the position of a user, but may also contain additional information like the exact position of each foot of a person inside the system or orientation data to further increase the capabilities of the system to adapt itself to each user [4].

In addition to providing required data for the system, tracking also needs to have sufficient precision to provide an accurate enough translation of user movement. Depending on the size of the tracked area and the amount of users, a tracking system also has to be able to identify single users without missing positional information or generating false tracking points amidst larger groups. Speed is also an important factor, as an interactive experience should respond in real time in order for a user to not feel any delay between his or her action and the reaction of the system.

Should the system provide interaction with objects inside the virtual space depicted by the system, further processing is needed to allow this.

Any content or object which would be the target of any user interaction defines its location in a virtual coordinate space dictated by the system used for displaying content on screen. In most cases, this will differ heavily from the tracking coordinate system. To allow for interaction between objects in both the physical and virtual world, the position of each partaking object needs to be translated to the same coordinate space. This can be done by either converting physical positional information to its virtual equivalent or performing the opposite translation from virtual coordinates to physical space [2, p. 329]. It might also be of advantage for some use cases to construct a new shared space and transform both positions into it.

3.2 View Assignment

Depending on the available hardware, an interactive system may be capable of rendering and displaying more than one different view of said system using the same output method. If it is desired to have more than one distinct rendering of the system available, users will have to be assigned to one of the potential views.

Depending on the capabilities of the system used, grouping can initially be done based on the collected tracking information or might be fixed at the whole experience and based on where a user enters the tracking space or which viewing device taken from a set of preassigned devices is chosen. If dynamic group assignments due to the tracking information is possible, users may also be able to get reassigned to another view group while interacting with the system. This however not only depends on the projection system used or screen but also on the viewing devices, if no autostereoscopic system is in use. In most cases, special hardware or adaptations of existing hardware will be needed for dynamic group changes and therefore might not be supported at all.

If multiple views are supported, some users may become potential candidates for more than one view, therefore not allowing a clear view assignment. These cases need to be handled gracefully, as they have the potential to degrade the experience for a large number of users. Handling of such outliers also has to have the ability to adapt to different use cases, as some more fast paced experiences may encourage this issue to occur more frequently.

3.3 Viewport Display

Should multiple different views of the provided interactive content be desired, some way of displaying that content only to users assigned to that specific view has to be found. In choosing the appropriate method, the following issues have to be considered, as defined by Fröhlich et al in [9, p. 140].

3.3.1 Brightness

Fröhlich et al [9, p. 140] cites the delivery of sufficient light to the user's eye as one of the main challenges of creating a system that displays multiple views simultaneously, as viewing of stereoscopic images becomes far less comfortable without a high enough perceived brightness. How much light is transmitted to each user's eye is dependent on three factors: the initial brightness of the image, the frequency at which the display switches between different views and the stereoscopic technique in use.

The projector or screen used to display each view will need to have sufficient brightness to overcome any ambient light. If an image shown by the display is not clear enough when viewed without any additional interference from the stereoscopic viewing device, its odds of being sufficiently bright with those systems in place are slim.

Displaying a different view to a specific group implies that others are not able to see this particular group's view. This can be done in a variety of ways, depending on the number of groups desired and available hardware. Even with the same hardware, different techniques can be employed to improve perceived brightness, as can be seen in figure 2.4, which depicts two different ways to alternate between multiple viewers when using active stereoscopy.

As Kane et al [10] and other have shown through their research, the choice of stereoscopic technique will have a large impact on the perceived brightness and should be chosen in accordance to the amount of distinct views desired. This will also have an impact on the frequency at which different views can be displayed and should be taken into account when deciding on a desired technique.

3.3.2 Crosstalk

Since different views are displayed at the same time or in close succession, some users may see traces of an image that was not intended for their assigned group or even view images with an eye that was not intended to receive that particular image. This phenomenon depicted in figure 3.1 is called crosstalk. It should ideally be prevented entirely, as it interferes with the desired perception of depth and may even cause discomfort in some users.

Crosstalk can originate from a variety of sources, be it the projector systems [23, p. 746] or the type of stereoscopic enabling glasses being used [15, p. 188:5]. Depending on the cause of this effect, Fröhlich et al [9, p. 140] categorize this effect as either static or dynamic crosstalk.

Static crosstalk originates from the underlying system used to achieve stereoscopy and remains constant throughout the experience, hence the name. In the case of passive direct view stereoscopy, this kind of interference could be caused by using a low quality polarization filter which does

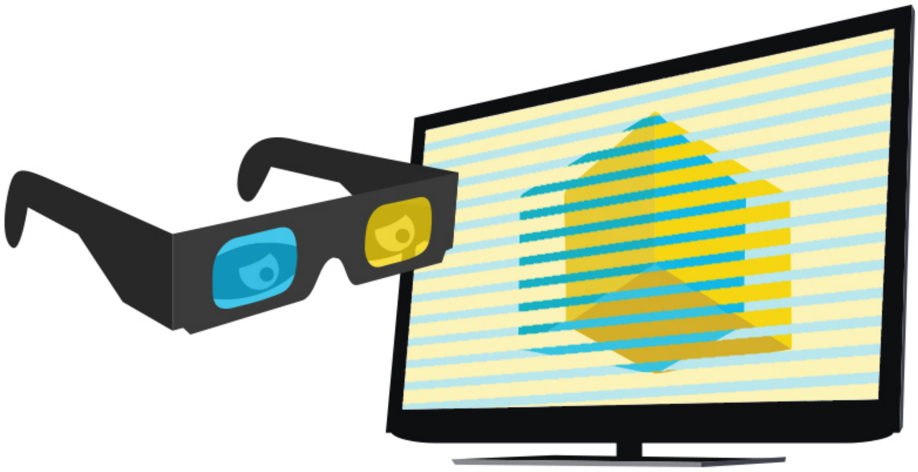


Figure 3.1: A stylistic depiction of the effect called crosstalk, taken from *Locafox.de*[25]. Both eyes are able to see images intended for only one of them, therefore making them uncomfortable to look at.

not entirely filter light polarized at a different angle as desired.

Dynamic crosstalk, on the other hand, is caused by any moving or dynamic parts of the system. Users of an active stereoscopic system could experience this phenomenon during the transition phases, when both shutters switch their state from opened to closed and vice versa, allowing one eye a glimpse of an image directed at the other and causing crosstalk.

3.3.3 Flicker

In order to experience the visuals and animation provided by the stereoscopic system as smooth, the amount of frames displayed for each individual view has to exceed a certain threshold. Even though that number slightly varies for each individual, a general consensus is that at least 30 frames per second[1] have to be displayed for each eye for a smooth experience. Flicker will be experienced if the frame count falls short of this number, which should ideally be prevented.

Apart from the frame rate of the rendered view, flicker can also be caused by brightness of the images and shutter frequency, if applicable to the system used[10, p. 15]. The amount of distinct views will also have a direct influence over the perceived flicker, as each addition causes the number of frames per eye for a single view to drop drastically. This cause of flicker is one of the greatest limiting factors to adding more than one view to a stereoscopic

[1] Not to be confused with the roughly 24 frames minimum required to perceive a sequence of images as motion.

system.

3.4 Viewport Continuity

In an interactive stereoscopic environment, some users may feel discomfort when moving while inside a system that provides the illusion of depth. This phenomena, more prominently known as cybersickness [5], can affect individuals to varying degrees, ranging from slight discomfort to vomit-inducing sickness. It is not entirely understood what causes the discomfort, but the most prevalent theory is the conflicting nature of the perceived motion in the rendered image and the absence of the experience of any such motion for each user [5].

Even though discomfort of this kind is mostly associated with head mounted displays and virtual reality systems, it seems that simply experiencing stereoscopic content can also cause nausea, eye strain and headaches, even though only a low level of presence is achieved [21, p. 785]. It is therefore important to keep any influences which might induce cybersickness to a minimum or prevent them entirely.

Since a group of users share a view, which is in turn adapted to tracking information gathered from that group, each movement by any of the users inside the system will translate into a position change of the associated view. In general, small amounts of movement within the system should result in equally small changes of the virtual view's position. If this simple constraint is violated, the movement of the rendered viewport will come unexpected so some users of the group which could bring forth discomfort [21]. It is therefore desirable to keep the discrepancy between movement of the users assigned to a viewport and the position changes of said viewport to a minimum to prevent large discrepancies between a previous and current rendering of the assigned viewport. This prevention of large viewport differences will be defined as viewport continuity, referencing a term for a similar property for mathematical functions [2].

Since the number of users inside the system is dynamic, their behavior and speed of movement will greatly vary. To cope with the diverse movement of users, the most common patterns which can cause problems with viewport continuity should be handled gracefully. These have been broken down into three distinct scenarios:

1. Since group assignments are updated continuously, some users may be assigned to a new group during these updates. Any simple viewport calculation performed by averaging each user's assigned position or a similar method will probably cause a dramatic change in each viewport's position, thereby breaking viewport continuity. An example of

[2] For more information on mathematical continuity, see https://en.wikipedia.org/wiki/Continuous_function

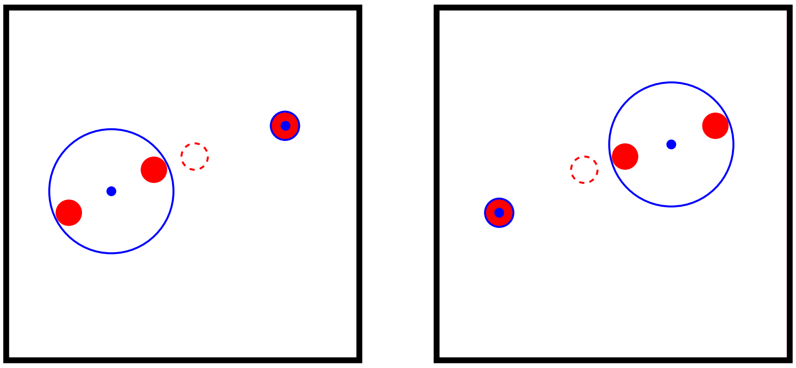
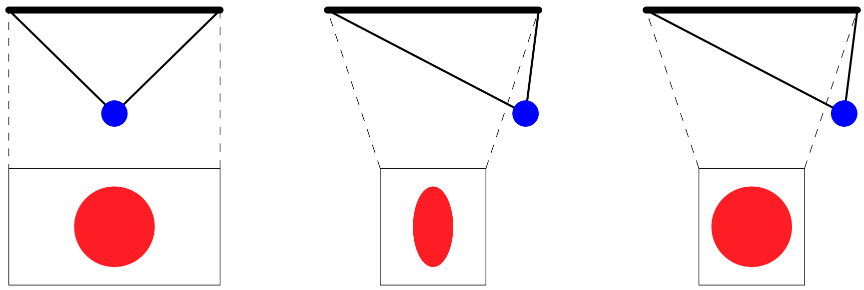


Figure 3.2: The images show blue view group assignments to red users during two successive view renderings, during which one user is moving. This examples calculates the new view position by averaging the location of each assigned user. This is insufficient to preserve viewport continuity, as can be seen by the sudden shift in view positions.

this can be seen in figure 3.2. As this issue degrades the experience of more than one group, it is especially important to consider and handle gracefully.

2. Similarly to users moving to another group, any user might enter or leave the tracked space of the system at any time, causing a similar change in view calculations and therefore a shift in perspective. Depending on the number of views supported by the current system, this may only cause problems for a single group of users, or all of them, if only one view is supported. In theory, this can be handled in a similar fashion to the previous scenario, but it has to be taken into account that a user might briefly exit the tracked spaced, only to enter it again shortly thereafter.
3. As the users inside the system might not always coordinate and move somewhat independently from each other, most users will not navigate the system at the same pace. A single, fast moving user or small sub-group of users may cause discomfort for more stationary users, if the moving group's position equally contributes to viewport location, causing the stationary group to experience a discrepancy between their actual movement and the visualized viewport position. Gracefully handling this proves difficult, as the way in which the speed of movement is handled will be highly dependent on the use case provided by the system. If fast movement is desired, it might not be possible to preserve viewport continuity and some users may experience increased discomfort as a result.



(a) Normal view (b) No correction (c) Correction

Figure 3.3: This is a simple depiction of the issues of viewing stereoscopic content on a system achieving stereoscopy via direct view, see section 2.1. The blue circle symbolizes the viewing position of the screen, the red circle a spherical object being displayed on screen. When perspective correction is applied, the red sphere will appear round, no matter where it is viewed from.

3.5 Perspective Correction

Depending on the medium used to achieve stereoscopy, some form of perspective correction has to be applied to the interactive content. This correction must use the data described in section 3.1 to ensure that the stereoscopic content is rendered in a correct way, in regards to the position of each user [2, 4]. If this correction is not applied, content may appear distorted when viewed from various angles, as can be seen in figure 3.3.

This issue applies exclusively to direct view stereoscopic systems, as head mounted displays can only be viewed from a single angle, due to the nature of how such devices are handled. Head mounted displays however suffer from one of the limitations initially described in chapter 2, namely needing an additional piece of hardware for each additional user that should experience the system and will therefore not be considered a valid option for a system with a dynamic number of users.

3.6 Consequences

Supplying multiple users with an acceptable experience within a stereoscopic environment is an elaborate problem, as detailed in this chapter. Some balance between each aspect of such a system will have to be found, since not all problems can be tackled independently.

Chapter 4

Accommodating Multiple Co-located Users

This chapter will detail an attempt to solve the problems describe in chapter 3. It is by no means a complete solution, as the set-up used and intended for the provided solution has physical as well as hardware limitations which cannot be overcome without adding additional hardware capabilities to it. Some suggestions on how additional capabilities could be added will be brought forth in chapter 6.

The last section of this chapter will also detail an implementation of the proposed solutions and a use case in the form of an interactive story experience implemented in `C#` using the *Unity* game engine. The implementation will also be made available as a *Unity* package, containing all the components needed for use in a custom use case.

4.1 Tracking

To track each user inside an area of width t_w and depth t_d, an array of laser rangers were used to retrieve positional information about each person inside the system. Since this system operates by sending and detecting reflection of infrared light, it is quite fast at supplying positional information [17, p. 16].

Due to the nature of the tracking, four laser rangers were arranged in a rectangular shape, in order to minimize any occlusion problems which might be caused by multiple users moving about the tracked area. This allows a large number of users inside the tracked area and can potentially be extended by adding additional laser rangers. However, this tracking system does not provide accurate enough information to discern the correct orientation of each user, therefore orientation data was not available.

To allow for easier interaction with the virtual environment, the incoming positional information is converted into a 2-D coordinate system described as (x, y) with $x, y \in [-1.0, +1.0]$. This was chosen to allow for easier handling of different sizes of tracking systems, as any calculations based on positional information will always be inside this defined range. Furthermore, any interactive objects will also be placed inside this coordinate space. Should a real world position be necessary for any calculations, such as positional sound, a simple multiplication of $(x \cdot t_w/2, y \cdot t_d/2)$ suffices to retrieve that position.

4.2 View Assignment

Partitioning data points into one or more groups based on their properties is the definition of a clustering problem [1]. Since this closely aligns with the problem described in section 3.2, it is a natural fit to employ clustering algorithms to solve the group assignment problem.

When clustering endless streams of data, classical clustering algorithms like *K-Means* have known problems. For one, it is assumed that the data to be clustered is fully known, does not change and that it is possible to access it multiple times [3]. Furthermore, those algorithms have no concept of the age of data which makes them unable to distinguish between older and newer data, therefore rendering them unable to detect evolving features of clusters over time [1].

Since the tracking information is updated in real time and the cluster will have to adapt in accordance to the position of each user inside the systems, a set of data stream clustering algorithms is employed to detect and capture the features of each group of users inside the system. In particular, the *D-Stream* algorithm described by Chen et al in [3] will be employed, as it has distinct advantages for the described problem.

D-Stream is able to detect arbitrary shapes in the clustered data, instead of simple spherical clusters. This allows the clusters to more closely resemble the actual groups of users inside the system, instead of assuming that each group is uniformly shaped.

Furthermore, *D-Stream* is also capable of automatically dealing with outliers. In the context of grouping positional information of users, this ensures that users are assigned to a single cluster and that this assignment is stable enough to prevent users which are potential candidates for multiple clusters from being reassigned rapidly, which may influence overall view assignment, display and continuity negatively.

Additional benefits of using the proposed will be discussed in the following sections, as they are not limited to view assignment. However, there is a drawback to using *D-Stream*. Since the algorithm is not only capable of detecting arbitrary shapes, but also an arbitrary number of clusters, the resulting group information will not be usable as is, as the number of groups directly correlates to the number of distinct views supported by the system. To mitigate this problem, a weighted *K-Means* algorithm is applied to the results of *D-Stream*, in order to receive the desired number of cluster K.

4.3 Viewport Display

Since the ability to display multiple distinct viewports is highly dependent on the available hardware, the employed method will have to adhere to the limitations imposed by it. The provided package will make sure to render as many viewports as indicated by the exposed configuration.

The environment used to develop this system consists of a stereoscopic projector capable of displaying at 120 Hz and providing image separation for each eye via active shutter glasses. Using the techniques detailed in [2, 9], it should theoretically be possible to serve two different groups with distinct renderings of the same scenes. However, this results in a refresh rate of 60 Hz per user or 30 Hz per eye, which is only barely acceptable, as it will most likely result in noticeable flickering of the image and reduced brightness per view [9].

To achieve the best possible results, a combination of active and passive stereoscopy is recommended to separate images for each group and the respective left and right eye images, as this results in less flicker and higher perceived brightness, albeit being more susceptible to crosstalk [9, 12].

4.4 Viewport Continuity

To make each user of the system as comfortable as possible, the position and rendering of each viewport should be as continuous as possible, as described in section 3.4. Since the described problems are mostly tied to the position of each user, the *D-Stream* algorithm in section 4.2 was not only chosen to solve the issue of group assignments, but also in order to combat the issues of viewport continuity.

Instead of simply clustering the current state of data, *D-Stream* uses a density decay mechanism to incorporate previous states of the data and therefore capture behavior of the different groups of users over time. It does so by processing the data in two steps, a fast online calculation phase which is applied to each incoming data record and a slower offline phase which updates the cluster data [3].

4.4.1 Online Phase

As a first step, the data space S is partitioned into p equally sized grids. This is done so that not every data record has to be stored individually, but is rather implicitly stored inside the discreet grid. This allows the algorithm to scale well with larger amounts of data, as each incoming data record will not have to be stored and processed separately.

Once a data record x arrives at time t_0, it is assigned a density coefficient $D(x, t_0) = 1$. This density coefficient will then decrease over time, as is dictated by the overall decay factor $\lambda \in [0, 1]$. The density D of a data record at time t can then be calculated as

$$D(x,t) = D(x,t_0) \cdot \lambda^{t-t_0} = \lambda^{t-t_0}. \tag{4.1}$$

Now, since not all data records should be stored, Chen et al [3] have proposed a way in which the combined density of a grid g can be calculated easily without storing the individual values for each data record $x \in g$. The density D at a time t after the time t_l which is the last time that particular grid has been updated, can easily be calculated as

$$D(g,t) = \sum_{x \in g} D(x,t) = \lambda^{t-t_l} \cdot D(g,t_l). \tag{4.2}$$

Once a data record x has been assigned to a grid g, and the density D of g has been updated, a grid will be assigned one of three possible states which will be used in the offline clustering phase. If the density D of a grid g at time t is

$$D(g,t) \geq \frac{C_m}{N \cdot (1-\lambda)}, \tag{4.3}$$

then it is deemed a *dense* grid, with the number of grids $N = \prod_{i=1}^{S} p$. C_m can be chosen in a range between $N > C_m > 1$ and controls the value at which a grid will become dense. In addition, if the density D is

$$D(g,t) \leq \frac{C_l}{N \cdot (1-\lambda)}, \tag{4.4}$$

then a grid is deemed *sparse*. C_l can be chosen in a range between $1 > C_l > 0$ and controls the value at which a grid will become sparse. Should the density D lie between

$$\frac{C_l}{N \cdot (1-\lambda)} \leq D(g,t) \leq \frac{C_m}{N \cdot (1-\lambda)}, \tag{4.5}$$

a grid is marked *transitional*.

With this simple method, the density of each grid is not only influenced by newer data records, but will also incorporate historical data, building an evolving database on which the next phase of the algorithm can be built.

4.4.2 Offline Phase

Rather than updating the clustering information every time a data record is received, the offline phase of *D-Stream* is only executed every time gap t_g. This is done to prevent excessive calculations if larger amounts of data records arrive in a small time window and allows the algorithm to scale well under heavy load.

Chen et al [3] describe the following steps which should be performed for every grid $g \in G$ at $t \mod t_g = 0$:

1. Update the density of g according to section 4.4.1.

2. If $t = t_g$, perform initial clustering by assigning all adjacent dense grids to a cluster and any transitional grids to the largest neighboring cluster.
3. If the state of g has changed since the last time gap, add it to a list of active grids A.
4. Check the state of each active grid $a \in A$ and perform the following operations:
   - If a is *sparse*, remove any assignment to a cluster c_a it may have had. If this severed the connection between other parts of c_a, split them into new clusters.
   - If a is *dense*, find a grid h which belongs to the largest cluster c_h adjacent to a. If a belongs to a smaller cluster c_a or no cluster at all, assign all $a \in c_a$ to cluster c_h. Otherwise, assign all grids $h \in c_h$ to c_a.
   - If a is *transitional*, find the largest neighboring cluster c_h and assign a to it.

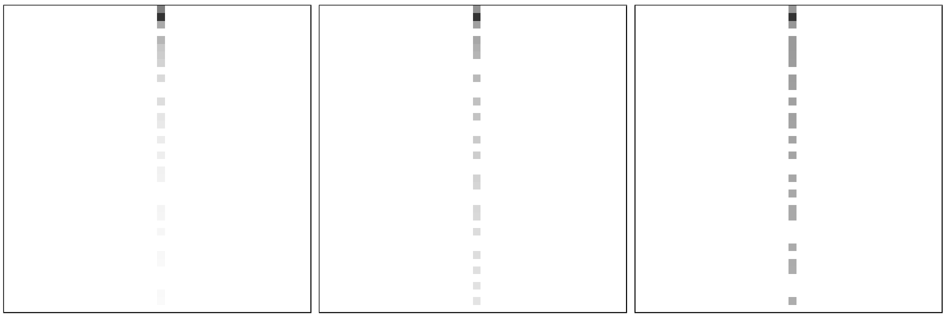
The largest clusters are identified by the combined density of each grid instead of their size. Since the density includes historical values, these simple steps provide clustering data for an arbitrary number of clusters and will evolve those according to new incoming data records, without disregarding the previous entries.

4.4.3 Post Processing

Considering that *D-Stream* provides an arbitrary number of clusters, some form of post processing is needed in order to combine the results to the desired number of k groups, as was already mentioned in section 4.2. One of the features of *D-Stream*, and density-based clustering in general, is the ability to deal with outliers and noise when forming clusters [3, 22]. This allows the resulting data to be combined using simpler algorithms without dealing with outliers. Since the density D of each cluster c is known, it is easy to apply a weighted *K-Means* algorithm, using D as a weighing factor to reach a desired amount of k clusters which can then be used as viewport positions for rendering.

4.4.4 Expected Behavior

D-Stream exposes a number of parameters which can be tweaked in order to mitigate the issues described in section 3.4. The most important factor influencing all three scenarios is the decay factor λ, as it directly controls the resilience of each cluster against change over time. Using values closer to 1 for λ will cause the density of data records to influence each cluster for a longer period of time, therefore more data is required to shift these clusters.



(a) $\lambda = 0.3$ **(b)** $\lambda = 0.6$ **(c)** $\lambda = 0.9$

Figure 4.1: A visualization of the density of all grids of the *D-Stream* algorithm after a user has crossed the tracking area from bottom to top in three seconds, with increasing values for the density decay rate λ. The data space $S = [-1.0, +1.0]$ has been partitioned into $p = 40$ grids. Higher density is indicated by darker pixels.

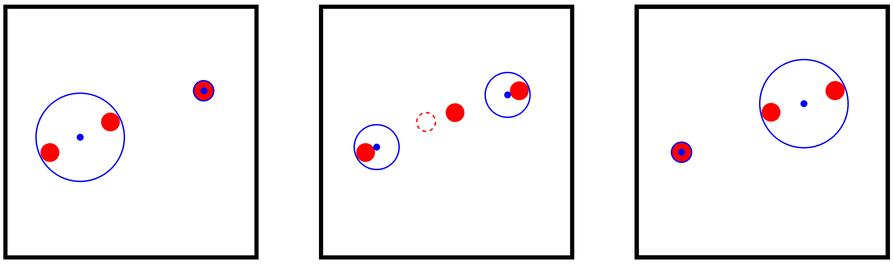
In contrast, using values closer to 0 for λ negates the stabilizing effect of a high λ value. The behavior of different values for λ on a single user moving through the tracked area can be seen in figure 4.1.

The first problem, described in section 3.4, is that of a user in cluster c_a being assigned to another cluster c_b while moving inside the tracking area. A simple grouping approach would lead to large changes in both groups viewports, as can be seen in figure 3.2. With the density decaying approach of *D-Stream*, the historical position information of a user does not simply vanish from the cluster c_a, but will remain influencing it until the associated grid is no longer dense enough to be included in clustering calculations. This causes the viewport of c_a to slowly move towards all active users of that group and the viewport of c_b to slowly expand to encompass the newly assigned user, as can be seen in figure 4.2. This behavior will not eliminate counter intuitive movements for users, who are not moving, but will help to mitigate the effects on them, as the change is not spontaneous, but rather continuously adapts to the new tracking information.

As with users being assigned to other view groups, the scenario of a user leaving the tracked space, and therefore the assigned group g, will have the same effect on group g as it had on the initial user group in figure 4.2. The cluster associated with g will slowly shrink to only encompass the still active users in that view group without rapid changes to the viewport.

The dense state threshold C_m and sparse state threshold C_l, in conjunction with appropriate values for λ can be used to control the third issue, fast moving users or groups.

Larger values for C_m and C_l cause a grid to only become dense once one or more users spend an increased amount of time in a position associated



(a) $t = 0$ **(b)** $t = 1$ **(c)** $t = 2$

Figure 4.2: Depiction of the two blue groups adapting to the position change and consequent group reassignment of the red user in the center of the black tracking area, when *D-Stream* is used as a clustering algorithm. Notice the gradual shrinking of the leaving group to the left of the tracking area and the incremental increase of the right, causing the viewport position, depicted as the blue dot inside each group, to slowly adapt to the current state.

with that particular grid. This has the effect of preventing fast moving users which will pass through multiple grid positions during their movement from having sufficient effect on a grid's density, as to have an impact on cluster calculations.

In detail, C_l controls the density, at which a grid g will switch from being sparse to transitional and vice versa. In terms of groups of users inside the tracked system, this will control the speed at which users will have to move away from the center of their current group in order to not enlarge the current cluster, but escape it and potentially be reassigned to another group.

C_m on the other hand controls the density at which a grid g will switch from being transitional to dense and vice versa. This controls the amount of time or users it takes to form a grid, which could become a new cluster and therefore a potential new viewport, depending on how many distinct views are supported by the system.

Depending on the use case, λ, C_m and C_l will have to be adapted to suit the desired movement patterns and expected number of users inside the system. There is no single optimal value for each given parameter, but it is possible to choose them in a way to optimize for certain movement patterns.

Slow moving, exploration-based use cases will want to set higher values for C_l and keep the gap between values for C_l and C_m relatively low. This allows users to slowly wander around the edges of their groups, without being separated from them, but if they desire to switch groups or form a new group on their own, they can easily escape their current group and quickly form a new point of view. A decay rate of $\lambda \geq 0.5$ will ensure a

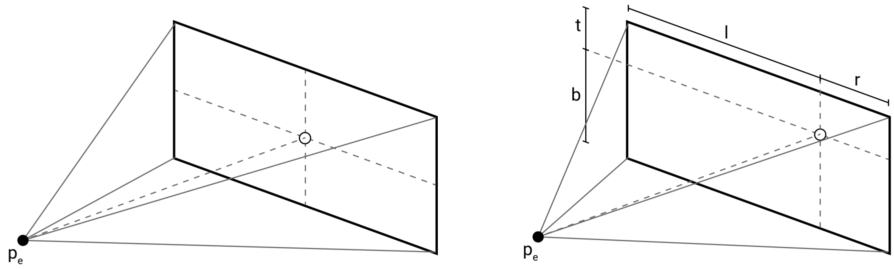


Figure 4.3: Visualization of the view frustum for on-axis and off-axis projection as described by Kooima et al in [13]. The on-axis projection on the left assumes the eye position p_e to be in the center of the image plane. The off-axis projection on the right however does not make that assumption, allowing p_e to be in an arbitrary position in front of the image plane.

more stable viewport, at the cost of a slower rate of adapting to the current positional information.

Use cases with fast moving users will want to set a lower decay rate of $\lambda \leq 0.3$ to make sure viewport updates will keep up with their users, at the cost of having a less stable viewport. Low values for C_m will make moving around without separating from a cluster easier, while high C_l values will encourage clusters to wander along transitional grids create by associated users, instead of the separation into new clusters.

4.5 Perspective Correction

To correctly render each viewport in accordance to each user's position, a well-established technique will be used. Off-axis projection, which is employed in *CAVE* environments, allows for view centric perspective when using direct view stereoscopy [4, 13].

Interactive 3-D graphics normally use so called on-axis projection. This projection assumes that the eye position p_e of a viewer is at the center of the image plane, therefore all edges of the image plane are the same distance from p_e and the resulting view frustum is symmetric. This would be a plausible assumption if the image is viewed on a computer monitor. However, since users of the system will walk in front of a stationary screen, the view frustum has to be modified in order to represent the relative change in position of p_e in relation to the images plane's center point, as depicted in figure 4.3.

To apply the correct perspective rendering, a projection matrix P has to be applied to the scene in order to transform the geometry in a way as to make objects that are farther away from the image plane appear smaller.

The standard 3-D perspective projection matrix is defined as

$$P = \begin{bmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & -\frac{f+n}{f-n} & -\frac{2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{4.6}$$

In equation 4.6, t, b, l and r represent the distance from the projected eye position p_e to the top, bottom, left and right edges of the image plane respectively, as seen in figure 4.3. n and f represent the distance of the near and far plane in relation to the camera position. Any geometry which is not between near and far plane will not be rendered in the final image.

For the on-axis projection, the values for t, b, l and r are easily calculated, as $t = b$, $l = r$, $t = h/2$ and $l = w/2$, with w and h being the width and height of the image plane. However, off-axis projection requires calculation of each value individually. Utilizing the properties of vectors, or to be more precise the dot product, the desired values can be calculated as

$$\begin{aligned} d &= -\left(v_n \cdot v_a\right), & l &= \left(v_r \cdot v_a\right) n/d, \\ r &= \left(v_r \cdot v_b\right) n/d, & b &= \left(v_u \cdot v_a\right) n/d, \\ t &= \left(v_u \cdot v_c\right) n/d, \end{aligned} \tag{4.7}$$

with v_n, v_r and v_u being vectors describing the local coordinate system of the image plane and v_a, v_b and v_c being vectors from p_e to three corners of the image plane, as seen in figure 4.4.

In addition, the scene has to be moved in accordance to the position of p_e. This can easily be achieved with a simple translation by the inverse of p_e, resulting in the translation matrix

$$T = \begin{bmatrix} 1 & 0 & 0 & -p_{ex} \\ 0 & 1 & 0 & -p_{ey} \\ 0 & 0 & 1 & -p_{ez} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.8}$$

If the display or projection screen is plain, and not rotated in any way, the previous steps suffice to achieve a perspective corrected image for the position p_e. However, should the screen be rotated, then an additional step of transforming the scene to the coordinate space of the target image plane is needed. Kooima achieves this in [13], by applying the inverse of what is

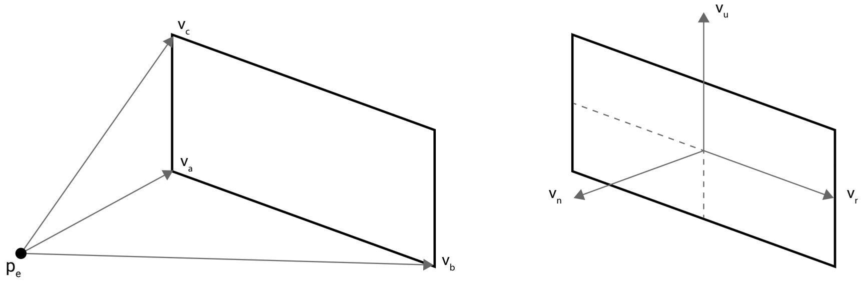


Figure 4.4: Screen corner vector v_a, v_b and v_c originating from the eye position p_e on the left and the local coordinate space of the image plane constructed by the three unit vectors v_n, v_r and v_u on the right, described by Kooima et al in [13].

normally a model-to-world matrix. Using the local coordinate system of the image plane, seen in figure 4.4, to construct this matrix results in

$$M^{-1} = \begin{bmatrix} v_{rx} & v_{ry} & v_{rz} & 0 \\ v_{ux} & v_{uy} & v_{uz} & 0 \\ v_{nx} & v_{ny} & v_{nz} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.9}$$

By combining P, T and M^{-1}, the complete off-axis projection matrix $P' = PM^{-1}T$ can be constructed and used instead of a standard on-axis projection.

4.6 Implementation

In order to test the proposed algorithms in a real world setting, a simple use case utilizing the described solutions was devised. Focusing on the ability to change the perspective view of a scene, the implementation provides an interactive environment to be explored by multiple people. While navigating the environment, a simple story is told by manipulating objects outside the view of all users and replacing them with certain elements which create a scene from the story. The fairy tale *Hansel and Gretel* was used to create those scenes, as its adaptation by the *Brothers Grimm* is well known and suitable for a wider audience.

To track each user inside the system, the *PHARUS* tracking system was used. An easy-to-use implementation is provided as a *Unity* package which was designed to allowed easy access to tracking data provided by laser rangers [17, p. 20].

The *D-Stream* algorithm described in section 4.4 was implemented inside the *Unity* engine using the C# programming language. It is highly customiz-

```
{
  "system": {
    "groupCount": 1,
    "partitionsPerDimension": 40,
    "clusterCalculationStep": 0.1,
    "decayFactor": 0.4,
    "denseThreshold": 3,
    "sparseThreshold": 0.8,
    "displayVisualization": false
  },
  "trackArea": {
    "width": 6.5,
    "depth": 6.5
  },
  "display": {
    "width": 4,
    "height": 2.5,
    "elevation": 0.5,
    "trackAreaDistance": 0.5
  },
  "camera": {
    "vrMode": false,
    "vrHeight": 1.72,
    "nearPlaneSync": true,
    "speed": 2.5,
    "manualStereoscopy": true,
    "topBottomMode": false,
    "alternatingFrameMode": true,
    "interPupilarDistanceM": 0.0635
  }
}
```

Figure 4.5: `config.json` file exposing various parameters for an application using the provided *Unity* package.

able by exposing each parameter of the algorithm, allowing it to be tailored to the expected use case; see section 4.4.4 for more details. In addition to allowing the parameters to be changed inside the *Unity* editor, the entire configuration is also exposed via a `config.json` file which is read during startup of any application using the provided implementation. In addition to configuration of the *D-Stream* algorithm, a few environment specific configurations, for example the dimensions of the tracking area and projection space, are also exposed in this file, see figure 4.5.

The environment for the story was built using various free assets from the *Unity* asset store and enhanced with various sound clips published under the *Creative Commons 0*[1] license taken from www.freesound.org. As the tale

[1] See https://creativecommons.org/publicdomain/zero/1.0/ for details of the *Creative Commons 0* license.

Figure 4.6: Screenshots of the interactive woods environment built for the *Hansel and Gretel* story.

of *Hansel and Gretel* mostly takes place in a forest, a stylized version of such an environment was created and serves as the backdrop of each scene, as can be seen in figure 4.6.

Players need to navigate the tracking space in order to uncover different elements of the story. While they move around, the time of day changes from dawn to dusk. At certain times, new story elements will appear, for example the witch's hut. Players need to find those elements in the scene and keep them in the view for a certain amount of time to progress the story, otherwise the time of day will not progress until they have done so. To spawn the next story element after the current one was found, players need to change their view of the scene,in order to lose sight of the current element. This causes the current object to disappear and the next one to appear somewhere outside the line of sight of the players, giving the impression that the forest changes magically while they are not looking.

Should players not move away from the current object after sufficient time, a ball of light will appear to grab their attention and guide them away so the next scene can be created. Furthermore, if players are not able to find the next element after it has appeared in the scene, some positional audio cues will give further aid to progress the story.

Testing was performed inside an installation provided by the *Playful Interactive Environments* research group. Figure 4.7 shows a development session in that exact space. This space was also used for evaluating the devised system and will be described in more detail in section 5.2.3.

In order to allow other use cases to be built around the system, a *Unity*

Figure 4.7: Photographs taken during a development testing session depicting an early design for the scene created for the *Hansel and Gretel* use case.

package is provided with all components needed to connect to any laser ranging system supported by the *PHARUS* tracking system. It provides visual representations of the tracking and display area and can easily be placed inside any *Unity* scene. It will then place virtual cameras at calculated group centers and take care of rendering a perspective correct image for each placed camera.

4.7 Result

Using the proposed solution and the applied use case in this section with recorded test data yields promising results. However, without further testing the results still remain theoretical and need to be evaluated in a more structured manner in order to determine their validity.

Chapter 5

System Evaluation

In order to test and confirm the expected behavior of the proposed solution in chapter 4, some form of evaluation is required. Since the main focus of the system is to provide an acceptable stereoscopic experience for multiple users of any system, less focus will be given to the application itself compared to the systems performance in handling multiple users. For this reason, an expert evaluation was chosen as a focused evaluation method specifically designed to detect usability issues, as evident by positive results acquired by Desurvire et al [6, p. 1510], Korhonen et al [14, p. 14] and Pinelle et al [19, p. 1455].

5.1 Expert Evaluation

Often also called heuristic evaluation, expert evaluation is a method developed to find problems as early as during the prototyping stages of an application. Instead of using thousands of rules for evaluating a certain application or scenario, a smaller subset of heuristics derived from a larger pool is used for evaluation; hence, expert evaluation is described as a discount usability method. Even though a small set of heuristics is used, it is already possible to find many usability problems of a system, even when using non experts [18, p. 20].

Even though Nielsen originally described expert evaluation as a tool to assess the usability of classic desktop applications, his methods have been successfully applied to games in the past, with Pinelle et al [19, p. 1460] calling heuristics " [. . .] well suited to uncovering important usability problems in the game context.". Desurvire et al [6, 7] and Korhonen et al [14] have also formulated heuristics specifically designed to examine usability in games and tested their work with positive results.

Even though previous work is a testament to the usefulness of expert evaluations, some drawbacks still occur. Due to the evaluation being based on a small set of heuristics, it is possible to miss problems that are not covered by one of the heuristics, encouraging evaluators to focus on the issues described by the prepared selection [14, p. 12].

5.2 Setup

To perform an expert evaluation, three things have to be considered before the actual evaluation can take place. First, heuristics have to be defined which form the basis of the whole evaluation procedure. Secondly, a number of experts has to be chosen to perform the evaluation with. Finally, the different steps during the procedure need to be defined and communicated to all participants of the evaluation.

5.2.1 Heuristics

To assess the usability of an application, appropriate heuristics have to be found which should cover as many elements and potential problems of the system as possible. To accomplish this, existing heuristics should be used as a basis to form heuristics more suited for the specific system which should be evaluated [18, p. 19].

As the application described in this thesis is an interactive one and the use case outlined in section 4.6 is a story-based game, heuristics defined for usability assessment in games were used as a basis to form eight new heuristics. *Heuristic Evaluation for Playability* and *Principles of Game Playability* by Desurvire et al [6, 7], as well as the heuristics defined by Pinelle et al in [19] were chosen due to them already being tested and used with promising results in various usability investigation.

The following eight heuristics were used in the evaluation, with the first six aiming to test the underlying system introduced in chapter 4 and the last two focusing on the interactive story use case described in section 4.6:

1. Players feel that their movement is intuitive and consistently mapped to the system

Any movement in the tracking space should translate into consistent and believable navigation of the interactive system. Users should not require additional information about how movement is mapped into the system, examining the mapping effort put forth in section 4.1.

2. Players experience the visual perspective provided by the system as believable

While users navigate the virtual space, their viewport of said space should match their position in front of the target screen, as if they would look through a window. This correlates to section 4.5 and the off-axis projection described there.

3. The system's reaction to multiple players moving inside of it is natural enough to not need any instructions

Coping with individual users moving around is the third scenario described in section 3.4 and as such, the system should handle such movement gracefully as to not disturb the natural movement of each user or imposed certain navigation rules on them.

4. The system provides a pleasant, consistent viewing experience

This heuristic is provided as a catch all for any underlying problems with viewport continuity from section 3.4 which are not specific to any of the three defined problem scenarios. It is included to detect any issues that were not expected, as viewport continuity is a central property of the system and any problems in that regard need to be detected.

5. Players can enter or leave the system at any time without breaking the experience for users already inside the systems

As users will sometimes, willing or unwillingly, exit and re-enter the tracking space, as described by the second scenario in section 3.4, this heuristic is formed in order to ensure issues regarding this feature of the system are caught.

6. Players feel that they can navigate the system, even when multiple users are present

Users should be able to navigate around the system own their own, allowing them to influence the current state even with multiple users inside the system. In addition, they should also be able to switch to another viewing group without problems, as described by the first scenario in section 3.4 and 3.2.

7. Players are able to understand and follow the story without additional information

In order to motivate players to explore the application and move around, they should be able to grasp the story outlined by the elements in the forest scene, as described in section 4.6. As no additional material is given, the combination of audio and visual elements need to transport the story on their own.

8. Players are able to progress the story, without additional instructions

Since the story might only become clear after multiple scenes have been seen, it is important that the users understand how to progress the story, without needing additional instruction. A central part of this are the audio cues and the guiding light described in section 4.6.

5.2.2 Experts

Even if it is already possible to acquire satisfying results with normal users [18, p. 20], experts bring forth deeper understanding and knowledge and are able to apply the proposed heuristic to greater results. The number of experts suggested by Nielsen in [18] are three to five users, which is further supported by positive results from Pinelle in [19, p. 1459]. Due to the limited space in the available test environment, which will be detailed in section 5.2.4, the number of evaluators was chosen to be three.

These experts were all chosen amongst members of the *Playful Interactive Environments* (*PIE*)[1] research group, which deals with a diverse set of topics concerning interaction in co-located spaces. Their experience was drawn upon to perform the evaluation procedure detailed in section 5.2.4. They will be introduced in the order in which they appear in the evaluation transcript, as can be seen in appendix A.

The first evaluator was Wolfgang Hochleitner, providing previous experience of a heuristic evaluation approach to determining the usability of an interactive system. With a background in human-computer interaction, Michael Lankes was also part of the evaluation. Having had first hand project experience in creating interactive, co-located systems, Georgi Y. Kostov was also part of the procedure and concludes the trio of experts.

5.2.3 Location

Due to the nature of the developed system, another part of the evaluation procedure was to find a suitable location with sufficient space and the required hardware installation needed to be found. Thankfully, the *PIE* research group was kind enough to provide such a space and made it available for development, testing and evaluation during the duration in which this thesis was created. The space is depicted in figure 5.1 and measures roughly 6 by 5 metres in size, with a tracked area of 4.5 by 3 meters and featuring a 4.5 by 2.5 meters projected image when using an aspect ratio of 16 : 9 for the projection. Four laser rangers are mounted on the walls of the tracking space to ensure satisfying tracking coverage when multiple users are using the system, as seen in figure 5.1.

[1] For more information, visit http://pie.fh-hagenberg.at

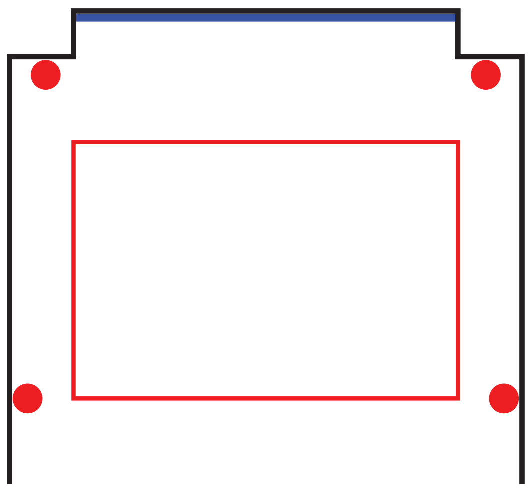


Figure 5.1: Schematics of the location provided by the *PIE* research group, with four laser rangers and the tracking space depicted in red and the projection screen in blue.

Another potential candidate for a suitable evaluation environment would have been the *Ars Electronica Center*'s *Deep Space*[2] located in Linz, Austria, as it provides a similar setup in terms of tracking technology and visual display, albeit a bit more ambitious. With its dimensions of 16 by 9 meters for both the tracking space and the projected image, it would have provided additional insight as to how the proposed system scales with increased tracking space and a potentially larger group of users. However, due to time constraints and the *Ars Electronica* festival taking place in close temporal proximity to the date of the planned evaluation, this was not possible.

5.2.4 Procedure

After the experts needed for the evaluation are found, they need to be introduced to the evaluation procedure, which Nielsen [18] splits into four main parts.

Kick-off Meeting

At the beginning of the evaluation each expert is introduced to the system and its current state. They should receive an overview of the expected capabilities of the system and any parts of it that might still be missing. In addition to the system introduction, the heuristics used as a basis for the

[2] For more information about the *Ars Electronica* center and *Deep Space*, visit http://www.aec.at/center/.

evaluation are reviewed and sufficient background information is given to understand them, if needed.

Expert Evaluation

Using the information given during the kick-off phase and their experience and wisdom in their respective fields, each expert is given sufficient time to experience and evaluate the system's issues, if any are found. Handouts describing the defined heuristics and rating scales can be provided to keep the evaluation procedure focused. In this step, the director of the evaluation will not partake or influence the experts, as they are encouraged to work alone and, if possible, independent from one another.

Review Session

After each expert has concluded their evaluation, their findings are brought together to describe each heuristic violation that has been found during the evaluation. These violations then receive a rating in severity, frequency of occurrence and presumable ease of solution using a scale for each attribute, which will be explained in more detail in section 5.3.1.

Reporting

The findings of the review session are used to compile a list of violations and their agreed upon ratings in the three properties described in the previous section. This issue compilation is ordered by an overall score which will be detailed in section 5.3.1, providing a suggested order in which to solve each issue.

5.3 Report

After the results from the evaluation session have been compiled into a list of unique violations of heuristics and rated, further processing is required to form the final, ordered list of issues and suggested solutions.

5.3.1 Data and processing

As already mentioned in section 5.2.4, each found violation is rated by its *severity*, *frequency of occurrence* and presumable *ease of solutions*, in order to allow further processing and ordering of the results.

Severity

To describe the weight that a particular issues has on usability, a severity rating is assigned to it, taking into account the following scale:

0: A heuristic has been violated, but no implications on the overall usability arise from the issue.
1: The problem is purely aesthetic and only serves as a distraction, not a real usability issue.
2: The violation causes a small impact on usability.
3: Usability is highly degraded by the issue, but the system is still somewhat usable.
4: The system is rendered unusable by the heuristic violation.

Frequency of Occurrence

To estimated how often an issue will occur during normal usage by end users, the following scale is used to assign a frequency of occurrence:

1: The issue occurs rarely or is hard to reproduce.
2: Problems only happen sometimes during normal use or is easily reproducible.
2: The violation happens often during normal usage.
4: The system is permanently affected by the issue.

Ease of Solution

To not only judge an issue by its impact on the usability, but also on its implications on the architecture and programming of the system, the following values are possible for ease of solution:

1: Trivial code or configuration changes are able to solve the issue.
2: The issue is resolvable by a small modification to one specific element.
3: A redesign of a small part of the system is needed to resolve the violation.
4: Fundamental modifications are necessary in order to overcome the issue.

Overall Score

In order to provide a prioritized list in which to work through the discovered violations, an overall score o which combines the three rating values for each issue is calculated as $o = s \cdot (f + e)$, with s being the severity of the issue, f the frequency of occurrence and e the ease of solution. This approach, described by Friedl in [8, p. 56], yields a suitable ordering of the found results, as high impact issues will be prioritized over infrequent or insubstantial ones. However, the resulting order deviates slightly from what Friedl describes as expected ordering, as he assigned higher numbers to more easily solvable tasks, therefore favoring the lower hanging fruits, as he calls them. The chosen scale for ease of solution in this paper will instead shift the focus on solving the higher impact issues first.

5.3.2 Results

Based on the findings and discussion during the review process, which were transcribed and are available in appendix A, seven heuristic violations have been found, rated and were compiled into an ordered list based on their overall score, as described in section 5.3.1.

Score 10: Navigation feels counter intuitive at the beginning

Violated Heuristic: 1, Severity: 2, Frequency of Occurrence: 2, Ease of Solution: 3
One of the most severe issues found during the evaluation was learning how navigation works in the provided system. Without proper guidance users may be confused as to how their movement influences the current state. Depending on the use case that is built on top of the system, players could use some time alone with the system to learn its behavior or need some form of introduction before entering the tracked space. However, once navigation is understood, it works well.

Score 10: Navigation with multiple users needs coordination to be understood

Violated Heuristic: 3, Severity: 2, Frequency of Occurrence: 2, Ease of Solution: 3
Deeply connected with the aforementioned violation, navigation with multiple users is also not understood immediately if the systems inner workings are not known to each user. This could be remedied by giving a brief overview of the system or, for a more elegant solution, the presented use case could be adapted to encourage communication and teamwork between users.

Score 9: Users sometimes experiences rapid changes in perspective when users enter

Violated Heuristic: 5, Severity: 3, Frequency of Occurrence: 2, Ease of Solution: 1
Experts also observed brief rapid changes in the rendered perspective which caused great discomfort, causing them to assign a high severity rating to this issue, even though it was only infrequently observed. It was agreed upon that simply applying more smoothing to the cameras movement or adapting the parameters used for the *D-Stream* algorithm, as described in section 4.4.4, should suffice as a solution.

Score 7: Players do not understand that a story is being told

Violated Heuristic: 7, Severity: 1, Frequency of Occurrence: 4,
Ease of Solution: 3
Related to the *Hansel and Gretel* use case, the fourth highest violation targets the storytelling experience. The elements used to build each key scene of the story were not clear enough to draw focus and therefore went unnoticed. A proposed solution was to increase the number of elements per scene or add more moving or animated objects to catch the attention of users.

Score 6: User navigation is slightly impaired with larger user groups

Violated Heuristic: 6, Severity: 1, Frequency of Occurrence: 4,
Ease of Solution: 2
Even though this issue was only extrapolated from the current setup, as the number of users was rather small, discussion showed that each expert agreed that the navigation might be impaired when larger groups of users are present inside the system. Depending on how big the available space is for the installation and how many users are present, a single user might no longer be able to influence anything, as his tracking information is overshadowed by the rest of the group. This could be prevented by encouraging group movement or limiting the maximum number of users inside the system.

Score 3: The story guide was not always visible

Violated Heuristic: 8, Severity: 1, Frequency of Occurrence: 2,
Ease of Solution: 1
The guiding light was a tool that greatly helped with advancing the story, should users be confused as to what actions will move the narrative forward. However, it only appearing after some time, when the system thinks users are stuck, brought about some confusion. Users might not be sure if the guide was just hidden behind an object and out of view or had actually vanished. Simply having the light stay on screen as a permanent guide would solve the issue.

Score 0: Users moving apart from larger groups in the system experience degraded perspective

Violated Heuristic: 2, Severity: 0, Frequency of Occurrence: 4,
Ease of Solution: 4
The only issue rated with a severity of zero was the user experience of a singular user who is apart from another group of users inside the system. Depending on how far away from the group the singular user is positioned, the rendered perspective would no longer line up with what could still be

considered believable. This issue is unsolvable given the physical constraints that such a stereoscopic system needs to adhere to. However, each expert agreed that this behavior is expected and will be understood by users of such a system, therefore receiving a zero rating.

5.4 Outcome

Looking at the violations found, it is interesting to note that from the heuristics defined in section 5.2.1, the second one, *Players experience the visual perspective provided by the system as believable*, and the fourth one, *The system provides a pleasant, consistent viewing experience*, seem to be the only ones fulfilled without significant compromises. This suggests that, although one severe issue regarding the viewing experience was found, the system provides a pleasant visual experience overall. In addition to the positive results of the non-violations, the issues that have been found serve as a basis for improvements to the system which will be discussed further in section 6.2.

Chapter 6

Conclusion

This thesis and the accompanying project aimed to propose a viable approach that allows multiple users to navigate a stereoscopic system in a way that provides a pleasant and consistent viewing experience. By drawing on existing research that was introduced in chapter 2 and performing initial test runs with the installation provided by the *PIE* research group, a set of problems was defined that needed to be solved in order to implement the desired system.

Building on the existing attempts on providing stereoscopic content to a smaller, known number of users, a plan of action was devised and implemented to cope with the increased complexity that is introduced by allowing any number of users inside a stereoscopic system. This included processing the positional information for each user and mapping it to a virtual representation of the tracking space, displaying content in a perspective-corrected manner according to the positional information and ascertaining that the user experience is still sufficient as to not cause discomfort or sickness for each user.

In order to showcase the designed system, a small use case was then devised which was based on navigating an interactive environment in stereoscopic 3-D. The fairy tale of *Hansel and Gretel* served as a backdrop to create an interesting scene that can be explored by utilizing the implemented techniques described in chapter 4.

The concluding step was to assess the viability of the devised system by performing an expert evaluation. By inviting members of the *PIE* research group, introducing them to the system and its function and having them scrutinize the system based on a set of eight heuristics devised to test each desired aspect, shortcomings of the systems were able to be identified. The results of the evaluation were then used to propose future enhancements for the system, which can be found in section 6.1.

6.1 Evaluation Retrospective

By performing the expert evaluation detailed in chapter 5, important issues with the real world usability of the proposed system were able to be observed. From the eight heuristics that were defined in section 5.2.1, six were violated to varying degrees, with the most prevalent issues affecting navigation inside the system. Reactions of the virtual viewport to the movement of each user were not immediately clear during the evaluation, requiring some time and coordination to understand and get used to. These issues were however highly dependent on the use case, as each expert agreed that in any exploration-based context, as was the case in the example use case, the coordination and additional effort enhanced the desired experience. Further research is needed in these regards, but as of now navigational issues seem to be better handled in the context of an actual use case built on top of the proposed system.

An issue which warrants further improvement of the systems concerns rapid changes in perspective which were sometimes experienced by the evaluators. Such a change can cause potential discomfort or even sickness and should be avoided at all cost, as was already detailed in section 3.4. Discussion about the issue with each expert already yielded a few possible remedies, from increased smoothing of viewport movement to adapting input parameters of the applied *D-Stream* algorithm. The problem may also arise while the clustering results from *D-Stream* are combined into the final data used for positioning the rendered viewport, as detailed in section 4.4.3. Additional testing effort will be needed to confirm or dismiss these recommendations and any successful solution should be incorporated into the current system.

As the interactive story experience of *Hansel and Gretel* was also subject to scrutiny, issues with this part of the application were also pointed out by the evaluation. The most crucial one has to be the inability to comprehend the story without any additional information, rendering the storytelling itself nearly useless and in need of fundamental changes. However, the navigational cues and systems established worked very well, allowing the experts to advance the story and encouraging them to do so, even though they did not realize the effects their actions had, as the story was not delivered in an accessible enough manner.

In addition to uncovering issues with the system, one of the two heuristics dealing with the comfort and visual aspect of the system was missing from the list of violations. The other one, while still being violated, was assigned a severity of zero, indicating that the violation can be categorized as a non-issue. This hints at a successful first attempt in delivering interactive stereoscopic content to groups users.

6.2 Additional Enhancements

Apart from the improvements that could be made based on the feedback extracted from the evaluation, additional enhancements to the system were planned, but due to time constraints and the limits of the available hardware, never implemented. As already mentioned in section 4.2, it is possible to assign different views to each detected group of users. This step is already handled within the system, but due to the projection system used, only one distinct view could be displayed. Due to this limitation, no concrete functionality of handling the display of each distinct viewport in sync with any projection system has been implemented, apart from creating and updating multiple viewport renderings inside the *Unity* engine. If a sufficiently sophisticated projection system became available, functionality to actually display the already rendered viewport could be added. This includes communication with the projection system to allow users to switch groups dynamically and have their visible viewport migrate to new group assignments.

To complement the visual experience, additional interface options could be made available by the system, so group assignment and position are available to third party libraries, allowing for positional audio to be generated for each group and further enhancing the interactive experience. This would also greatly benefit the *Hansel and Gretel* use case described in section 4.6, as sound cues were given to help advance the story.

To further optimize the experience and prevent sudden changes to viewport positions, movement pattern recognition and prediction similar to what Naderer describes in [17] could be employed. This would allow for better accuracy when calculating viewport positions and could also potentially allow each viewport to anticipate movement and shift its position accordingly, reducing the latency that is inherent with only being able to process positional data as it arrives.

6.3 Outlook

The system devised in this thesis tries to tackle many issues that emerge when dealing with multiple users in a stereoscopic environment. Even though some issues were found by performing a heuristic evaluation, it also indicates that the groundwork for providing a pleasant experience for each user inside such a system was successfully formed. The information contained in this thesis and the provided implementation should therefore be able to serve as the basis for further improvements and research into the area of interactive stereoscopy for multiple co-located users.

Appendix A

Evaluation Review Session Transcript

The transcript was written in English, to make it more widely accessible. The actual conversation was held in German.

I – Interviewer
W – Wolfgang Hochleitner, MSc
M – DI (FH) Dr. Michael Lankes,
G – Georgi Yordanov Kostov, MSc

Introduction for experts at the beginning.
I: Wolfgang, why don't we start with you? But of course, everyone is welcome to interrupt at any time and add information as well.
W: Okay. About the controls, because I am looking at the first heuristic, "Players feel that their movement is intuitive and consistently mapped", and I do not feel it has been violated. But at the beginning I was not sure how the controls work, meaning at the beginning it was not intuitive for me, until I realized what it meant when I walked forward, to the right or backwards. It just took some time for me to realize. I don't know if the first heuristic is violated because of this, at the beginning yes, but later on everything was clear. Also by walking in a group everything was kind of made easier, because one could coordinate with the others, but it also made things kind of harder, as I did not know exactly what my own movement affected. As the system tracks the whole group, if I move to the right and the rest of the group moves to the left the system will react in a way that might even be considered counter intuitive. So I would see the first heuristic as violated but not very severely. I would say it causes a small impact, so severity 2. About the frequency of occurrence, I would say easily reproducible, so 2 …
Stops mid-sentence and starts again.
W: … but actually it is hard to tell, as this only happens in the beginning

and then, once the system is understood, it doesn't happen again. At the beginning I was not entirely sure what actually happened, how I am able to move and how this works, so I would still say 2. But still, it does not really happen during the entire experience, just at the beginning. Ease of solution. That is a hard one.
I: Would you think an introduction at the beginning would suffice? Do you think this is more a problem of the system, or the use case?
W: Depends on what one would like to achieve. If it is part of exploring the system at the use case level, then it would be part of the experience, if that was the intention. Similar to the game Journey, where you are simply put inside the world and have to figure out everything on your own, how everything works. If that is the intention, the system works fine. But if users should be able to quickly use the system then some form of instruction would be nice. It does not need to be text, but could be communicated by the experience in some way. That is why it is hard to set an ease of solution, it could range from one to four, depending on how the system should be experienced. If experimenting is part of the use case, nothing needs to be changed. If the aim is to have users grasp the system as quickly as possible, how the controls work, then big changes would be needed, at least that is how I see it. But that highly depends on the focus, how the application is intended to feel. I was a bit confused at the beginning, about what happens and what I needed to do, and I instinctively observed the other users' behavior, if they seemed to have grasped how the system works. That were my thoughts on issues with the first heuristic.
The other experts are expressing their approval via nodding.
I: How do the others feel about this issue?
M: Should we walk through the issues by checking each heuristic? That would give a clearer structure.
I: Sure.
M: Okay. What irritated me was inverted movement, meaning if I walked backwards I actually moved forwards and if I walked forwards I moved back. So I do not think it actually felt intuitive, that was not clear for me. I got a hold of it in the end, but it was not clear for me. The mapping made sense for me, but it did not feel natural throughout the whole experience. So for me it would be a severity of 2 and as it happened all the time ...
*Cut off mid-sentence by **W**.*
W: Well, there are two paradigms opposed here, on one hand walking forwards to move forward, and on the other to walk backwards to get a better overview, so if you move backwards it kind of zooms out, increasing the field of view, which makes sense in a way, because if I walk backwards, I will see more of the scene. On the other hand, the urge to walk forward ...
*Cut off mid-sentence by **M**.*
M: Well, for me it is not so much the walking backwards and seeing more of the scene aspect, but if I walk towards the screen, that I move away from

it, does not make sense for me.
I: Might that be an issue with the second heuristics, as in maybe the perspective was not correct?
M: No, not really. It really is about the controls not feeling intuitive for me, as I realize my expectation of movement in the real world do not map to the system. It is not about the visual aspect, that is why I would say it is a mapping problem.
W: Yes, but that might just more suit the heuristic "Players feel that they can navigate the system" ...
Stops mid-sentence and starts again.
W: This might be a violation of more than one heuristic.
M: Maybe, but I was specifically talking about the intuitive aspect of the controls. As I understand the implementation, the field of view of the camera is manipulated for this effect, right?
I: Not exactly, but the resulting effect my look similar.
Short refresh on how the camera system is implemented.
G: It is actually easier to experience the system alone for a short time, as it gets a little harder to realize how exactly the camera system is influenced by each person, when not alone. The problem is, if you move away from the central perspective, the default, by moving left or right you will not experience view changes as fast with other people. But if you are standing at the back and move alone, the camera moves faster and it is easier to understand how you influence the camera. But if you move left or right next to a group, the camera moves a bit too slow to understand, that you are actually influencing the camera.
W: Of course, but it is expected that a single person moving away from the rest of the group will have a worse experience, there is no other way to handle this. So yes, heuristic two is violated, but I do not see a way how this could be fixed. The camera has to be positioned somewhere and it makes sense to place the camera to larger groups of users, to provide more users with a better perspective. This is more a user's problem than a system issue.
M: But even so it always looked three-dimensional and believable.
W: Yes, I would only note this as a remark.
I: So heuristic 2 was violated, but it is a non-issue?
W, G, **M**: Yes.
W: Sure it is a bit of an issue, but I think this is the preferred way. The other way would be to degrade the experience for all three users, but that is not how I experienced it. Of course, if you are not near the biggest group, your experience will be compromised. But I think this is preferred.
G: Of course, if you are not aware of this it still might disrupt you, as you are dependent on the other users, so it is very subjective.
***W** reads the title of the third heuristics, "The system's reaction to multiple players moving inside of it is natural enough to not need any instructions".*
W: The same applies as for heuristic one. In the beginning it was hard to

understand how movement is mapped, but by coordination within the group and some users leaving the space, it became clear very fast. As said, if the users move in a group the reaction of the system is clear, but if they move on their own, the system will try to react and depending on how each user is positioned, the resulting movement might not even be noticeable in some circumstances. If one user is walking to the right and the other to the left, what is supposed to change? The behavior is okay, but depending on the focus some instruction might be necessary. But I think the reactions will be simply understood.
M: I actually think it works better without instructions, as the group coordination works in favor of the system. Instruction might actually, how should I say, feel out of place and force users into a certain behavioral pattern. To be honest, the group coordination was more interesting and efficient than following instructions.
W: This of course depends on how the system is used.
M: Of course, as we understood and experienced the system, instruction would destroy the experience.
W: Yes, as said before, I am reminded of Journey, as you are dropped into the world and have no idea what is going on and just have to experiment.
Each expert hums approvingly.
I: So by communication with each other and moving around, the system's behavior was clear?
M: The interaction nearly becomes a game on its own, which is really good. The camera experience and coordinating works well.
I: So, how severe would you say is the third heuristic violated at the beginning?
M: I do not see a problem at all. At least in my opinion.
I: Do you see no problem in the system context alone, or within the whole experience?
M: Just looking at the system context there is a problem, yes.
W: I also have to say regarding the story, it really did not matter too much. To be honest, if you asked me what the story was, I would be hard pressed to tell it now.
M: Yes, if you would not have mentioned the story and just let us play, I would not have guessed there was a story and that the elements in the back were of any importance.
W: For me the light thing was a kind of goal, to chase it in order to not have the experience be random, otherwise one would probably be confused about what to do after a minute, so it was important. But if there was an overarching story, apart from the figure, the house and the cage, I would not be able to tell.
G: Well, we did not experience the story until the end.
W: That is true, but it was still too vague and I did not really pay attention to it, to be honest.

M: For me these elements were simply part of the scenery, that they had no special meaning and were just there for decoration.
W: I would have liked to actually go there, I felt a strong urge to do so, but was not able to.
M: To more closely inspect elements yes, but not to understand any kind of story.
***W** reads the title of the seventh heuristic, "Players are able to understand and follow the story without additional information".*
W: This heuristic is clearly violated. I would not even have guessed that there was a story without knowing so. It does not really hinder the usability, as I did not really care about the story, therefore I think it is more of a minor, esthetical problem, that I did not understand the problem. Frequency of occurrence was permanent, so severity 1, frequency 4 because it is permanent. For ease of solution, I would have to rate it a 3, as there is some work to be done in order to communicate the story in a clearer way. Maybe one has to work with voice-overs or just make it more clear, what actually happens, but it should also be considered that changes might not even be desired. If the story should be fully understood, then something has to be rebuilt. If the story should be open to interpretation, then the story experience is fine as is.
I: So do you thinks simply giving the information that a story is told is sufficient, or also an additional indicator about the genre of the story, in this case a fairy tale? So users would know to keep a look out?
M: Well, if people need to be told that there is a story, then this in itself is a problem. A story should speak for itself, otherwise it will not work.
W: If I think back to the experience, a fairy tale makes sense and I would have guessed that. But again, the story did not seem to be the focus of the experience and was overshadowed by the control aspect, which is not a critique in any way. But if the story should be the focus, it did not work.
M: This may also be a problem with the aesthetics, as all the objects were placed rather far away from the camera. There was nothing moving to the foreground, indicating it is of importance for any kind of interaction. This is why we tried to navigate towards each object all the time.
G: Maybe a narrator would be good? Every time a new object of the story appears, a small portion of the story could be read by a narrator, so each user is able to understand, what the current scene is about. Just the visuals and some sound was a bit too little information.
W: But again, it is important to know if this is wanted, if it should be cleared up that this was, for example, Hansel and Gretel or whatever, then it could be read after each event, but that is the forceful approach.
I: It actually was Hansel and Gretel.
M: If this is about Hansel and Gretel, one could use breadcrumbs and similar visual objects to give hints to the users and guide them to the objects.
W: Maybe one needs to work with more than one objects instead of just

having that house. Adding more objects from the fairy tale to the scene would probably clear things up.
G: What was very important was the light, because it lead the audience from one object to the next. But what I noticed is that the light only appeared after two or three minutes and it was not always there, so it appeared for a bit, disappeared and then something happened. Maybe it would be good if the light is always there and always leads the users.
***W** reads the title of the eighth heuristic, "Players are able to progress the story, without additional instructions."*
W: Well as soon as the light is introduced as a guide that I have to follow, then no instructions are needed as we all figured that out pretty fast. If that light, or the wisp or whatever, appears you want to follow it, I think this behavior is automatic, and this really works well. It was a bit confusing that it actually disappeared sometimes, so I thought it may just be hidden by a tree and my view was not right or it might have disappeared. This would be the only issue that I would see that violates the last heuristic, but only very slightly.
I: So a cosmetic issue?
W: Yes, exactly. I was often unsure if it actually disappeared and if it would actually appear again, especially after the first time that was not clear. Maybe if it would just stay on screen somewhere, this would help, but as I said it is only an aesthetic issue. So, what are we still missing?
***W** reads the title of the fifth heuristic, "Players can enter or leave the system at any time without breaking the experience".*
W: Yes, sometimes I felt that if someone was entering that the camera would suddenly move a lot. It did not happen all the time, which seems interesting. I could not entirely say why this happened, you (**I**) would probably know better, but sometimes one would perceive a boost, or a jump, were the field of view, or whatever is actually adapted, changed a whole lot and this was disruptive. So for me, I would rate this as a three, heavily degraded, because this really takes you out of the experience, if there is such a shift in perspective. Frequency, that is hard. I would say sometimes, but I am not sure if it is easily reproducible, as I did not fully understand why it happened. Michael entered the space once and it did not happen and then Georgi entered and it did happen, so probably how the users are placed and some situation cause this, so I would still rate it a two. Ease of solution, you (**I**) would probably also know that better, but I would say a small modification, it would probably suffice if the transition in this situation is slower. If a big change in perspective occurs, it should not happen as quickly, so the transition is a bit stretched over time. This might cause the perspective to be slightly incorrect for a short time, but I do not think anyone would notice, so the focus should be that such big changes should not be as noticeable I think, as this really breaks the experience a lot. It is probably better if the perspective is not correct for a short amount of time and the transition

takes two or three seconds and then most people probably will not notice the transition, or you may notice something changes, because someone entered the space, but it is not as obvious.
M: I agree, so some ease in and ease out maybe.
I: So for ease of solution, maybe a bit more detail on the actual algorithm.
Short explanation about the parameters for the D-Stream algorithm for each expert, see section 4.4.
W: So if the system already supports what we suggested, I would rate this one, if there is some additional work or code change necessary, some additional conditions or similar, then I would rate it a severity of two. But this issue was something that really disturbed me, because, well sometimes you do not always realize that someone enters. When I was engrossed in the system and I did not realize that the others entered or left and suddenly the camera jumps and you do not know what happened.
***G** hums approvingly.*
***W** reads the title of the sixth heuristic, "Players feel that they can navigate the system, even when multiple users are present".*
W: We actually already talked about that. I think this is not violated, but you have to learn some understanding about how to navigate. If one understands that with the clustering, that you have more influence if not everyone is in a different corner, but I think that is part of the system and I do not see a violation here.
M: Yes, I also see it that way.
I: I guess the third heuristic also plays a part in this, right?
M: Of course, some of these issues are not wholly independent from one another.
W: The remaining question is, even though it is hard to say, if you do not have three people, what happens if there are 10 people, which do not coordinate as easily and will also spread out over a larger area. So for the area downstairs maybe 5 or 6 people would already nearly fill up the whole space and that leads to the question, which possibilities that algorithm would have left if they are all placed more or less equally distributed. I guess the influence from a single user would be pretty slim, so what should one do about that? But if everyone acts together and coordinates, it would work, but it might be a problem with an increased amount of users.
G: Also with a lot of people, there is the possibility that someone will obstruct your view and you will not be able to see the projection as a whole, that already happened with us three and that also influences the effect of the system.
W: True, but with three people it worked really okay.
***W** reads the title of the fourth heuristic, "The system provides a pleasant, consistent viewing experience".*
M: Well, that is influenced by some technical limitations, if you stand to close, it will degrade.

W: Yes, and if you stand far away from the group, your experience will not be as pleasant, but we already talked about that. Of course this is a violation, but that will not be possible to solve.
G: Well, it was actually also pleasant to just step back for a moment and let the others navigate for a bit.
M: Yes, it actually might be impressive enough to just stand around see the view change. Austrian way of interacting.
All experts laughing.
W: Yes, so that works really well. Apart from that, I would have liked to be able to move a bit faster, but I am not sure which heuristic that would be assigned to, maybe the first one? I mean, the movement is intuitive, but I sometimes wished that it would react a bit faster and more direct. But again, probably also part of how the use case is designed, if navigation is too fast it might be over too soon or boring, so I would have liked some more action, do not know how you felt about that. And also if you could get closer to some objects, like I was not sure if we could actually go there or not, that was a bit hard to deduce. Sometimes we just waited for the wisp to signal us, so we knew that we can progress.
I: How did the others feel about the speed?
M: I actually thought it was pleasant.
G: The speed of the camera?
I: Yes. Did you feel like the position mapping did not work, or something else?
M: Well no, but something was ...
G: Left and right was slower than forwards and backwards, that irritated me a bit.
M: I also felt everything was more stretched, like a tunnel, the more we progressed. So we moved backwards. Was that the case, or was that just me?
W, **G**: No.
I: The camera moves to wherever each user is standing.
The discussion then moved to a technical level on how everything was implemented and concluded. No more issues were named by the experts.

Appendix B

Contents of Supplied Media

Format: CD-ROM, Single Layer, ISO9660-Format

B.1 Documents

Path: /

Interactive Stereoscopic Content for Multiple Co-located Users.pdf
: Thesis document

B.2 Project source

Path: /src/

Root folder . . . . Should be selected to open the sources as *Unity* project in the open project dialog

config.json . . . . Example configuration file used by the system

gameConfig.json . . . . Example configuration file used by the *Hansel and Gretel* use case

B.3 Project package

Path: /lib/

MultiUserStereo.unitypackage *Unity* package that can be used as a basis to create a new use case built upon the provided system

B.4 Binaries

Path: /bin/

InteractiveWoods.exe . Use case application compiled for *Windows* architecture

InteractiveWoods_Data/ Contains compiled assets created during *Unity* build

config.json . . . . System configuration file read by the application during startup

gameConfig.json . . . . Game configuration file read during startup and used to control the *Hansel and Gretel* use case

B.5 Online sources

Path: /online/

3D-Fernseher_locafox.de.pdf Print view copy of *Locafox.de*[25]

References

Literature

[1] Charu C. Aggarwal et al. “A Framework for Clustering Evolving Data Streams”. In: *Proceedings of the 29th International Conference on Very Large Data Bases*. Vol. 29. Berlin, Germany, 2003, pp. 81–92 (cit. on p. 19).

[2] Maneesh Agrawala et al. “The Two-User Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space”. In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH ’97. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 327–332 (cit. on pp. 8, 9, 12, 17, 20).

[3] Yixin Chen and Li Tu. “Density-based Clustering for Real-time Stream Data”. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Jose, California, USA: ACM, 2007, pp. 133–142 (cit. on pp. 19–22).

[4] Carolina Cruz-Neira et al. “The CAVE: Audio Visual Experience Automatic Virtual Environment”. *Communications of the ACM* 35.6 (June 1992), pp. 64–72 (cit. on pp. 7, 11, 17, 25).

[5] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. “A Systematic Review of Cybersickness”. In: *Proceedings of the 2014 Conference on Interactive Entertainment*. IE2014. Newcastle, NSW, Australia: ACM, 2014, 8:1–8:9 (cit. on pp. 6, 15).

[6] Heather Desurvire, Martin Caplan, and Jozsef A. Toth. “Using Heuristics to Evaluate the Playability of Games”. In: *CHI ’04 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’04. Vienna, Austria: ACM, 2004, pp. 1509–1512 (cit. on pp. 31, 32).

[7] Heather Desurvire and Charlotte Wiberg. “Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration”. In: *Online Communities and Social Computing: Third International Conference*. OCSC ’09. San Diego, CA: Springer, 2009, pp. 557–566 (cit. on pp. 31, 32).

[8] Andreas Friedl. "Integration of Mobile Devices into a Floor-Based Game to Increase Player Dynamics". MA thesis. Hagenberg, Austria: University of Applied Sciences Upper Austria, School of Informatics, Communications and Media, 2015 (cit. on p. 37).

[9] Bernd Fröhlich et al. "Implementing Multi-Viewer Stereo Displays". In: *WSCG 2005 Full Papers Conference Proceedings*. Plzen, Czech Republic: UNION Agency, 2005, pp. 139–146 (cit. on pp. 4, 6, 9, 12, 13, 20).

[10] Paul J. Kane, Patrick A. Cosgrove, and Cathleen D. Cerosaletti. *Stereoscopic display system with flexible rendering for multiple simultaneous observers*. US Patent 8,217,996. New York, USA, 2012 (cit. on pp. 4, 13, 14).

[11] Yoshifumi Kitamura et al. "Interactive Stereoscopic Display for Three or More Users". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 231–240 (cit. on pp. 9, 10).

[12] Yoshifumi Kitamura et al. "The Illusionhole with Polarization Filters". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '06. Limassol, Cyprus: ACM, 2006, pp. 244–251 (cit. on pp. 9, 10, 20).

[13] Robert Kooima. *Generalized Perspective Projection*. Tech. rep. June. Baton Rouge, LA, USA: Louisiana State University, Computer Science and Engineering Division, 2009, pp. 1–7 (cit. on pp. 25–27).

[14] Hannu Korhonen and Elina M. I. Koivisto. "Playability Heuristics for Mobile Games". In: *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*. MobileHCI '06. Helsinki, Finland: ACM, 2006, pp. 9–16 (cit. on pp. 31, 32).

[15] Alexander Kulik et al. "C1x6: A Stereoscopic Six-user Display for Co-located Collaboration in Shared Virtual Environments". In: *Proceedings of the 2011 SIGGRAPH Asia Conference*. Hong Kong, China: ACM, 2011, 188:1–188:12 (cit. on p. 13).

[16] Hammond Laurens. *Stereoscopic motion-picture device*. US Patent 1,506,524. New York, USA, Aug. 1924 (cit. on p. 4).

[17] Otto Naderer. "Crowd Tracking and Movement Pattern Recognition". MA thesis. Linz, Austria: Johaness Kepler University Linz, 2015 (cit. on pp. 18, 27, 43).

[18] Jakob Nielsen. *Usability Engineering*. Elsevier, 1994 (cit. on pp. 31, 32, 34, 35).

[19] David Pinelle, Nelson Wong, and Tadeusz Stach. "Heuristic Evaluation for Games: Usability Principles for Video Game Design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08. Florence, Italy: ACM, 2008, pp. 1453–1462 (cit. on pp. 31, 32, 34).

[20] Armando de la Re et al. "An Inexpensive 3D Stereo Head-tracked Dual Viewing Single Display". In: *Proceedings of the International Conference on Computer Graphics Theory and Applications (VISIGRAPP 2012)*. Rome, Italy, 2012, pp. 503–506 (cit. on pp. 4, 8).

[21] A. Fleming Seay et al. "Simulator Sickness and Presence in a High Field-of-view Virtual Environment". In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '02. Minneapolis, Minnesota, USA: ACM, 2002, pp. 784–785 (cit. on p. 15).

[22] N. K. Sharma, R. C. Jain, and Manoj Yadav. "Efficient and Fast Clustering Algorithm for Real Time Data". *International Journal of Emerging Technology and Advanced Engineering* 2.10 (2012), pp. 231–236 (cit. on p. 22).

[23] Phil Surman et al. "Laser-based multi-user 3-D display". *Journal of the Society for Information Display* 16.7 (2008), pp. 743–753 (cit. on pp. 5, 9, 13).

[24] Hakan Urey et al. "State of the Art in Stereoscopic and Autostereoscopic Displays". *Proceedings of the IEEE* 99.4 (Apr. 2011), pp. 540–555 (cit. on pp. 4, 6).

Online sources

[25] Locafox.de. *3D-Fernseher: Der Unterschied zwischen Aktiv- und Passiv-3D*. Images licensed under CreativeCommons Attribution-NoDerivatives 4.0 International, found at https://creativecommons.org/licenses/by-nd/4.0/. 2015. URL: https://www.locafox.de/magazin/3d-fernseher-aktiv-oder-passiv-3d/ (visited on 09/15/2016) (cit. on pp. 14, 53).

Messbox zur Druckkontrolle

— Druckgröße kontrollieren! —

Breite = 100 mm
Höhe = 50 mm

— Diese Seite nach dem Druck entfernen! —