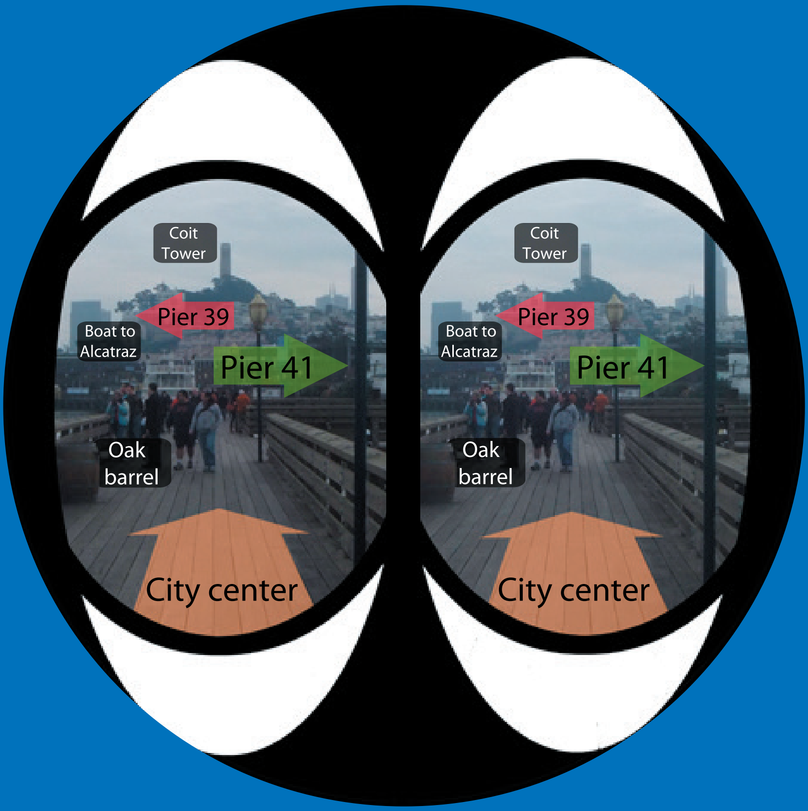
Department of Media Technology

Depth Perception of Augmented and Natural Scenes through Stereoscopic Systems

---

Mikko Kytö



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 25/2014

Depth Perception of Augmented and Natural Scenes through Stereoscopic Systems

Mikko Kytö

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall AS1 of the school on 21 March 2014 at 12.

Aalto University
School of Science
Department of Media Technology
Visual Media Research Group

Supervising professor
Prof. Pirkko Oittinen

Thesis advisor
Dr. Jukka Häkkinen

Preliminary examiners
Prof. J. Edward Swan II, Mississippi State University, United States

Dr. Ernst Kruijff, Bonn-Rhein-Sieg University of Applied Sciences, Germany

Opponent
Prof. Nicolas Holliman, The University of York, United Kingdom

Aalto University publication series
DOCTORAL DISSERTATIONS 25/2014

ISBN 978-952-60-5584-8
ISBN 978-952-60-5585-5 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-5585-5

Unigrafia Oy
Helsinki 2014

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Mikko Kytö

Name of the doctoral dissertation
Depth Perception of Augmented and Natural Scenes through Stereoscopic Systems

Publisher School of Science

Unit Department of Media Technology

Series Aalto University publication series DOCTORAL DISSERTATIONS 25/2014

Field of research Media Technology

Manuscript submitted 12 November 2013 **Date of the defence** 21 March 2014

Permission to publish granted (date) 17 February 2014 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

The ability to perceive the world with two eyes is called stereoscopic perception. The slightly different viewing positions of the two eyes provide three-dimensional information about their overlapping field of view. To utilize this common human ability technologically, stereoscopic three-dimensional (S3D) systems, for both imaging the real world with cameras and visualization in displays, need to be designed and developed. However, there is limited quantitative knowledge regarding visual performance and viewing experience, and more information is needed to better define the requirements for S3D systems. The overall research question of the dissertation addresses the benefits of stereoscopic functionality and its incorporation into guidelines for the design of stereoscopic systems. In this dissertation, human depth perception through S3D systems is evaluated based on performance measures and viewing experience attributes.

Typically, depth perception through S3D systems is investigated at distances that are within arm's reach; however, we focus primarily on distances from 2 m to 30 m away from the observer, which is called action space. Within the action space, we study the depth perception in augmented reality (AR) and natural scenes. In the visual AR applications, the observer sees virtual objects that are aligned with the surrounding 3D reality in real time. Two types of AR scenes are investigated: augmented objects above the ground plane and behind the wall. For both AR scene types, the added performance of stereoscopic viewing was examined and a novel visualization approach was conceptualized and tested. In the case of natural scenes, the stereoscopic perception was investigated with crowds and typical mobile imaging. In the crowds scene type, we studied the discrimination accuracy in a head counting task. In typical indoor imaging scene types, the viewing experience was evaluated by varying the camera separation. All the experiments were conducted in set-ups constructed from software and hardware components.

The results of the experiments indicate that the benefits of stereoscopic viewing are significant within the chosen scene types. The dissertation contributes three key results that are applicable to stereoscopic systems: 1) a novel visualization approach for augmented reality that involves adding virtual reference objects to the scene; 2) quantifying the performance of stereoscopic perception within the action space for virtual objects that lie above the ground plane or behind the wall and for perceiving crowds; and 3) identifying the effect of the camera separation range on the experience of viewing stereoscopic photographs.

Keywords Depth perception, stereoscopic perception, depth judgment, augmented reality, stereoscopic imaging, stereoscopic visualization, crowd perception, depth cue integration, stereo camera measurement, viewing experience, action space

ISBN (printed) 978-952-60-5584-8 **ISBN (pdf)** 978-952-60-5585-5

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2014

Pages 176 **urn** http://urn.fi/URN:ISBN:978-952-60-5585-5

Tiivistelmä

Tekijä
Mikko Kytö

Väitöskirjan nimi
Syvyyshavaitseminen lisätyssä todellisuudessa ja luonnollisissa näkymissä stereoskooppisten systeemien läpi katsottuna

Julkaisija Perustieteiden korkeakoulu

Yksikkö Mediatekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 25/2014

Tutkimusala Mediatekniikka

Käsikirjoituksen pvm 12.11.2013 **Väitöspäivä** 21.03.2014

Julkaisuluvan myöntämispäivä 17.02.2014 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Ihmisen kykyä muodostaa kolmiulotteinen havainto kahdella silmällä kutsutaan stereoskooppiseksi havaitsemiseksi. Kaksi hieman toisistaan poikkeavaa silmien sijaintia mahdollistaa kolmiulotteisen havainnon silmien yhteisestä näkökentästä, mikä on kasvattanut visuaalista suoriutumiskykyä tietyissä tilanteissa. Stereoskooppisen havaitsemisen hyödyntäminen vaatii stereoskooppisesti kolmiulotteisten (S3D) systeemien kehittämistä, mitä varten tarvitaan lisää kvantitatiivista tietämystä ihmisen visuaalisesta suoriutumiskyvystä ja katsomiskokemuksesta S3D-systeemien kanssa. Väitöskirjan tutkimuskysymyksenä on, että kuinka suuret ovat stereoskooppisen toiminnallisuuden tuomat hyödyt ja kuinka niitä voi soveltaa systeemien suunnittelussa. Tässä väitöskirjassa mitataan ihmisen suoriutumiskykyä ja katselukokemusta S3D-systeemien läpi syvyyshavaitsemisen näkökulmasta.
Tyypillisesti stereoskooppista syvyyshavaintoa on tutkittu lyhyillä, käden ulottuvilla olevilla etäisyyksillä. Tässä työssä keskitytään kuitenkin pääsääntöisesti etäisyyksiin, jotka ovat 2 - 30 m etäisyydellä havaitsijasta. Tätä etäisyysaluetta kutsutaan toiminnalliseksi etäisyysalueeksi. Valitut näkymätyypit toiminnallisella etäisyysalueeen sisällä ovat lisätty todellisuus, ihmisjoukot ja tyypilliset mobiilin valokuvauksen näkymät. Lisätyssä todellisuudessa havaitsija näkee näkökentässään virtuaalisia objekteja, jotka ovat sijoitettu todellinen ympäristö huomioiden. Tässä työssä tutkitaan kahta lisätyn todellisuuden näkymätyyppiä: virtuaaliset objektit maatason yläpuolella ja virtuaaliset objektit seinän takana. Syvyyden tarkempaa havaitsemista varten kehitettiin visualisointitapa, jota kokeiltiin valituissa lisätyn todellisuuden näkymätyypeissä. Ihmisjoukkojen tapauksessa tutkittiin stereoskooppisen havaitsemisen hyötyjä. Tyypillisten mobiilikuvauksen näkymien tapauksessa mitattiin katsomiskokemusta. Kaikki väitöskirjan kokeet on tehty koehenkilötestein, joita varten on rakennettu koeympäristöjä ohjelmisto- ja laitteistokomponenteista.
Kokeiden tulokset osoittavat, että stereoskooppisen havaitsemisen tuoma hyöty on hyvin merkittävää valituissa näkymätyypeissä. Väitöskirja tarjoaa kolme päätulosta, jotka ovat sovellettavissa S3D-systeemien kehittämiseen: 1) uuden lisättyyn todellisuuteen kehitetyn visualisointitavan, jossa näkymään lisätään virtuaalisia referenssiobjekteja, 2) stereoskooppisen havaitsemisen tuoman hyödyn kvantifioinnin toiminnallisella etäisyysalueella valituissa lisätyn todellisuuden näkymissä sekä ihmisjoukkonäkymissä ja 3) kameroiden välisen etäisyyden vaikutuksen katselukokemukseen.

Avainsanat Syvyyshavaitseminen, stereoskooppinen havaitseminen, syvyysarvio, lisätty todellisuus, stereoskooppinen kuvaus, stereoskooppinen visualisointi, ihmisjoukkojen havaitseminen, syvyysvihjeintegraatio, mittaus stereokameralla, katsomiskokemus, toiminnallinen etäisyysalue

ISBN (painettu) 978-952-60-5584-8 **ISBN (pdf)** 978-952-60-5585-5

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2014

Sivumäärä 176 **urn** http://urn.fi/URN:ISBN:978-952-60-5585-5

Preface

The research for this dissertation was conducted in the Visual Media research group at the Department of Media Technology. The research was mainly conducted as a part of the UI-ART (Urban contextual interfaces with multimodal augmented reality) project, and it was funded by the MIDE (Multidisciplinary Institute of Digitalization and Energy) program.

I am grateful to my supervisor, Prof. Pirkko Oittinen, for offering me this interesting challenge and giving me her time, support and valuable feedback during the writing process. I thank Prof. J. Edward Swan II and Dr. Ernst Kruijff for their preliminary examination of the dissertation. I welcome Prof. Nick Holliman from the University of York to act as my opponent at the dissertation. I value the professional comments and feedback from my instructor, Dr. Jukka Häkkinen.

I thank my colleagues for offering peer support and welcome lunch discussions. I am especially grateful to Aleksi, with whom I have collaborated on much of my work. I thank my friends in scouting and sailing. Scouting has offered me a variety of things to think about and do. Sailing with Nefertiti has given me relaxing weeks during which I have not thought about this dissertation at all, and I thank you, friends, for those moments. I want to thank my parents and brothers for being supportive. Last but not least, I thank Karoliina for encouraging me to finish the dissertation.

Helsinki, February 27, 2014,

Mikko Kytö

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Kytö, M., Nuutinen, M., and Oittinen, P. Method for measuring stereo camera depth accuracy based on stereoscopic vision. In *SPIE/IS&T Electronic Imaging (vol. 7864), Three-Dimensional Imaging, Interaction, and Measurement*, San Francisco, 7864I-1-9, 2011.

II Kytö, M., Mäkinen, A., Tossavainen, T., and Oittinen, P. Stereoscopic Depth Perception in Video See-through Augmented Reality within Action Space. *Journal of Electronic Imaging (to be published)*, 23, 1, 16 pages, 2014.

III Kytö, M., Mäkinen, A., Häkkinen, J., and Oittinen, P. Improving Relative Depth Judgments in Augmented Reality with Auxiliary Augmentations. *ACM Transcations on Applied Perception*, 10, 1, 6:1-22, 2013.

IV Kytö, M., Häkkinen, J., and Oittinen, P. Stereoscopic viewing facilitates the perception of crowds. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Klagenfurt, 49-53, 2011.

V Kytö, M., Hakala, J., Oittinen, P., and Häkkinen, J. Effect of stereo camera separation on viewing experience of stereoscopic photographs. *Journal of Electronic Imaging*, 21, 1, 011011-1-9, 2012.

Author's Contribution

Publication I: "Method for measuring stereo camera depth accuracy based on stereoscopic vision"

The present author designed and conducted the experiments and analyzed the data. The study was the continuation of a Master's thesis that was instructed by Dr. Mikko Nuutinen, who gave some background information about how to start developing the test target. The present author wrote the publication, and Dr. Mikko Nuutinen and Prof. Pirkko Oittinen have participated in proof reading.

Publication II: "Stereoscopic Depth Perception in Video See-through Augmented Reality within Action Space"

The present author developed the concept, designed and conducted the experiments and analyzed the data. Dr. Timo Tossavainen and Aleksi Mäkinen were the main developers of the software used in the experiments. The author wrote the publication, except the software, which was written by Aleksi Mäkinen. Prof. Pirkko Oittinen helped to increase the readability of the text. The all authors participated in proof reading.

Publication III: "Improving Relative Depth Judgments in Augmented Reality with Auxiliary Augmentations"

The present author developed the concept, designed and conducted the experiments and analyzed the data. Aleksi Mäkinen was the main developer of the software which was used in the experiments. The present author wrote the publication, except the software part. Prof. Pirkko Oittinen helped to increase the readability of the text. Prof. Pirkko Oittinen and Dr. Jukka Häkkinen participated in writing the discussion and proof reading.

Publication IV: "Stereoscopic viewing facilitates the perception of crowds"

The present author designed and conducted the experiments and analyzed the data. The author wrote the publication. Prof. Pirkko Oittinen has helped to increase the readability of the text. Prof. Pirkko Oittinen and Dr. Jukka Häkkinen have participated in proof reading.

Publication V: "Effect of stereo camera separation on viewing experience of stereoscopic photographs"

The present author conducted experiments, analyzed data and wrote the publication. The present author has partly designed the experimental setup. M.Sc. Jussi Hakala wrote the testing software and developed the recall attention method that was used in the experiments. The content used was inspired by the camera benchmarking project, which uses similar scenes. Prof. Pirkko Oittinen helped to increase the readability of the text. Prof. Pirkko Oittinen and Dr. Jukka Häkkinen have participated in proof reading.

1. Introduction

1.1 Brief Historical Background

Humans perceive the world using two eyes. The ability to see three-dimensionality from two views is called stereopsis. From a historical standpoint, the fundamental understanding of stereopsis was established when Wheatstone (1838) invented the stereoscope. The Wheatstone stereoscope, shown in Figure 1.1, was a remarkable invention because stereoscopic three-dimensional (S3D) image could be represented with two flat two-dimensional (2D) images. The invention of the stereoscope and photography made stereoscopic viewing popular in the 1850s. Viewing devices developed by Brewster and Holmes, shown in Figure 1.2, made the viewing of S3D photographs more popular. S3D photographs became a common form of entertainment for laymen, especially in Great Britain and France. Queen Victoria showed interest in S3D images at the Great Exhibition of 1851, and a quarter-million Brewster viewers were sold in London and Paris in the three months following the event (Howard, 2012*a*, p.85). In addition to stereophotography, stereoscopy has a long tradition in the movie industry, as the first patents were developed in the 1880s. S3D movies have been publicly shown using movie projectors since 1903, when Auguste and Louis Lumière showed an anaglyph movie about a train arriving at a station; it lasted only few seconds. The first commercially succesful movies were shown in 1922 using anaglyphs. (Howard, 2012*a*, p.90) Movie theater technology has evolved from anaglyphs to passive polarization and to shutter glasses. Polarization was first utilized in movie

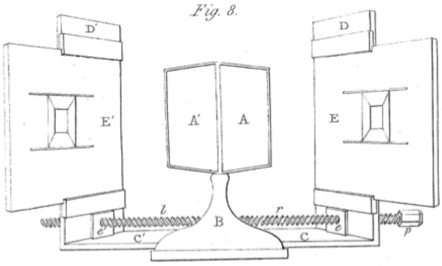


Figure 1.1. The Wheatstone (1838) stereoscope. The left image (E') is shown to the left eye through the left mirror (A'), and the right image (E) is shown to the right eye through the right mirror (A).

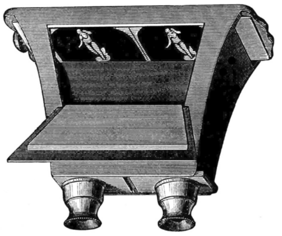


Figure 1.2. The Brewster stereoscope (Brewster, 1856).

theaters in the 1930s and again in the 1950s, when there was a boom for S3D movies. Due to viewing discomfort and poor image quality, the boom ended before the beginning of 1960s. The next boom of S3D movies took place at the beginning of the 1980s. (Howard, 2012*a*, p.91)

Electronic S3D desk-top displays were developed during the late 1970s and early 1980s, and the first electronic shutter glasses entered markets afterward (Lipton, 2012). Glasses-free S3D displays, called autostereoscopic displays, have not become common because of the difficulties in producing good stereoscopic image quality. Image quality has been a problem, especially in multiview autostereoscopic displays which can be viewed without glasses by multiple persons. However, much effort is currently being focused on the development of autostereoscopic technologies (Dodgson, 2013).

Recently, S3D movie theaters have become the most well-known venues for S3D media. The current research interest of the movie industry is focused on converting existing 2D films to S3D (e.g. Zhang et al., 2011). Classic films (including Star Wars and Titanic) have been presented again as S3D versions in the 2010s[1]. In addition to the movie industry, S3D representation has become more common with S3D televisions[2]. S3D games[3,4] and mobile phones[5] have also entered the market.

In addition to entertainment, stereoscopy has been utilized for functional purposes. Especially, stereoscopy has been used to measure distances computationally, an approach known as stereophotogrammetry (e.g., Heike et al., 2010), and visually (Barr and Stroud, 1924). In addition, stereoscopy has been utilized in detecting objects in photographs. For example a photo-finish camera was implemented stereoscopically for the Berlin Olympic games in 1936 (Stöckel, 2009) and stereoscopy was used to interpret aerial images during World War II (Kelly, 2011).

The first head-mounted display (HMD) system was introduced by Sutherland (1968). This system was able to display S3D images that were updated at an interactive speed. This invention can be considered as the beginning of augmented reality (AR), where computer graphics are aligned with the real world and shown according to the movements of the observer. The three requirements for AR applications were defined by Azuma (1997): 1) an AR application must mix real and virtual imagery, 2) it should run in real time and 3) virtual objects must be aligned (registered) with real world structures. In other words, AR is composed of an environment where reality is enhanced with augmented stimuli in real time. The augmentations - visual, aural or haptic - can give new contextual information to the user and advise the user to act in a certain way. Aural AR refers to using hearing as modality for perception augmented information, and haptic AR refers to using touch as the modality.

The research on AR has been growing over the last twenty years, and it supports advances in a wide range of applications from entertainment to endoscopic surgery. The most common applications of AR are related to the medical, manufacturing, scientific visualization, path planning, entertainment and military domains (Azuma, 1997).

---

[1] http://www.stereoscopy.com/database/movies.html (accessed 24.10.2013)

[2] http://www.sky.com/shop/tv/3d/ (accessed 24.10.2013)

[3] http://us.playstation.com/ps3/accessories/sony-playstation-3d-display-ps3/index.html (accessed 24.10.2013)

[4] http://www.nintendo.com/3ds (accessed 24.10.2013)

[5] http://www.lg.com/uk/mobile-phones/lg-P920-optimus-3d (accessed 24.10.2013)

1.2 Research Background

1.2.1 Stereoscopic System

A system that utilizes stereoscopic perception is called a stereoscopic system. Broadly defined, stereoscopic systems constitute a unifying concept in this dissertation. Stereoscopic systems address situations where stereoscopic perception is formed through a mediating technology, instead of natural viewing. Thus, the Wheatsone stereoscope (shown in Figure 1.1) can be considered the first stereoscopic system.

The initial costs for stereoscopic systems are higher compared with similar non-stereoscopic systems because specific hardware is required (Drascic and Grodski, 1993; Ware, 2004; Uratani et al., 2005). The higher initial cost is one reason why stereoscopic imaging and display have mainly been utilized in application domains where the costs of erroneous perception are high, such as medical and military domains. In medical uses, examples of increased visual performance with stereoscopic systems include mammography (Hsu et al., 1993; Goodsitt et al., 2000; Chan et al., 2005; Getty et al., 2008), the detection of fetal bony structures using ultrasound (Nelson et al., 2008) and surgery (Hanna et al., 1998), as reviewed by Van Beurden et al. (2009). In military applications, stereoscopic imaging and display have improved the tele-operation of unmanned vehicles (Drascic, 1991; Drascic and Grodski, 1993; Merritt et al., 2005). In addition to studies on performance, research on stereoscopic photography has focused on geometric issues (e.g., Jones et al., 2001; Yamanoue et al., 2006) and viewing experience (e.g., Ijsselsteijn, 2000) and viewing comfort (e.g., Lambooij et al., 2009).

In addition to stereoscopic imaging, stereoscopic content can be created using stereoscopic computer graphics. Stereoscopic computer graphics have been found to increase visual performance compared with non-stereoscopic graphics in certain applications, such as path tracing (e.g., Ware and Mitchell, 2005), the detection of overlapping labels (e.g., Peterson et al., 2009) and the accurate perception of object position (e.g., Hubona et al., 1999). With computer graphics, it is possible to generate visual virtual reality (VR) environments that are completely synthetic and interactive spaces (Milgram and Kishino, 1994). In VR environments, stereoscopic perception has been shown to increase accuracy in matching and placement tasks (e.g., Hu et al., 2002).

In stereoscopic AR, computer graphics are mixed with the real world in real time. In this dissertation, a stereoscopic system consists of four main components (shown as the green rounded box in Figure 1.3): stereoscopic imaging, computing, generating stereoscopic computer graphics and rendering on a stereoscopic display. There are two main functions for stereoscopic imaging in AR. First, stereo cameras can be used to measure a location for adding an augmented object, i.e., computer graphics (PI). Second, they can be used to capture the scene to be displayed stereoscopically (PII and PIII). Stereoscopic imaging involves the capturing of image pairs, with left images displayed to the left eye and right images to the right eye. The computing component includes geometric calibration, tracking of the user and depth measurement. Stereoscopic computer graphics are used to align additional information over the images of real world scenes. Displays are used to show scenes, whether augmented or natural, to the observer. All components of stereoscopic systems are used in the case of augmented scenes whereas a subset is relevant for natural scenes.

1.2.2 Research Approach

The dissertation has a multidisciplinary approach: it combines perceptual psychology and technological systems. In perceptual psychology, the human perception is studied empirically. The technological systems presented were constructed using software and hardware components. This approach, where

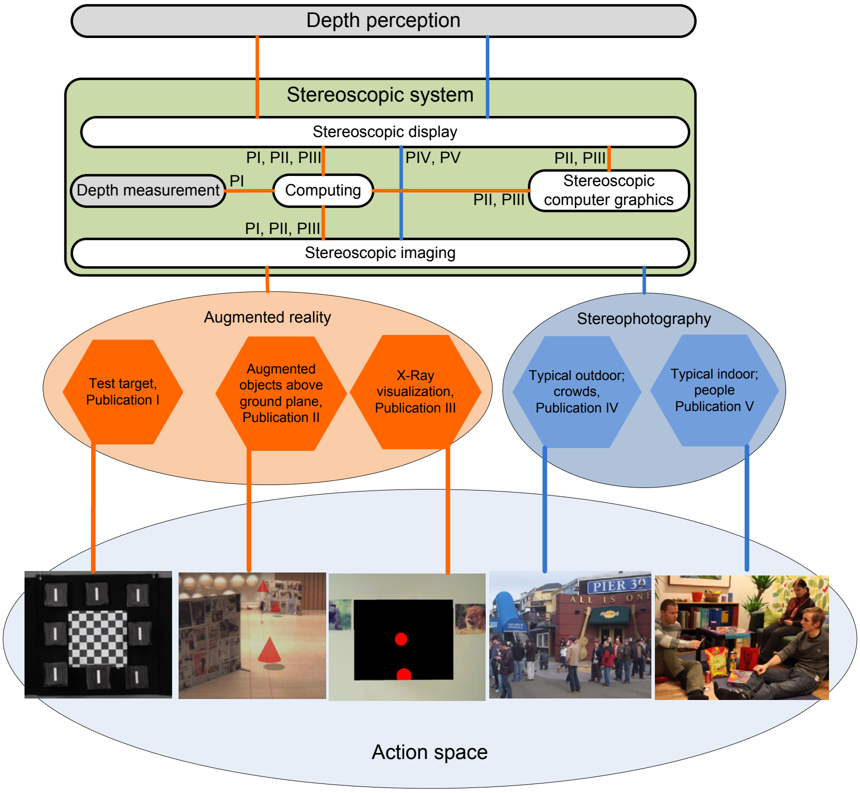


Figure 1.3. The framework of the dissertation. Depth is perceived through the stereoscopic system. The scene types are marked with hexagons, and an example scene is shown for each type. The scenes dealing with AR are marked with orange, and scenes related to stereoscopic imaging are marked with blue.

perceptual psychology is applied to construct and evaluate technological systems, is called applied perception.

The research methods in this dissertation are a combination of constructive and empirical methods. Constructive methods are employed in building setups to study human perception through stereoscopic systems. On the other hand, the developed stereoscopic systems are evaluated empirically, where the end-to-end performance and experience are measured. The performance and experience through S3D systems are evaluated by varying the stimuli and measuring depth perception responses from the participants using psychophysical methods.

Space perception refers to how humans process the sensation of physical space as a perceived space. A cognitive map of the scene is created to enable navigation and interaction. Space perception has become an active research area in both real and virtual environments. A significant research field within space perception is depth perception.

Two closely related terms, depth perception and distance perception, are used throughout the dissertation. Depth perception is either egocentric (distance from the viewpoint to an object) or exocentric (distance between two objects). Depth usually means the exocentric distance along the visual axis that is directed away from the user (Loomis et al., 1992), and it is especially used when studying the human ability to detect exocentric distance differences with stereoscopic perception (Julesz, 1971; Palmisano et al., 2010). Distance perception is usually more closely related to egocentric distance perception, meaning the absolute distance from the observer. Distance perception is also used for the perception of exocentric distances that are not along the visual axis of the user (e.g., Foley, 1980; Foley et al., 2004). The term extent perception is also used in such situations (Durgin et al., 2012).

Depth perception is a cognitive process, which is why a psychophysical method that conveys judged depth to an action is needed. Such a method can be either motoric (e.g., walking) or cognitive (e.g., verbal reporting). The judgments can be open-loop (i.e., the observers do not receive feedback regarding their judgments) or closed-loop (i.e., observers receive feedback about their judgments). The appropriate method depends on the distance and the nature of the study. If the test observers want to be aware of their judgments, for example, when users are adapted to using a system, the closed-loop task can be used (Mohler et al., 2006; Singh et al., 2010). If the effect of different conditions on depth judgment are evaluated without awareness of judgment, open-loop tasks are preferred. In this dissertation both types of tasks are used.

1.3 Scope

The design space for technological visual systems is very wide; thus, the question arises of when stereoscopic systems are preferable to non-stereoscopic systems. An extensive amount of research has been carried out about the visual performance of S3D systems compared with 2D systems, as the review by McIntire et al. (2012) proves. The benefits of S3D displays have mostly been evaluated at short distances that are within arm's reach (McIntire et al., 2012). In prehension tasks, more accurate results have been obtained with stereoscopic viewing compared with monoscopic viewing (Loftus et al., 2004; Knill, 2005; Hu and Knill, 2011). Moreover, stereoscopic systems in AR applications have been found to be useful mainly in applications within arm's reach, especially in the medical domain, as reviewed by Sielhorst et al. (2008). This study focuses on utilizing stereoscopic systems within the action space, which means a perceptual space in which distances are 2 m - 30 m from the observer (Cutting and Vishton, 1995).

The perceived space can be divided into three ranges with respect to distance: personal space (distances below 2 m), action space (distances from 2 m to 30 m) and vista space (distances over 30 m) (Cutting and Vishton, 1995). In the context of human-computer interaction, other notions have also been used to divide space, namely proxemics (Greenberg et al., 2011). Proxemics relate to the effect of distance between the user and the system's response on engagement and intimacy. The concept of proxemics is typically used when studying the effect of distance between a user and a display. Thus, the concept of proxemics has not been widely adapted to AR depth perception studies, where distances are typically investigated between a user and an augmentation. In AR, the display is typically close to the user, carried using the head or hand.

In this dissertation, the focus is on the action space, which is where humans communicate by talking and are able to move quickly (Cutting and Vishton, 1995). Ground perception is very important for movement, and thus, the visibility of the ground has a great effect on depth perception (Gibson, 1950). Within the action space, visual performance and experience through stereoscopic systems are investigated from the perspective of depth perception in the case of different scene types.

1.3.1 Scene Types

A scene type is defined as a collection of scenes that share some common and prevailing factor. The scene types were selected based on the challenges they pose to stereoscopic systems and on the capabilities of human stereoscopic vision. For stereoscopic systems, the test target-scene type (PI) is designed to be used for evaluating the depth threshold of stereo cameras.

Stereoscopic perception can be especially important in scenes that have a reduced number of monoscopic depth cues (Creem-Regehr et al., 2005), which is usually the case in AR scenes. The original review about perceptual issues in AR by Drascic and Milgram (1996) and the updated review by Kruijff et al. (2010) show how the perceived depth is affected by multiple sources in AR scenes. The latter review states that "incorrect depth interpretation is the most common perceptual problem in AR applications." In this dissertation, AR scene types are studied that can be utilized within the action space in urban environments. The scene types included are augmented objects above the ground plane (PII) and behind the wall (PIII). Perceiving augmented objects behind occluding physical surfaces, such as walls, called ”X-ray vision" (Feiner et al., 1995), is one potential application for AR.

In addition to scene types where number of depth cues are reduced, it has been shown that stereoscopic systems have great potential in visually complex scene types, such as information labels (Peterson et al., 2008) and tree graphs (Ware and Mitchell, 2005). In visually complex scenes within the action space, improved visual performance arising from stereoscopic vision has been shown in leafy environments, such as forests (Changizi and Shimojo, 2008). In this dissertation, the urban environment defines the context for the scene types. In urban scenes, the role stereoscopic perception plays in improving visual performance or experience is largely unknown. However, it can be expected that stereoscopic perception to some degree increases visual performance within the action space in urban environments. In this dissertation this is investigated with perception of crowds. Stereoscopic perception of crowds is compared with non-stereoscopic perception (PIV).

Crowds are an integral part of visually complex scenes of urban environments, and Mojsilovic et al. (2004) listed crowds as one major typical imaging category. Crowd perception has a strong position in video-based surveillance, which is an area where the accurate representation of crowds can improve the detection of humans and their actions. It is also an example of an application area, in addition to military and medical areas, where the costs of wrong decisions can be very high.

To complement the AR and crowds scene types, a typical indoor scene type (PV) is studied. The indoor scenes primarily include people. The viewing experience of stereoscopic photographs of indoor scenes is measured by varying camera separation.

1.3.2 Visual Performance Measures

The performance of stereoscopic systems is evaluated with performance measures. Howard (2012*a*, p. 93) states that the "basic parameters of performance on any task involving a response to a stimulus are accuracy (constant error), precision (mean of deviation scores), magnitude, sign and speed". For depth cues, Howard (2012*b*, p. 148) lists the following performance measures: detectability, reliability, gain, accuracy, range, spatial resolution, latency and temporal resolution. To limit the scope of this dissertation, the performance measures are narrowed down to four: detectability (measured using the depth threshold), gain (later called depth magnitude), accuracy (measured using detection accuracy) and latency (measured using performance time). The viewing experience is another factor addressed in the dissertation.

For each scene type, illustrated as hexagons in Figure 1.3, the focus is on exocentric depth perception. The measures of this thesis are shown in bold type in Figure 1.4.

Depth Threshold
In this dissertation, the depth threshold measure is applied to determine how small depth differences between objects can be perceived. The depth threshold is defined as the smallest depth difference between objects that can be correctly resolved.

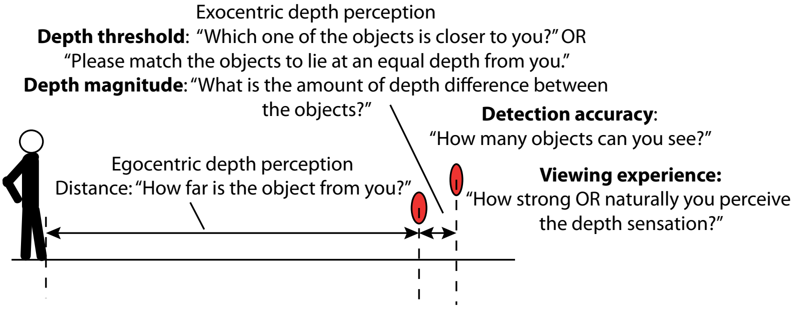


Figure 1.4. Human depth perception was measured with visual performance measures and viewing experience attributes. The measured responses of this dissertation are shown in bold type with example questions that can be asked of observers before recording their responses.

The terms depth resolution (as in PI), depth sensitivity, depth detectability and depth contrast have also been used in the literature (Cutting and Vishton, 1995) and are inversely proportional to the depth threshold. Both constant stimuli and adjustments methods have been used to measure depth thresholds through stereoscopic systems (Rolland et al., 2002).

Method of constant stimuli is usually implemented as a forced choice task, where the observers are asked to select an object that is farther or nearer than another object (e.g., McKee and Taylor, 2010). Usually 75 % is used as a criterion for the depth threshold derived from JND (just-noticeable-difference) (e.g., Howard, 1919; Jennings and Charman, 1994). We used the same criterion in our studies conducted in PI and PIII. However, other criteria have been used; for example McKee and Taylor (2010) used 67 %, and Badcock and Schor (1985) used 71%.

With the method of adjustment, the depth threshold is usually measured by giving observers the task of matching two objects to be an equal depth level. Using the method of adjustment, the standard deviation can be used to represent the difference threshold (Gescheider, 1997, p.66). Typically, depth judgments with the method of adjustment are recorded using a Howard–Dolman apparatus, which has been used in various stereoscopic depth perception studies, along with HMDs (Rolland et al., 2002, 2007).

The method of adjustments is more common when measuring the observer's depth threshold while performing a task, and it is usually implemented as a matching task. Matching tasks have been conducted in many different contexts and devices. In situations where the amount of pictorial depth cues has been limited, stereoscopic perception has decreased the depth threshold in matching tasks in teleoperation (Drascic and Milgram, 1991; Drascic, 1991), AR (Ellis and Menges, 1998; Sands et al., 2004) and virtual reality (Hu et al., 2002). Although perception through stereoscopic systems enables a low depth threshold within the personal space, it has not decreased the depth thresholds in all matching tasks. For example, Hendrix and Barfield (1995) did not find improved depth judgments in a matching task with a desktop display. In surgical tasks, Passmore et al. (2001) did not observe any improvement in accuracy with stereoscopic viewing when using a haptic feedback. For AR, Jurgens et al. (2006) showed no improvement of pointing accuracy with stereoscopic viewing in a matching task when shadows were present.

Depth Magnitude

It is important to distinguish between depth threshold and depth magnitude measures. The former measure depicts a depth difference where the depths of two objects are perceived as equal, and the latter is the magnitude of the difference between two objects. The depth magnitude is the metrical depth judgment that can be compared with the veridical depth.

Numerous protocols have been used to measure the depth magnitude. In egocentric tasks, blind walking in real environments conveys accurate results between judged and actual distances (Thomson, 1983; Rieser et al., 1990; Loomis et al., 1992; Philbeck and Loomis, 1997) up to 20 m. Another protocol that is close to blind walking is triangulation-by-walking, which is based on measuring observers' walking direction after walking and turning (e.g., Fukusima et al., 1997). Throwing has also been used as a motoric protocol (Eby and Loomis, 1987). Verbal reporting and matching techniques are cognitive responses (Foley, 1980).

In egocentric tasks, verbal reporting has been shown to result in the underestimation of distances compared with reaching responses (Gogel, 1976) and compared with blind walking responses (Loomis and Philbeck, 2008). The matching technique is usually implemented by asking observers to match pointers so that they are at an equal distance interval (Gilinsky, 1951), by matching the depth interval to the horizontal distance (Loomis et al., 1992; Ooi et al., 2001) or by matching the size of a known object (Holway and Boring, 1941; Gogel, 1976). Overall, the matching and verbal reporting tasks have yielded more errors in estimates compared with motoric responses (e.g., Loomis et al., 1992).

In exocentric tasks, the verbal reporting (e.g., Allison et al., 2009; Palmisano et al., 2010), forced-choice (e.g. McKee and Taylor, 2010) and matching techniques (e.g., Foley, 1980) are the most commonly used.

In this dissertation, closed-loop matching tasks, where objects are aligned to lie at the same distance from the viewer, are considered as depth threshold tasks. Tasks that require scaling of depth are considered depth magnitude tasks. Swan II et al. (2006) defined the closed-loop matching task as depth magnitude task, but we categorize it as a depth threshold task. We categorize open-loop matching task as egocentric task (e.g., Jerome and Witmer, 2005). In this task, the target disappears before judging the depth by matching, and the observers are unable to compare the depths of the pointer and target at the same time.

Instead of egocentric depth magnitude, this study focuses on exocentric depth magnitude. This was done for two reasons. First, the effect of stereoscopic vision on exocentric depth perception seems to be higher than on egocentric depth perception, as shown by Loomis et al. (2002). They found dissociation between egocentric and exocentric depth magnitudes and the potential of stereoscopic perception for improving the latter. Although the perception of egocentric depth was similar in both conditions, binocular and monocular, the binocular condition improved the perceived exocentric depth magnitude. Second, perceptual studies about exocentric depth magnitudes in AR within the action space are largely lacking (Dey et al., 2012). Most of the AR depth perception studies address the perception of egocentric depth (e.g., Jerome and Witmer, 2005; Jones et al., 2008; Grechkin et al., 2010; Jones et al., 2011) or the depth threshold (e.g., Rolland et al., 1995, 2002).

Generally, the greater the depth, the greater the errors in judgment of depth magnitudes. Even with real-world objects, errors in judging depth magnitudes have been present with matching responses (Gilinsky, 1951; Loomis et al., 1992). In virtual environments, it has been found that egocentric depth magnitudes are underestimated (i.e., objects are perceived as too near) when measured with blind walking (e.g., Grechkin et al., 2010; Napieralski et al., 2011), triangulation by walking (Willemsen et al., 2009), imagined walking and throwing (e.g., Witmer and Kline, 1998; Sahm et al., 2005).

The study by Dey et al. (2012) seems to be the only one that has investigated issues affecting the perception of exocentric depth magnitudes in AR with a monoscopic display. Studies that have compared the perception of exocentric depth magnitudes in the case of stereoscopic and non-stereoscopic AR systems could not be found. PIII reports the results of a study on the perception of exocentric depth magnitudes using stereoscopic perception with verbal reporting.

Detection Accuracy

Detection accuracy can be evaluated as error rate in recognizing the presence of target. In this dissertation, the detection accuracy is evaluated by comparing the judged number of targets with the actual number of targets, similar to study by Tsirlin et al. (2008).

The detection of half-occluded objects is an important capability of stereoscopic vision (Nakayama et al., 1989). Tsirlin et al. (2008) studied the detection of half-occluded depth layers by asking observers to count the number of depth layers made of glass. They found that the segregation ability varied from 3 to 6 layers. The segregation accuracy was improved as a function of disparity. However, studies are lacking on the stereoscopic detection of half-occluded objects in natural settings, where other depth cues in addition to disparity are present. Changizi and Shimojo (2008) presented a theory of why the stereoscopic detection of half-occluded objects is an important capability in leafy environments.

Detection accuracy through stereoscopic systems has been evaluated primarily in medical uses, where the detection of complex structures from complex scenes is needed. Stereoscopic perception has increased object detection accuracy in mammography and ultrasound imaging (e.g., Hsu et al., 1993; Goodsitt et al., 2000; Chan et al., 2005; Getty et al., 2008; Nelson et al., 2008). These scenes typically contain multiple overlapping objects, where the stereoscopic detection of half-occluded objects has the potential to improve object detection.

Stereoscopic detection of half-occluded objects is useful in crowded scenes in which there are multiple objects of interest in the same view. For example, stereoscopic viewing has been found to be useful in the detection of labels in crowded scenes (Peterson et al., 2009) and in path following in complex graphs (Ware and Mitchell, 2005). We studied the stereoscopic detection of half-occluded objects in an urban environment using a crowd perception task in PIV. The crowd perception task involved counting people in monoscopic and stereoscopic images of crowds.

With stereoscopic perception, the detection accuracy can also be improved in tasks that do not require depth perception. Increased detection accuracy with stereoscopic perception was experimentally shown with the detection of letters (Jones and Lee, 1981). However, this issue is beyond the scope of the dissertation, as its focus is on depth perception.

Performance Time

Performance time is the time interval that is required for an individual to complete a task. There are contrasting results on how stereoscopic depth affects performance times. Overall, stereoscopic perception has been shown to decrease the performance time compared with monoscopic perception in situations with natural viewing that require a motoric response. Hayhoe et al. (2009) showed that people walk faster with stereoscopic viewing, and reaching and 3D positioning tasks have been accomplished faster with stereoscopic perception (Loftus et al., 2004; Teather and Stuerzlinger, 2011).

This has also been the case when the scene is viewed through stereoscopic systems. Stereoscopic viewing has decreased operation times in motoric tasks compared to monoscopic tasks in 3D manipulation tasks with a virtual crane (McWhorter et al., 1991), in telesurgery tasks (Bowersox et al., 1998) and in wire tracing tasks (Barfield et al., 1999). Although opposite results exist for surgeon tasks (Hanna et al., 1998). For cognitive responses, Ware and Mitchell (2005) showed a decreased performance time in a path tracing task, and Jones and Lee (1981) showed a decreased time in a breaking camouflage task with binocular viewing.

Thus, stereoscopic systems can be expected to decrease performance time, especially in motoric tasks. It has been suggested that this decrease is due to the reduced uncertainty in the motoric tasks (Loftus et al., 2004). In this dissertation, the effect of stereoscopic perception on the performance time of depth judgments was measured with cognitive responses in PIII and PIV. It can also be expected that the reduced uncertainty in depth judgments reduces the performance time in cognitive tasks.

Viewing Experience
In this dissertation, the previous four measures constitute the visual capabilities of stereoscopic perception. The viewing experience through a stereoscopic system is also addressed. Viewing experience is difficult to measure and define, as noted by Seuntiens (2006). We concentrate on depth perception, as it has been suggested to be an important factor in the viewing experience in stereoscopic systems. Viewing experience is studied with the strength and naturalness of depth sensation. These two attributes are measured in PV, where participants viewed stereoscopic photographs taken with different camera separations.

In some cases, the viewing experience with S3D content has been shown to be higher compared with 2D content. For example, Seuntiens et al. (2005) found an improved viewing experience with the stereoscopic viewing condition compared with the monoscopic viewing condition when viewing still images. However, the potential of improving the viewing experience by using stereoscopic systems is not in the core of this dissertation and thus is not dealt with extensively. Extensive work on the added value of stereoscopy on the viewing experience has been carried out as exemplified by the studies of Seuntiens (2006) and Häkkinen et al. (2008).

In addition to the viewing experience, the feeling of presence has been found to be increased with stereoscopic displays compared with monoscopic displays. Prussog et al. (1994) discovered that stereoscopic viewing increased the impression of telepresence in video communications. IJsselsteijn et al. (2001) found an increased feeling of presence when participants played a rally game with stereoscopic viewing.

However, when S3D systems are improperly designed, the viewing experience can be impaired due to visual fatigue and discomfort. However, this dissertation does not provide an in-depth assessment of viewing comfort issues associated with S3D displays. The assumption is made that when the recommendations from the literature are followed (Lambooij et al., 2009; Shibata et al., 2011), the viewing experience associated with stereoscopic displays remains high.

1.4 Objectives

The overall aim of the dissertation is to help fill the knowledge gap hampering stereoscopic system design for AR and photographic applications. This leads to research questions of the benefits stereoscopic functionality brings to systems and of the design of such systems. The mission is to introduce stereoscopic systems to new application areas in which depth perception plays an important role. There is a lack of knowledge regarding how much depth perception through stereoscopic systems is improved within the chosen scene types. Although stereoscopic perception is most likely the most studied source of depth (Cutting and Vishton, 1995), studies comparing depth perception through non-stereoscopic and stereoscopic systems within the action space are limited.

Veridical depth perception is essential in many application domains. The retinal size of an object decreases as a function of distance, and therefore, veridical estimates of distance enable the veridical perception of size, which is crucial in such applications as architectural design processes and moving within a space. For example, in wayfinding, veridical distance perception is important for evaluating travel distances and for the perception of signs at their correct locations. In simulators, errors in depth perception bias the perception of space, which may lead to misjudgments in a real environment. Depth perception issues should be considered, for example, in cockpit simulators for airplanes and ships. While research on virtual environments is limited to synthetic environments, augmented reality (AR) can offer real interaction with a physical world. However, space perception issues in AR are not well understood.

The aim is operationalized under objectives O1-O5. The relations between the objectives and stereoscopic system design process are indicated in Figure 1.5. The depicted development process is a simplification, as it does not take into account issues related to displays, camera calibration and tracking. These issues are discussed in Chapter 3.

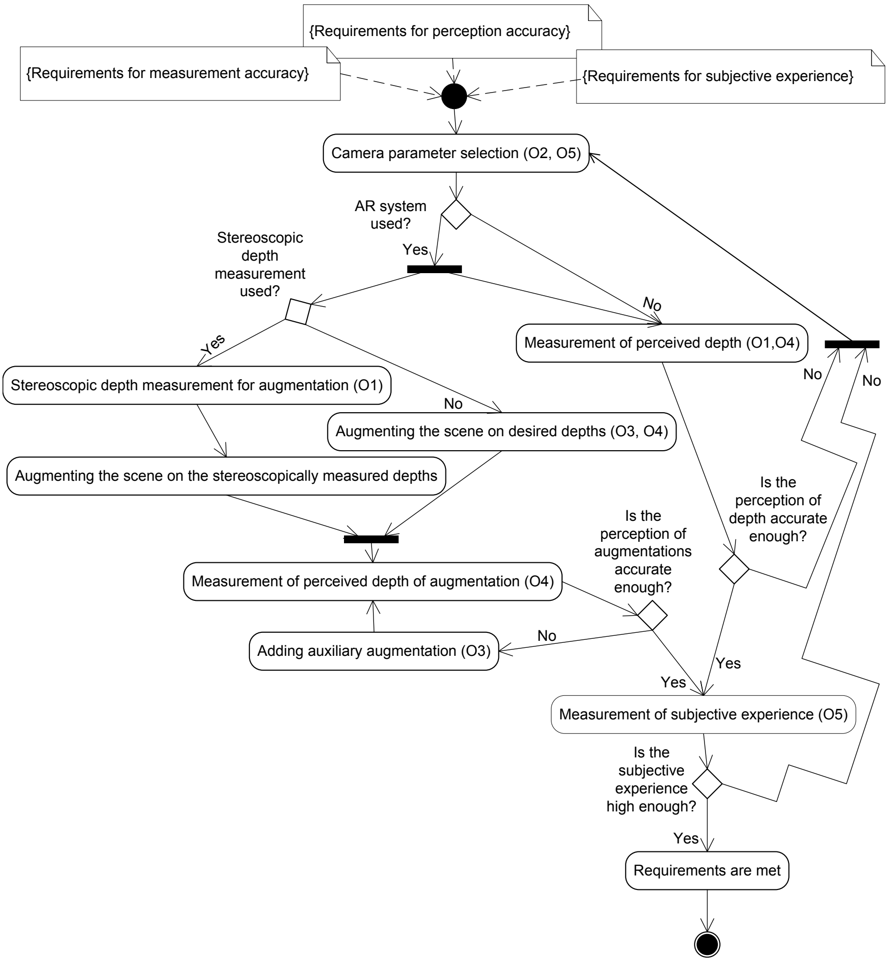


Figure 1.5. UML activity diagram for stereoscopic system development, with mapping of objectives O1-O5.

With reference to Figure 1.5, stereoscopic systems in this dissertation are used in AR and natural environments. Stereo cameras can be used to measure the distances between observers and the physical world. The measurement data obtained can be used to augment the scene. The first objective, O1, is to develop a method that allows the examination of the depth threshold of a stereoscopic system. This method can be used to select stereo camera parameters, as the stereoscopic system involves capturing images for both eyes. The selection of camera parameters for stereoscopic systems is of fundamental importance, especially the choice of camera separation. Objectives O2 and 05 relate to the selection of camera separation. O2 and O5 approach the task from two different angles: measurement and perception. Both the depth threshold and the viewing experience vary with the camera separation. In AR environments, in addition to imaging parameters, the visual scene can be augmented with virtual objects. However, the depth judgment errors in AR are significant. O3 aims to reduce depth perception errors by developing a visualization approach that is applicable to stereoscopic systems. O4 is designed to cover the added performance of stereoscopic systems compared to non-stereoscopic systems. The difference in visual performance between stereoscopic and non-stereoscopic conditions is compared in an urban environment with AR and crowd scenes. The objectives are described in greater detail below, and their study in the publications of this dissertation is indicated in Table 1.1.

O1: Conceptualize and develop a method for measuring the depth threshold of human and stereoscopic measurements

The first objective of the study is to design and develop a method that can be used to determine the depth threshold of stereo camera measurements and human viewing through stereoscopic systems. Accuracy determinations are needed when selecting camera parameters for stereoscopic AR systems, as the system performance should exceed the visual performance.

O2: Find the effect of camera separation on depth thresholds within the action space

Theoretically, the depth threshold of human stereoscopic perception and stereo cameras increases with the squared distance and an increase in camera separation decreases the depth threshold linearly. Thus, within the action space, it is tempting to increase the camera separation to decrease the depth threshold. The second objective is to determine the effect of camera separation on the perceived depth threshold and the depth threshold of stereo camera measurements. The effect was evaluated in practice with the test target, which was the outcome from the first objective.

O3: Design and implement a method to improve depth judgments concerning augmented objects within the action space

In AR applications, it is difficult to visualize depths in a manner that supports veridical perception. The errors in depth judgments are significant, and depth perception is affected by multiple factors. The third objective is to develop a visualization approach that is applicable for visualizing the relative depths between objects. This visualization approach is evaluated in the case of two types of AR scenes: augmented objects above the ground plane and X-ray visualization. The visualization approach involves adding reference augmentation to the scene and is called auxiliary augmentation (AA). AA resembles the target augmentation, called the augmented object of interest (AOI), the distances of which are judged by the participants.

O4: Establish the visual performance of perception through stereoscopic systems within the action space

The fourth objective is related to the visual performance by using a stereoscopic system in comparison with a non-stereoscopic system in the AR and outdoor crowd scene types. Performance differences between stereoscopic and non-stereoscopic systems are evaluated depending on the scene type, as presented in Table 1.1. As mentioned, in this dissertation, stereoscopic perception is studied mainly within the action space.

O5: Discover the effect of camera separation on the viewing experience of stereoscopic photographs

The separation between the left and right cameras is an important factor in stereoscopic photography, as it can be used for controlling the perceived depth of stereoscopic photographs. The fifth objective is to find out how the camera separation affects the viewing experience, where depth perception plays an important role.

Table 1.1. The publications and objectives of the dissertation. The performance and experience of depth perception through stereoscopic systems is investigated in the case of augmented and natural scenes. The measured responses are gathered depending on scene type. The responses in bold type are addressed in this dissertation.

Pub.	Obj.	Scene type	Measured response
I	1, 2	AR: Test target	**Depth threshold of stereo camera measurement and humans**, The effect of focal length on depth threshold in stereo camera measurement
II	3, 4	AR: Augmented objects above the ground plane	**Depth threshold**, Uncertainty
III	3, 4	AR: X-ray visualization	**Depth threshold, Depth magnitude, Performance time**, Uncertainty
IV	4	Typical outdoor setting; crowds	**Detection accuracy, Performance time**, Perceived ease, Preference
V	5	Typical indoor setting; people	**Viewing experience, Strength of depth sensation, Naturalness of depth sensation**, Attitude towards S3D photographs, Qualitative data

1.5 Main Contributions

Overall, the dissertation increases critical knowledge regarding the design of stereoscopic systems for use within the action space. The knowledge was gained by combining constructive and psychophysical methods. The main contribution of the study corresponds to O3 and O4. A novel visualization approach improves the depth perception in AR. Depth perception is currently the largest perceptual problem in AR applications (Kruijff et al., 2010), and using the visualization approach coupled with stereoscopic perception was found to reduce the depth judgment errors markedly. Adding the AAs to the scenes showed additive depth cue integration with binocular disparity and relative size, making it an efficient approach for visualizing the relative depth between objects.

For augmented objects above the ground plane, the auxiliary augmentations lowered the depth threshold values (PII). In the X-ray visualization case, the depth threshold and judgments of depth magnitude were improved using stereoscopic viewing with auxiliary augmentations (PIII). Stereo AR glasses were constructed for the experimentation. Stereoscopic perception is one way to further improve the depth perception in AR applications and within the action space.

In the case of outdoor crowds, the detection accuracy of counting humans was found to increase with the stereoscopic condition (PIV), suggesting that crowd analysis benefits from stereoscopic presentation (corresponds to O4). The study showed that people were distinguished from each other more easily and the depths between people were assessed to be easiest with stereoscopic viewing. This revealed a potential application area for stereoscopic systems, namely, the visual inspection of crowds in video-based surveillance applications.

The method for comparing the depth thresholds of human and stereo camera measurements corresponds to O1. The method uses a self-built test target that allows determining the depth threshold by using a rank ordering of multiple depth levels. One advantage of the method is that rank ordering is a faster way to measure the depth threshold than pair-wise comparisons. The results using this method indicated that increased camera separation did not significantly decrease the depth threshold of stereoscopic measurements or of humans (corresponding to O2). However, this result should considered as suggestive, as the measurements were not repeated.

The dissertation also contributes findings (corresponding to O5) related to the role of camera separation as a key factor not only in controlling perceived depth in stereoscopic systems but also the visual experience. The effect of camera separation on the subjective experience of stereoscopic photographs, in addition to the depth threshold, was explored in the case of typical indoor imaging scenes (PV). The results showed that smaller camera separations were preferred and that the cardboard effect (quantified with a roundness factor) did not affect the naturalness of depth sensation.

This dissertation is relevant for system developers from both software and hardware perspectives. The developed visualization approach, auxiliary augmentations, relates to both software and visualization, while imaging systems developers can use the results from the stereoscopic imaging studies in hardware design.

1.6 Structure of the Dissertation

The structure of the study is as follows. The theoretical foundation and related research is presented in Chapters 2 and 3. The essential properties of human depth perception are introduced in Chapter 2. An understanding of human depth perception is needed when evaluating factors affecting depth perception through stereoscopic systems, which is the focus of Chapter 3.

The research of the dissertation is reviewed in Chapter 4. First, we concentrate on AR. The method for measuring the depth threshold in stereoscopic systems is presented, and then, the visualization approach of auxiliary augmentations is introduced. AR is followed by stereoscopic photography within the context of crowd perception and typical indoor scenes. The results are discussed in Chapter 5, and conclusions are drawn in Chapter 6.

2. Human Depth Perception

2.1 Depth Cues

Space perception refers to the process by which the sensation of the physical space is transformed into the perceived space. Humans form a cognitive map of scenes, which enables navigation and interaction with the scenes. Depth perception is a more specific research field within space perception.

Human depth perception is a complex issue and is usually explained with a combination of several depth cues. Helmholz (1867) was the first to introduce classical depth cues, which have been present in the arts for centuries (Cutting and Vishton, 1995). Table 2.1 provides a list of depth cues and their properties.

Depth cues can be monocular, binocular or oculomotor, as indicated in the first column in Table 2.1. Monocular depth cues can be perceived with one eye, binocular depth cues can be perceived with two eyes (stereoscopic perception). Oculomotor depth cues are perceived through muscles in the eye. There are also pictorial depth cues (shown in Figure 2.2), which can be perceived by viewing a 2D image; they are a subset of monocular depth cues.

Table 2.1. The different depth cues and their properties.

Depth cue	Cue type: (M)onocular (B)inocular (O)culomotor	Effective distance range: (P)ersonal (A)ction (V)ista	Type of distance: (A)bsolute (R)elative
1. Occlusion	M	P, A, V	R
2. Height in visual field	M	A, V	A, R
3. Relative size	M	P, A, V	R
4. Familiar size	M	P, A, V	A
5. Linear perspective	M	P, A, V	R
6. Texture density	M	P, A, V	R
7. Foreshortening	M	P, A, V	R
8. Shading	M	P, A, V	R
9. Cast shadows	M	P, A, V	R
10. Aerial perspective	M	V	R
11. Brightness	M	P, A, V	R
12. Color	M	P, A, V	R
13. Binocular disparities	B	P, A	R
14. Accommodation	M, O	P	A
15. Convergence	B, O	P	A
16. Motion parallax	M	P, A	A, R
17. Kinetic depth effect	M	P, A	R

A division of the perceptual space into depth ranges is important because the depth threshold of certain depth cues depends on the associated distance (Nagata, 1991; Cutting and Vishton, 1995), as illustrated in Figure 2.1. Figure 2.1 involves simplifications, as the depth thresholds depend on more detailed properties of the scenes and observer. However, the figure represents an important effort toward quantifying depth thresholds as a function of distance. The second column in Table 2.1 shows the distance range where the depth cues are effective.

Depth cues have differences in the type of depth (absolute, relative, ordinal) they cue, as shown in the last column of Table 2.1. Absolute depth is the depth magnitude from the observer, relative depth is the depth magnitude between objects and in ordinal depth the order of depths can be distinguished but not their magnitude. All the relative depth cues in Table 2.1 are also ordinal cues.

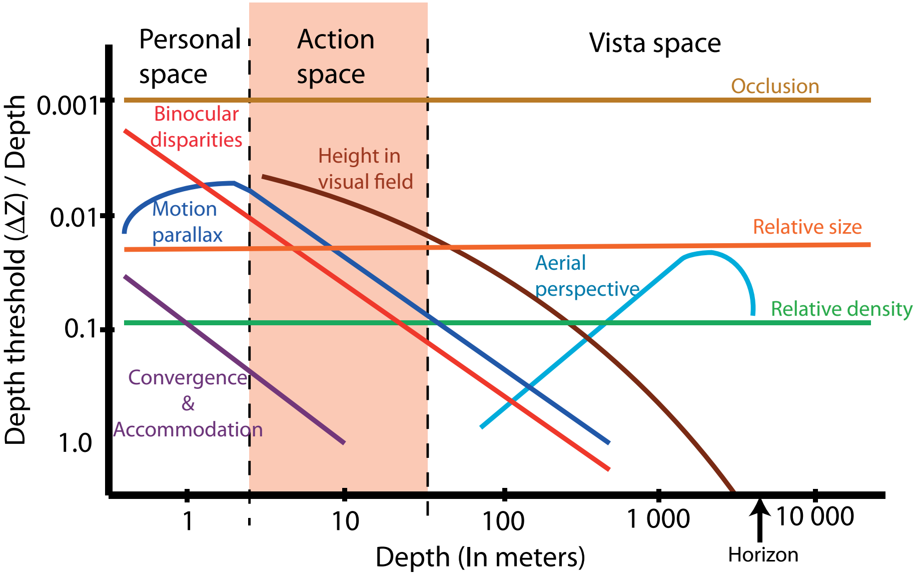


Figure 2.1. Depth thresholds ΔZ (note the inversed and logarithmic scale) of depth cues as a function of distance according to Cutting and Vishton (1995). Reprinted with permission.

Howard (2012*a*) categorized pictorial depth cues into the following categories: occlusion (also known as interposition) (Table 2.1, Cue 1), perspective (Cues 2-7), lighting (Cues 8 and 9), aerial effects (Cues 10-12) and focusing (Cues 13-15). A motion category (Cues 16-17) was also added. The categorization by Howard (2012*a*) is used to introduce depth cues.

From the list of depth cues in Table 2.1, binocular disparity is present in every publication, motion parallax and relative size in PII and PIII. An additional depth cue for PII was cast shadows. PIII has additional depth cues of occlusion and height in visual field (expressed as angular declination in the publication). These depth cues were chosen because they can offer accurate depth perception within personal and action spaces (see Figure 2.1).

Occlusion

Occlusion (Cue 1) occurs if an opaque surface hides a surface behind it. Occlusion is the only cue that is ordinal but not relative. With the occlusion cue, it is only possible to see that one object is occluding another object; it is not possible to see the relative distance between them. Occlusion is considered the most dominant depth cue (Cutting and Vishton, 1995; Howard, 2012*b*).

Transparency is opposite for occlusion. Transparency is present in everyday scenes, such as when watching through windows. Tsirlin et al. (2008) listed three types of transparency: glass-transparency, semi-transparency (called translucency by the author) and pseudo-transparency. Glass-transparency is based on light passing through a transparent surface. In semi-transparency, the scene behind the surface is partially seen through a translucent material. In pseudo-transparency, a scene is viewed through gaps in opaque objects. In this type of transparency stereoscopic perception plays an important role because stereoscopic vision facilitates the recognition of partially occluded objects (Nakayama et al., 1989). It has been claimed that the stereoscopic detection of half-occluded objects is one of the evolutionary reasons for binocular vision (Changizi and Shimojo, 2008). Stereoscopic pseudo-transparency is important in leafy environments, such as in forests, where there are numerous gaps between opaque objects (Changizi and Shimojo, 2008).

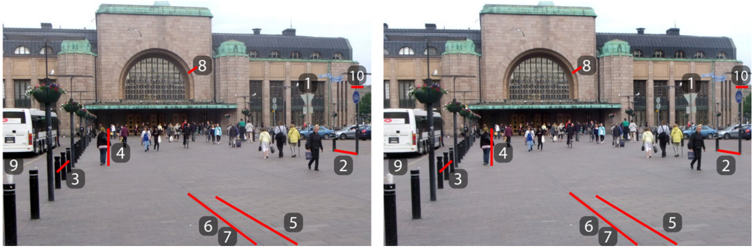


Figure 2.2. An example of pictorial depth cues in urban environments: 1. Occlusion (the traffic sign occludes the wall; therefore, it is nearer). 2. Height in visual field (the base of right pole is lower in the visual field than the base of left pole; therefore, it is nearer). 3. Relative size (the left pole appears larger than right pole; therefore, it is nearer). 4. Familiar size (the approximate size of the human is familiar; therefore, it can be used for scaling the distance). 5. Linear perspective (the edges of the paving stones collide toward infinity). 6. Texture density (the closer the paving stones, the lower their density). 7. Foreshortening (the aspect-ratio of paving stones changes according to depth). 8. Shading (the shading follows the geometrical forms of the building). 9. Cast shadow (the shadow of a bus can be seen on the ground). 10. Aerial perspective (aerial haze affects the building; therefore, it is perceived as farther away). The image can be viewed stereoscopically with the uncrossed (parallel) viewing method; therefore, binocular disparity is present.

Perspective cues

The height in a visual field (Cue 2) cue is based on the difference of the vertical positions of objects below and above the horizon. Below the horizon, the usual arrangement is that the lower the position of the object, the nearer it is perceived to be; above the horizon, the arrangement is the opposite. The relative size (Cue 3) cue is based on the fact that the retinal size of an object decreases with distance. The familiar size (Cue 4) cue is dependent on a priori knowledge of the size of an object. If the physical size of an object is known, the depth can be scaled based on the retinal size, as the retinal size decreases as a function of distance. The linear perspective (Cue 5) cue is based on the property of parallel lines converging toward infinity. The texture density (Cue 6) changes as a function of distance; the closer the object, the less dense the texture. A cue similar to texture density is foreshortening (Cue 7), which is also called the aspect-ratio. Foreshortening is based on the change of aspect-ratio of object dimensions as a function of distance; for example, a circle flattens to an oval as a function of distance (Howard, 2012*b*, p.23).

Lighting cues

Depth from shading (Cue 8) is based on the light falling on a surface and its reflectance. Another cue based on lighting is cast shadow (Cue 9). Shadows are cast by objects on surfaces, and they offer information about the spatial relations of objects. Cast shadow can also be used as a depth cue. Aerial perspective (Cue 10) is a cue that is based on decreased contrast and color saturation and an increase in blue as a function of distance. This occurs on the vista space due to optical haze and mist and is dependent on weather conditions. Brightness (Cue 11) is a cue in which the brighter the object appears the closer it is perceived to be. In addition to brightness, color has an effect on perceived depth. In this context, color (Cue 12) consists of hue and saturation. Hue has an impact on depth perception: a red area on a blue area may appear to stand out, an effect caused by chromostereopsis. Chromostereopsis may affect perceived depth due to chromatic aberration (McClain et al., 1990; Howard and Rogers, 2012, p.284). Saturation also has an effect on depth perception: as the differences between saturations increase, the depth increases (Egusa, 1981).

Focusing cues

Focusing cues rely on the human ability to fixate the optical axis of the eyes at the fixation point. Binocular disparity (Cue 13) is based on the human ability to perceive the world from two slightly different viewing points. We will focus on binocular disparity in Section 2.2.

Accommodation (Cue 14) and convergence (Cue 15) are oculomotoric cues, which are perceived through changes in muscle activity. Accommodation creates depth cues in two ways. First, the image blur increases according to the depth difference from the fixation point. Second, the eye focuses on a certain depth level by changing the shape of the lens. The muscle activity in shaping the lens can also be used as a depth cue.

The ability of the eyes to rotate toward each other is called convergence. The lines of sight collide at a point where the eyes are converged. In addition to normal fixation processes, the convergence can be perceived as a depth cue within personal space.

Motion

Motion parallax (16) is based on the movement of an observer or objects. When the observer moves, the change of object positions is usually larger for nearer objects, which can be used as a depth cue. When objects are moving, the observer perceives a greater difference in position for nearer than farther objects. The kinetic depth effect (17) is based on the rotation of objects. The surface speed depends on the radius, which can also be a cue for depth.

Depth cue integration

Human depth perception has been explained with depth cue integration theories, which can be categorized as weak and strong observer theories (Landy et al., 1995). According to the weak observer theory, depth perception is a linear combination of modules that correspond to depth cues, so the depth cues support depth perception with different weights. In the strong observer approach, the depth cues are combined in a single module, in which depth cue integration occurs. In the strong observer approach, depth perception is more challenging to predict because the depth perception is not based on a modular structure.

The debate over depth cue integration theories has continued for decades. According to current knowledge, depth perception is based on maximizing the likelihood of an object location according to the reliability of depth cues. When an observer performs this integration in an optimal way, she is called an optimal observer. The models for the optimal observer can be linear, Bayesian or non-linear (Landy et al., 2011).

Bülthoff and Mallot (1988) listed four possible integration mechanisms that are not mutually exclusive: accumulation, cooperation, disambiguation and veto. The accumulation mechanism predicts that depth perception is formed from accumulating depth cues. It is also called additive depth cue integration. Accumulation occurs if the sum of gains of different cues is more than the cue with the highest gain. In cooperation integration, the effect of the integration of depth cues is more than adding them separately. Disambiguation occurs if the sign of perceived depth changes with the integration of depth cues. In the veto mechanism, one depth cue dominates the depth perception, and the perceived depth is based one depth cue. Usually, the veto integration is a result of occlusion, which is the most dominant depth cue (Cutting and Vishton, 1995). In this dissertation, accumulative depth cue integration was found in PII and PIII.

Depth cue integration can be investigated according to performance measures (see Section 1.3.2). Jameson and Hurvich (1959) noted that the depth threshold decreases more with the combination of disparity, size, accommodation and motion parallax than if summed separately. Bruno and Cutting (1988) showed that the depth magnitude increased as a weighted sum of depth cues, which were size, angular height, occlusion and motion parallax. Moreover, the working range of depth cues can be extended by using them additionally according to distance, where, for example, disparity is used for small distances and linear perspective is used for large distances (Howard, 2012*b*).

Locally, depth perception is affected by binocular surface perception. Examples of surface perception include a depth "propagation" phenomenon (Takeichi et al., 1992), interpolation and extrapolation (Collett, 1985), and the Da Vinci phenomenon (Nakayama, 1996). The depth cues provide higher-level depth information for human visual processing, whereas the surface perception mechanisms affect lower-level vision. Surface perception is addressed in this thesis. The depth propagation phenomenon can partly explain the improved performance with stereoscopic perception in X-Ray visualization (see PIII).

Moreover, the regularities of space have been used to explain human depth perception. This probabilistic approach is based on the fact that certain structures are more common in scenes than others. The varying probabilities also have an effect on our depth perception. Yang and Purves (2003) emphasized the effect of probabilistic information on space perception. Using scene probabilities, they explained why a dip in the terrain between the observer and a target causes the target to be evaluated as farther away. A review about visual perception from the standpoint of the statistical properties of natural scenes can be found in Geisler (2008). The probabilistic approach has also been used in lower level vision research with surface perception (Nakayama and Shimojo, 1992).

2.2 Stereoscopic Perception

Stereoscopic perception is based on the two slightly different viewing points of two eyes. These two points cause a disparity due to the interpupillary distance (IPD). Human stereoscopic vision shares similarities with that of owls (Van der Willigen, 2011). The eyes of humans and owls point in nearly the same direction and thus the common field-of-view with both eyes is large (approx. 120°). This offers a wide stereoscopic viewing volume for evaluating distances binocularly. Stereopsis is important for predators for evaluating the distance and motion of the prey (Julesz, 1971). The stereoscopic vision of humans and owls is opposite that of many birds and preys. For example, pigeons have a common field of view of only 27° (Howard, 2012*b*, p. 249) but they have a wider overall field of view (340°).

Not all humans are able perceive stereoscopic depth; approximately 10 % of people are stereoblind, with estimations varying from as low as 3 % (Richards, 1970) to as high as 20 % of people (Ware, 2004). Impaired stereo vision is usually a consequence of strabismus, although it does not completely prevent stereoscopic perception (Henson and Williams, 1980; Leske and Holmes, 2004).

When two views are correlated with each other, the human visual system is efficient at finding the corresponding points between the views. Depth can be perceived even without monocular depth cues based solely on the correspondences between left and right views from random dot stereograms (Julesz, 1960, 1971). Random dot stereograms are figures that contain only randomly distributed dots and have no recognizable object when viewed monocularly. The stereoscopic perception is highly dependent on geometric and spatial issues, which are described next.

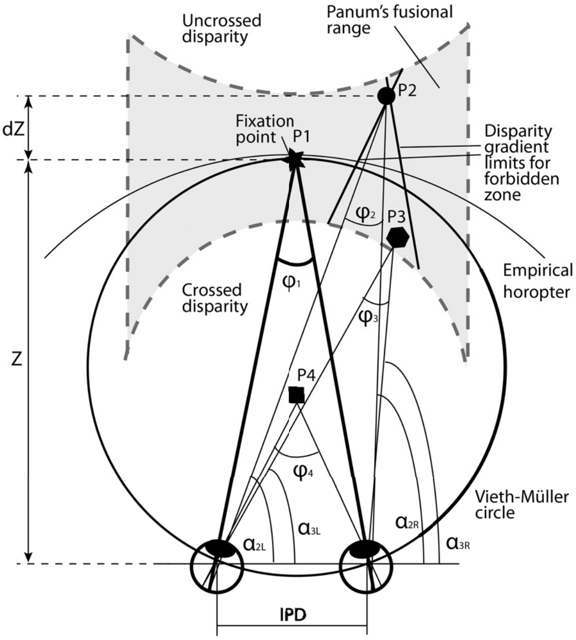


Figure 2.3. The basic geometry of stereoscopic perception. Adapted from Howard and Rogers (2012).

2.2.1 Binocular Disparities

The horizontal binocularity disparity between objects P1 and P2, shown in Figure 2.3, is the angular difference on the retina: $\varphi_{1,2} = \varphi_1 - \varphi_2$. It can be derived from geometrical principles and is calculated according to Equation 2.1

$$\varphi_{1,2} \approx \frac{IPD * dZ}{Z^2}, \tag{2.1}$$

where dZ is the depth magnitude between the objects, Z is the distance to the fixated object and IPD is the interpupillary distance. The amount of disparity is inversely proportional to the square of distance. By rearranging the equation, it can be observed that smaller depth thresholds should be detected by increasing the IPD. If a stereoscopic image is shown with a certain disparity, the smaller the IPD, the longer the depth magnitudes should appear to the user.

Equation 2.1 is a starting point for predicting the perception of depth magnitude based on stereoscopic perception. Scaling the depth magnitude (dZ) from disparity (φ) requires knowledge of the absolute distance of a fixated object (Z). The absolute distance can be perceived through oculomotor cues (accommodation and convergence) (Wallach and Zuckerman, 1963) or pictorial cues (Wallach and Zuckerman, 1963; Allison et al., 2009). An incorrectly perceived Z causes distortions in perceived depth magnitudes, as illustrated by Johnston (1991).

In environments where pictorial depth cues are limited or distances are short, disparity can improve the perception of exocentric depth magnitudes. This ability has also been observed beyond the action space in natural viewing. However, the perceived depth magnitudes are compressed (Allison et al., 2009; Palmisano et al., 2010) and affected by monocular depth cues.

In the real world with reduced pictorial cues, binocular disparity is a significant depth cue for exocentric depth magnitudes up to at least 18 m (Allison et al., 2009). Palmisano et al. (2010) studied the effect of stereoscopic perception on depth magnitude at distances over 20 m. A disparity of 3.5

arcmin caused a statistically significant improvement in depth magnitude judgments at 40 m compared with judgments with monocular viewing. With the presence of monocular depth cues, binocular disparity of 1 arcmin caused a statistically significant difference compared with the monocular condition. Foley et al. (2004) did not find improved exocentric depth magnitude judgments with stereoscopic viewing when objects were on the ground plane. To conclude, the additional performance in perception of depth magnitudes with stereoscopic perception seems to be dependent on monocular depth cues.

In Table 2.1, the binocular disparity is in plural form, because in addition to horizontal disparity, there are other disparities such as half-occlusions and vertical disparity. The horizontal disparity left and right views cannot be established from the areas that are visible only for other eye. These areas are called half-occlusions. These areas have special properties that can be used as sources of depth, as reviewed by Harris and Wilcox (2009).

Vertical disparity arises because objects appear to be of different sizes for retinas if they are not located in the middle of the visual axis. The size difference causes vertical, size and shear disparities (Howard and Rogers, 2012). Vertical disparity is also a source of absolute distance. However, within action space, which is the main focus distance range of the dissertation, the disparity is mainly horizontal. Within action space, the vertical size ratios are below 1.03 (Howard and Rogers, 2012). In this dissertation, therefore, binocular disparity means horizontal disparity, as the vertical disparity is diminished at distances beyond the personal space.

2.2.2 Panum's Fusional Range

A stereoscopic image is fused into a single image within Panum's fusional range, shown as the grey area in Figure 2.3. Within that range objects are perceived as single without diplopia (double vision). In Figure 2.3, objects P1 and P2 are perceived as single, but P4 is seen as diplopic. The Panum's fusional area is around the horopter, which is horizontally a curve. The horizontal empirical horopter deviates from the theoretical horizontal horopter, which is called the Vieth-Müller circle. The Vieth-Müller circle is formed with three points: the fixation point and the two nodal points of the eyes.

The empirical horopter has been found to be less convex than the Vieth-Müller circle (Schreiber et al., 2008). Two methods, the nonius and apparent fronto parallel plane (AFPP) methods, have been suggested for defining the empirical horopter. In the nonius method, the positions of partially occluded rods are adjusted to have same visual direction as the fixation point (Shipley et al., 1970).

In the AFPP method, the emprical horopter is based on the equidistance horopter, where distances that are perceived at equal distances from the observer form the horopter. Foley (1980) studied depth perception using AFPP method and showed that horizontal frontal distance has an effect on distance estimates in stereoscopic depth perception. Foley (1980) found that in a relative distance task, the user misevaluates the distances. The participants were asked to align points to lie at the AFPP, and the results showed that judgment errors increased as a function of horizontal frontal distance from the center. It was concluded that this increase in position errors arise from the distance evaluation of the center point. At greater distances (more than 1.8 m), the points were perceived to be farther away with respect to the center point (with an error in binocular disparity of approximately 1 arcmin at 10°). At shorter distances (less than 1.8 m), the points were perceived to be closer with respect to the center point. Moreover, the perceptual space is elliptical at smaller distances and is hyperbolic at greater distances.

The size of Panum's fusional range depends on the spatial frequency, eccentricity, sharpness, size, temporal frequency and movement of stimuli (Schor, 1993; Howard and Rogers, 2012). In the foveal area (within an eccentricity of 1°), the Panum's fusional range is approx. $\pm$ 5 arcmin (Schor and Wood, 1983). There is a hysteresis effect in stereoscopic fusion, which means objects are fused at larger disparities if they are initially fused. Panum's fusional range is usually less than a disparity of 1°; however, its length decreases as a function of spatial frequency up to 2 cycles/°, as illustrated in Figure 2.4. Temporal frequency also has an effect on Panum's fusional range. Stereoscopic fusion can occur even when the left and right views are not shown simultaneously. This fusion is possible with an interocular latency of 50 - 60 ms with exposure times of 10 – 25 ms (Ogle, 1963).

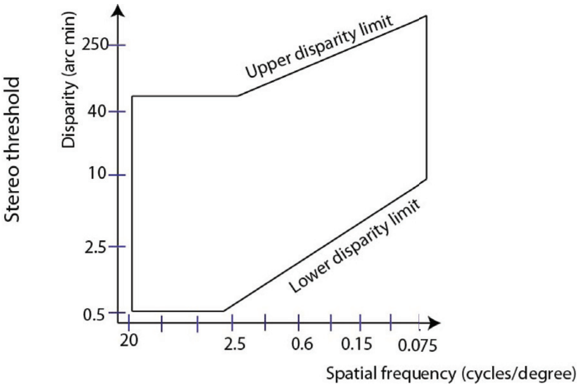


Figure 2.4. Stereo thresholds of stereoscopic vision (adapted from Schor and Wood (1983), reprinted with permission). The upper limit is determined by Panum's fusional range and the lower limit by stereo acuity (see Section 2.2.3). The upper limit increases when the spatial frequency decreases because the tolerance of stereoscopic vision for fusing low frequency images is higher. Additionally, the lower disparity limit increases when the spatial frequency decreases, as the threshold for stereoscopic depth increases.

Disparity gradient G depicts how rapidly the disparity φ changes as a function of angular separation θ. The linear horizontal disparity gradient between objects P2 and P3 is computed as the ratio between horizontal disparity $\varphi_{2,3}$ and average horizontal separation $\theta_{2,3}$ according to Equation 2.2:

$$G = \frac{\varphi_{2,3}}{\theta_{2,3}} = \frac{(\alpha_{2L} - \alpha_{2R}) - (\alpha_{3L} - \alpha_{3R})}{((\alpha_{2L} - \alpha_{2R}) + (\alpha_{3L} - \alpha_{3R}))/2}, \tag{2.2}$$

where α_{2L} and α_{2R} are the horizontal angles of P2, and α_{3L} and α_{3R} are the horizontal angles of P3, as shown in Figure 2.3 (Howard and Rogers, 2012).

If the disparity gradient between two fusible objects is greater than 1, the objects are seen as diplopic (Burt and Julesz, 1980). Thus, in Figure 2.3 object P3 is not observed as fused with object P2 if the disparity gradient exceeds 1. The area P2 forms is called the "forbidden zone". However, it is important to note that the forbidden zone is only a limit for binocular fusion. The stereoscopic perception with diplopic images is primarily qualitative, meaning that only ordinal depth can be perceived (Westheimer and Tanzman, 1956).

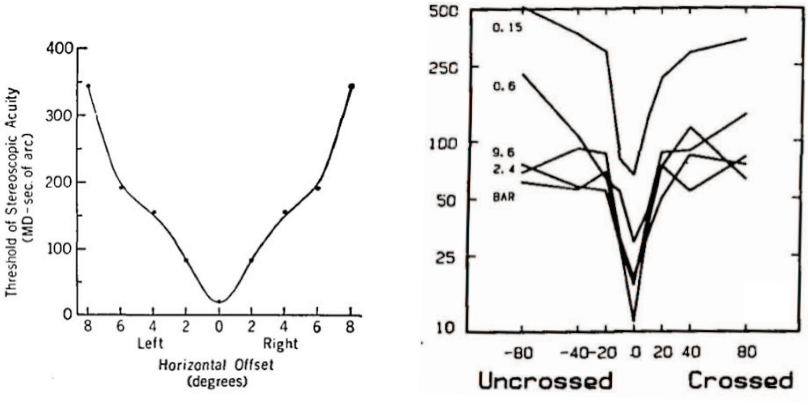
The properties of Panum's fusional range are addressed in PII and PIII, where the stereoscopic visualization approach is presented.

2.2.3 Stereo Acuity

With stereoscopic perception, the depth discrimination threshold is limited by stereo acuity. A widely used average disparity value for stereo acuity is 20 arcsec, but values between 3 – 60 arcsec have been presented in the literature depending on the viewing distance and type of experiment (Nagata, 1991). There are many different methods for testing stereo acuity (e.g., Frisby, Randot E, TNO, Titmus[1] and Howard-Dolman[2] tests), and tests can be used to measure local or global stereopsis. In local stereopsis tests, the targets are detectable from the monoscopic view, whereas in global stereopsis two views need to be fused to detect targets. The Frisby and Howard-Dolman tests are examples of local stereopsis tests. The Howard-Dolman apparatus is a physical box with two rods. The user is asked to align the position of the rods to be at the same distance from the user. In global stereo acuity tests, the images usually are fused from 2D images by using the polarization (e.g., Randot-tests) or anaglyph (e.g., TNO) methods.

In this dissertation, global stereopsis and stereo acuity were measured using the TNO test in every publication, and local stereopsis and stereo acuity were measured with the Howard-Dolman test in PII and PIII. In addition, stereo acuity through a stereoscopic system was measured with a method developed in PI.

Figure 2.5 shows stereo acuity (in arcsec) as a function of eccentricity (Rawlings and Shipley, 1969) and disparity (Badcock and Schor, 1985). Clearly, stereo acuity decreases according to eccentricity and disparity. Figure 2.5a shows how the threshold increases as a function of horizontal distance from fixation point. The stereo acuity also decreases (and depth threshold increases) as a function of vertical offset. The threshold is approximately doubled 30 arcmin away from the fixation point compared with the value at the very center of the fovea (McKee, 1983). Westheimer and McKee (1980) found that stereo acuity is deteriorated more than ordinary visual acuity when blurring was present.



(a) (b)

Figure 2.5. Threshold for stereo acuity (in arcsec) as a function of spatial dimension. (a) Along the eccentricity from the fixation point (in degrees) (Rawlings and Shipley, 1969), and (b) disparity (in arcmins) (Badcock and Schor, 1985). The numbers next to the curves represent spatial frequencies (see Figure 2.4).

Stereo acuity also changes as a function of spatial frequency, as illustrated in Figure 2.4, and luminance amplitude (contrast). When spatial frequency increases the depth threshold decreases (and stereo acuity increases) (Schor and Wood, 1983). The stereo acuity increases (and the depth threshold decreases) as a function of spatial frequency until approximately 2.4 cycles/°, after which it is saturated. The performance of stereo acuity increases as a function of contrast, but after 21 dB the effect of contrast on stereo acuity is saturated (Halpern and Blake, 1988).

---

[1] http://www.ophthalmologyweb.com/Pediatric-Ophthalmology/5649-Stereoscopic-Vision-and-Depth-Perception-Testing/ (accessed 24.10.2013)

[2] http://www.bernell.com/product/2013/126 (accessed 24.10.2013)

3. Depth Perception Through Stereoscopic Systems

3.1 Background

This chapter focuses on how human depth perception is affected when reality and computer graphics are viewed through a stereoscopic system that consists of the components illustrated in Figure 3.1. The depth perception through a stereoscopic system suffers from numerous error sources, which will be discussed in this chapter.

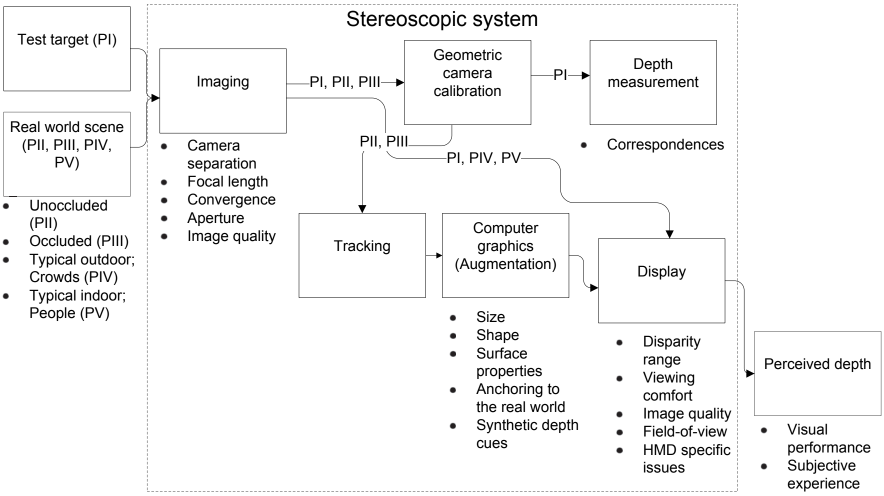


Figure 3.1. The stereoscopic systems of the dissertation (an IDEF0 type diagram); with mapping of the Publications PI - PV.

In augmented reality (AR) scenes, the real world scene is captured with a stereo camera and augmented with virtual objects according to the perspective of the observer. The position and orientation of the observer is tracked by tracking software, which uses the same video for tracking as that displayed to the observer using a handheld or head mounted display (HMD). These are the two distinct ways to display AR. Desktop displays are uncommon because they are usually viewed from fixed positions. Although mobility is not included in the definition of AR, it is usually an integral part. In AR, information is typically aligned with the real world in diverse viewing locations and situations. Thus far, handheld devices have typically been used in applications that require high mobility from the user, as the mobility of HMDs has been restricted.

See-through capability separates AR and VR displays. In VR, HMDs are opaque and do not have the see-through capability. The design of VR HMDs is different from AR HMDs, as VR HMDs aim for a larger field of view to achieve higher immersion (Cakmakci and Rolland, 2006). Moreover, VR HMDs typically occlude the space that lies outside the field of view of the display and inside the field of view of the human. VR HMDs are not typically designed to be light, as the applications do not require much mobility. In AR, video and optical see-through displays represent two different approaches. In a video see-through display, the scene is viewed through a video input, whereas in an optical see-through display, the scene is viewed through a semi-transparent lens. In a video see-through display, the real world and augmented graphical objects are at the same focal distance, and in an optical see-through display, the graphical objects and real world are at different focal distances.

The benefit of video see-through displays is that the rendering of scenes, e.g., latency and image quality, can be matched to real-world imagery (Rolland and Fuchs, 2000; Oskam et al., 2012; Xueting and Ogawa, 2013). However, the perception of the real world with a video see-through display is adversely affected by certain characteristics of cameras and displays, such as resolution limitations and optical distortions (Woods et al., 1993; Rolland and Fuchs, 2000; Takagi et al., 2000). In this dissertation, video see-through HMDs were used in PII and PIII. HMDs with still images were used in PI and PIV. In PV, an autostereoscopic display was used.

The type of HMD has a marked effect on depth perception, as the selection determines whether stereoscopic depth cues are available. The display can be either monocular (one camera and display shown to one eye), biocular (one camera and the same image shown to both eyes) or binocular (two cameras and a separate images shown to each eye). In this dissertation, the biocular displays are referred to as monoscopic or non-stereoscopic to avoid confusion with binocular displays.

Although binocular disparity is likely the most studied human capability for providing depth information about scenes, the combination of AR and stereoscopy has only been partially covered. Studies on stereoscopic AR depth perception within the action space have mainly used optical see-through displays and virtual objects on the ground plane (Jerome and Witmer, 2005; Swan II et al., 2007; Jones et al., 2008; Livingston, Zhuming and Decker, 2009; Grechkin et al., 2010). The errors have varied from 1 % to 37 % depending on the experiment. In addition to visual stimuli, the measurement protocol has a significant effect on depth judgments. For example, Jerome and Witmer (2005) studied depth perception within the action space (1.5 – 25 m) with optical see-through HMD and found a 37 % error regarding the actual distance of an augmented object with verbal reporting and 15 % error with a matching task. However, none of these studies compared non-stereoscopic and stereoscopic conditions. In VR, Willemsen et al. (2008) found no difference in depth judgments between non-stereoscopic and stereoscopic conditions when the objects were on the ground plane. This result cannot be generalized to X-ray visualization or to AR scenes where virtual objects are above the ground plane

The studies by Ellis and Menges (1998); Livingston et al. (2003) seem to be the only ones that compare non-stereoscopic and stereoscopic depth perception in X-ray visualization. Ellis and Menges (1998) investigated the effect of a partially occluding real-world object on depth judgments regarding a virtual object at near distances (< 2 m). Stereoscopic viewing improved perceptual matching accuracy when the virtual object was occluded. In addition, they found that an occluding surface could affect distance judgments by changing convergence. Livingston et al. (2003) used binocular disparity as one depth cue to visualize multiple occluded layers at longer distances of 60 m to 600 m. In that study, binocular disparity did not improve accuracy in distance judgments within the vista space.

AR applications are usually constructed by using one outward pointing camera, and the same image is shown to both eyes. AR systems suffer from rather poor depth perception within the action space. Wither and Hollerer (2005) found errors of 30 % regarding the actual distance when objects were aligned at distances from 38 to 65 m.

As for handheld systems, whether stereoscopic and non-stereoscopic, depth perception research has been limited. Dey et al. (2010) studied distances in a range of 70 – 117 m. The relative signed error was between 0 % and 30 % depending on the distance from the real-world reference point and the visualization conditions. In the latter study, Dey et al. (2012) investigated depth perception with a handheld mobile phone and tablet at distances from 20 to 120 m. They found that errors increase as a function of distance, so that for larger distances, the error was more than 50 % of the distance. They also discovered that the distances were underestimated less with the mobile phone. This result was ascribed to the smaller size of objects on the mobile phone screen compared to the tablet screen. Kerber et al. (2013) studied recently the depth threshold using a handheld autostereoscopic AR system. They examined the effect of binocular disparity and relative size on depth threshold and found that the relative size cue dominated the depth threshold, and binocular disparity did not decrease the depth threshold.

The depth perception through a display is affected by multiple factors, as compiled in Figure 3.1. The goal of this chapter is to identify the technological factors that affect depth perception stereoscopic systems. The factors are discussed in terms of the following categories: capturing, computing, augmenting and viewing.

3.2 Capturing

Stereoscopic capturing can have two main purposes: recording a scene for stereoscopic viewing and measuring depths within the scene. In video see-through displays, stereoscopic systems and humans can use the same stereo camera. The stereoscopic system uses the camera for measurement, and humans use it for perception. The use of stereoscopic video-see through systems also for measuring purposes has not been widely investigated (Ferrari et al., 2009). The capturing geometry has an effect on both purposes, but finding geometric properties (e.g., camera separation) for a stereoscopic system that fulfil the requirements from both standpoints has not been addressed.

Stereo cameras for capturing can be aligned either in a parallel (see Figure 3.2a) or toed-in (see Figure 3.2b) configuration (Woods et al., 1993). The parallel configuration uses a horizontal built-in sensor shift, or captured images are shifted afterward. In the toed-in method, the cameras are physically rotated toward each other. The parallel configuration is preferred because the vertical disparity due to keystone distortion and depth plane curvature can be avoided. If converging cameras are used, then displays should also be converged. This is rarely the case with HMDs (Takagi et al., 2000). The convergence distance is the point where the optical axes of the cameras intersect. The disparity at the convergence distance is zero.

3.2.1 Perceived Depth Threshold through a Stereoscopic System

The human depth threshold through a stereoscopic system dZ_{st} (see Figure 3.2) can be approximated according to Equation 3.1:

$$dZ_{st} = \frac{Z^2}{MNI} dp_{st}, \tag{3.1}$$

where M is the magnification between viewing and capturing, N is the ratio between camera separation and interpupillary distance, I is interpupillary distance and dp_{st} is the stereo acuity through the stereoscopic system (Jennings and Charman, 1994). Magnification M is computed as $M = ViewingFOV/CapturingFOV$. In orthostereoscopic systems, the errors in the capturing and display process are minimized. Thus $ViewingFOV = CapturingFOV$ and $IPD = b$ (parameters are shown in Figures 3.2a and 3.2d), and Equation 3.1 becomes the same as Equation 2.1 in natural viewing, except that the perceived depth threshold is affected by stereo acuity

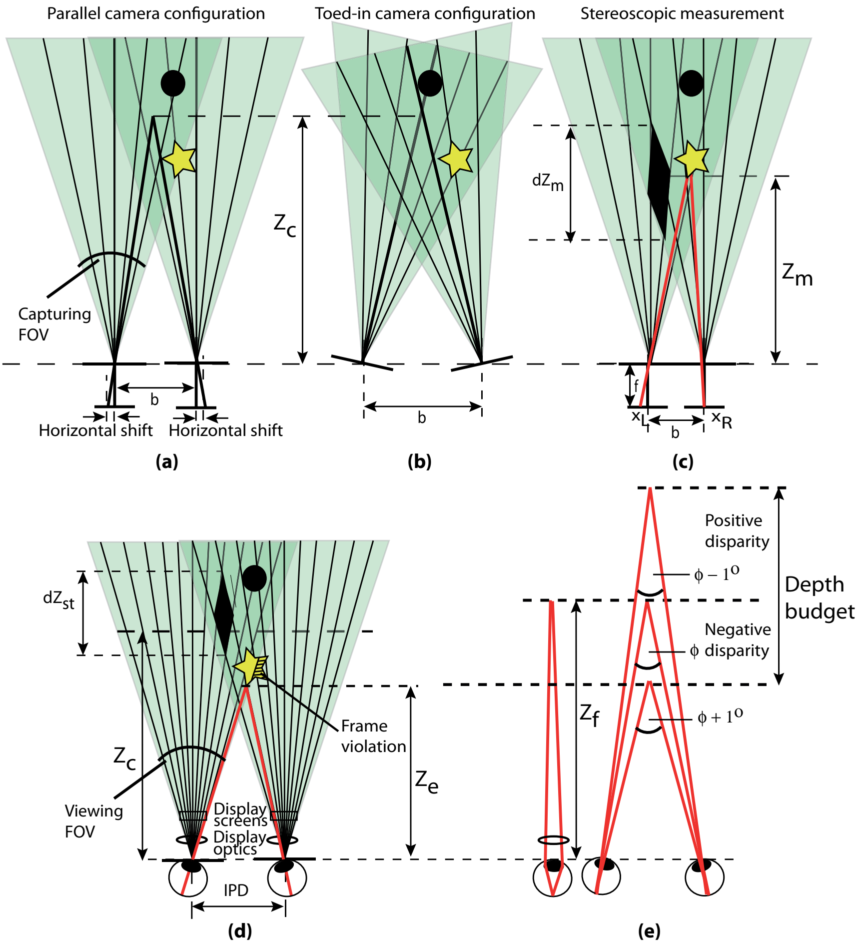


Figure 3.2. Two capturing configurations, parallel (a) and toed-in (b). At the convergence distance (Z_c), the disparity is zero. The stereo camera measurement is shown in (c), and the viewing geometry measurement is shown in (d). The depth budget of the stereoscopic display according to a disparity range of one degree is shown in (e). The accommodation-convergence mismatch occurs with stereoscopic displays. The convergence is at Z_e, but the focal distance is at Z_f.

through the stereoscopic system. The depth threshold through the stereoscopic system is limited by the angular resolution of the camera and the display. The overall depth threshold can be considered the higher value of the angular resolutions of the camera and display. For example if the FOV of the display and camera is 40°, the horizontal resolution of the display is 800 pixels and the horizontal resolution of the camera is 640 pixels, then the depth threshold through the stereoscopic system is limited by the angular resolution of the camera, which in this case is 3.75 arcmin. The angular resolution can be used as a limit for the depth threshold, but the actual depth threshold varies according to contrast.

3.2.2 Depth Measurement by Stereo Camera

The depth measurement by stereo camera Z_m in a parallel configuration (see Figure 3.2c) can be computed with Equation 3.2:

$$Z_m = \frac{bf}{p_x}, \tag{3.2}$$

where b is camera separation, f is focal length and p_x is disparity. Disparity can be computed as the difference between x-coordinates in the left and right images as $p_x = x_L - x_R$ (parameters shown in Figure 3.2c). To derive the dependence of the depth threshold of a stereo camera measurement dZ_m on disparity accuracy dp_x, Equation 3.2 is differentiated with respect to disparity p_x, leading to:

$$\frac{dZ_m}{dp_x} = \frac{-bf}{p_x^2}. \tag{3.3}$$

When Equation 3.3 is multiplied by dp_x and when p_x is substituded as from Equation 3.2, the following Equation 3.4 is obtained for the absolute theoretical depth threshold of a stereo camera measurement:

$$dZ_m = \frac{Z_m^2}{bf} dp_x. \tag{3.4}$$

The threshold is dependent on the capturing geometry and is linearly proportional to the disparity accuracy, which is dependent on the spatial resolution, as illustrated in Figure 3.2c. In addition, the stereo camera measurement is affected by calibration errors (see Section 3.3.1). Trade-offs exist between measurement and display purposes with the geometric factors of camera separation, focal length, convergence and aperture. The following sections make it evident that a trade-off exists between optimal camera settings for stereo camera measurement and human viewing.

3.2.3 Camera Separation

The cameras should be set far enough apart to achieve a small depth threshold. The theoretical depth threshold of human and stereo camera measurement are inversely proportional to the length of the camera separation[1]. According to Equation 3.1, the perceived depth threshold through a stereoscopic system is decreased as a function of camera separation, as was also empirically shown by Jennings and Charman (1994) at a viewing distance of 3 m. However, Rosenberg (1993) did not find this effect beyond a camera separation of 3 cm at a viewing distance of 0.8 m. Stringer (2003) studied the effect of increased camera separation on the precision of depth judgments at a viewing distance of 2 m. He found an improvement in precision, but it was not linear with increased disparity.

The disparity range increases as the camera separation increases, which increases the perceived depth magnitude (Ijsselsteijn, 2000). With a short camera separation, the objects may look like they are made of cardboard, which is called the "cardboard effect" (see PV). Increasing the camera separation also gives rise to a phenomenon called the “puppet-theater effect”, where objects look unnaturally small. The probability that foreground objects look unnaturally small increases as a function of camera separation (Yamanoue et al., 2006). In the puppet-theater effect, the magnification of near objects is less than that of far away objects. The puppet-theater effect is also called “gigantism”, as the world appears like with the IPD of a giant. Yamanoue et al. (2006) predicted that the degree of the puppet theater effect is proportional to the camera separation in a parallel configuration. In a toed-in configuration, the viewing distance, focal length and the depth of the object from the convergence point also influence the magnitude of the effect.

Camera separation has a wide variety of subjective effects on the perception of S3D content. If the camera separation is wider than the IPD of humans, a condition called hyperstereo arises. It is the preferred viewing condition when a magnification in disparities is desired. For example, with a stereoscopic range finder the distance between views is set to be much wider than the IPD (Barr and Stroud, 1924). However, the use of hyperstereo causes distortions in the perception of motion, which have been shown

---

[1] Also called the baseline, as in PI.

to be a benefit of stereoscopic perception when, for example, catching a ball (Servos et al., 1992; Mazyn et al., 2004, 2007) and judging a heading (Van den Berg and Brenner, 1994). In addition, using other camera separations than IPD can distort the enhanced perception of the glossiness of materials with stereoscopic perception (Sakano and Ando, 2010) as the rays of light are perceived differently.

As camera separation increases, the differences between the images increase, which makes crosstalk more visible. In crosstalk, the image that should be directed to only one eye leaks to the other eye. Crosstalk is perceived as ghosting in stereoscopic images (Woods, 2011). While ghosting has been shown to affect the viewing experience and comfort of stereoscopic displays (e.g., Kooi and Toet, 2004; Seuntiens et al., 2005), it also decreases judgments of depth magnitudes (Tsirlin et al., 2011).

In this dissertation, the effect of camera separation was evaluated within the context of both measurement and perception. The depth threshold by stereo camera measurement (dZ_m) and the perceived depth threshold through the stereoscopic system (dZ_{st}) were compared with a method that was developed in PI.

To understand how the viewing experience behaves with changes in camera separation, the model for stereoscopic visual experience developed by Seuntiens (2006) is used. In PV, subjective experience is measured using a Likert-scale that addresses viewing experience, strength of depth sensation and naturalness of depth sensation.

3.2.4 Focal Length

The theoretical perceived depth threshold, expressed in Equation 3.1, is inversely proportional to the focal length. However, increasing the focal length causes the cardboard effect, which makes objects look unnaturally thin (see PV). Yamanoue et al. (2006) predicted that the degree of this effect is proportional to the focal length in the parallel camera configuration.

Increasing the focal length reduces the field of view, which has been observed to influence depth judgments in certain situations (see Section 3.5.4).

The amount of geometric distortion in lenses typically increases when the focal length decreases. To the author's knowledge, the effect of lens distortions of the cameras on perceived depth has not been systematically researched. However, the effect of pincushion distortion of the display on the egocentric depth perception was studied, but Kuhl et al. (2009) did not find any effect of distortion on depth judgments. Klein and Murray (2008) proposed a method to match the lens distortions of live feed video and computer graphics, but the effect of lens distortions on depth perception was not investigated. Lens distortions are usually compensated for by polynoms. Radial distortion in particular has to be compensated because it causes vertical disparity at the edges of the images. However, accurate compensation is difficult in the case of low-level lenses.

The focal length in PI was varied, and its effect on the depth threshold of stereo camera measurement was found to be more dominant than that of camera separation.

3.2.5 Convergence

Two methods for implementing camera convergence are illustrated in Figures 3.2a and 3.2b. The convergence distance affects the location of the perceived disparity range. It is very important at close distances to reduce the convergence distance to limit the disparity range. Convergence can be adjusted dynamically by using software (Chen et al., 2010; Sherstyuk et al., 2012).

As the convergence distance increases, the probability of conflicts from frame violations increases. Frame violations occur if the disparity at the edge of the screen is negative (in front of the screen) and if the frame of the display occludes the region with negative disparity. Frame violation is illustrated in Figure 3.2d, where the star is cut by the frame. Based on frame violations, the convergence distance should be set so it is near.

The depth uncertainty of stereo camera measurement at the convergence point is large because the disparity there is zero (Mulligan et al., 2001). To reduce this problem, the resolution of the sensor can be higher in the middle to simulate the human foveation (Sahabi and Basu, 1996). For humans the most accurate stereo acuity is at the point of convergence (zero disparity) at the focal depth (Blakemore, 1970). The stereo acuity decreases rapidly as a function of disparity, as shown in Figure 2.5.

In this dissertation, the effect of the convergence distance on perceived depth was not examined. The convergence angle was adjusted to limit the disparity range within the limits for comfortable viewing. The parallel configuration was used for capturing in every study.

3.2.6 Aperture

Aperture affects the depth-of-field (DOF) of images. From a photogrammetric point of view, a small aperture (extended DOF) is used in stereoscopic systems to achieve overall sharp images. The aperture size is limited by diffraction. From a perceptual point of view, the focal distance of the eyes is at the screen plane and the DOF depends on the aperture of the camera. Thus, an overall sharp image without image blur from DOF of the camera, may cause distracting diplopic perception (Drascic and Milgram, 1996).

The effect of DOF on depth perception was not investigated in this dissertation. In AR and crowded scenes, the DOFs were large to maintain sharpness over the desired distance range. For typical indoor scenes, the DOF was varied between scenes.

3.3 Computing

3.3.1 Geometric Camera Calibration

Geometric camera calibration is a process in which intrinsic parameters, such as focal length and sensor size, and extrinsic parameters, such as camera separation and rotations between cameras, are computed. A detailed explanation of the camera calibration process is beyond the scope of the dissertation, but geometric calibration is a fundamental procedure that needs to be performed in AR systems. Camera calibration enables stereo camera measurements and tracking of the observers position. Inaccuracies in calibration cause errors in depth measurement (Zhao and Nandhakumar, 1996) and a misperception of the locations of objects (Livingston and Ai, 2008) and, thus, effects the perceived depth.

In this dissertation, standard camera calibration procedures are used, and image alignment and field-of-view matching between cameras and display are performed visually. Rolland et al. (1995) developed a procedure that can be used for aligning see-through world and computer graphics, which could be used in future studies.

3.3.2 Tracking

In AR, the camera movements can be tracked optically (Klein and Murray, 2007) and the augmentations are rendered from camera's perspective. The tracking parameters have mutual dependencies. For example, there is usually a trade-off between accuracy and speed with optical tracking. An increase in resolution slows down the frame rate of tracking, as there are more measuring points to be calculated. In addition to tracking computation, the frame rate of the cameras and display have an effect on latency.

Latency in AR systems deteriorates depth perception. With monoscopic displays, it has been shown that the latency of the system flattens the perceived depth magnitudes due to its effect on motion parallax (McCandless et al., 2000).

3.3.3 Correspondences for Measurement

Establishing correspondences between left and right views enables stereo camera measurement without a priori knowledge of the scene, while a non-stereoscopic system requires the integration of temporally and spatially separated images to measure the distances in a scene. There are two main approaches to matching the correspondences: feature-based and signal-based. In feature-based matching, features are extracted from the image and then matched based on the shortest Euclidian distance between the features. In signal-based matching, a window of one image is compared with the window of the other image. This allows for more dense matching, but finding correspondences with a good accuracy is difficult in real time.

In AR, stereo camera measurement can be used to measure the distances within a scene and augment the scene accordingly. In addition, stereo camera measurement can be utilized in handling occlusions between the observer and augmentations without an external depth sensor (Wloka and Anderson, 1995; Zhu et al., 2010). For detecting occlusions, only the order of the depths matter, as the occlusion does not offer a cue for the magnitude of depth.

3.4 Augmenting

3.4.1 Size of Augmented Objects

The real environment, observed through a display, should be as visible as possible to maintain the observers' awareness of the surrounding environment. The reason for this is that real world perception suffers from loss in screen space due to occlusions by the augmentations, and the observer's attention is divided between the real world and the augmentations. This issue must be taken into account when designing depth cueing in AR. For example, in a wayfinding application, the augmentations can be used to guide the observer to walk within the appropriate space. The augmented information should be provided with as little occlusion of the real world as possible to avoid a loss of perception of the real world. This principle should be applied to most applications in which the observer is guided and the visibility of the real world is essential for completing a task. This principle does not hold for diminished reality, where the real-world objects are removed from the field of view of the observer.

The ratio between seeing the real environment and the augmented graphics can be expressed as the see-through-graphics (CT-G) ratio. The CT-G ratio is analogous to the data-ink ratio design rule in information visualization, which applies to the amount of redundant information compared with data. Just as the data-ink ratio should be maximized (Tufte, 1983), the CT-G ratio should also be maximized.

The augmentations should be large enough to detect the relative size difference between them but small enough to avoid loss in screen space. The depth threshold ΔZ at depth (Z) can be computed according to Equation 3.5 (Nagata, 1991):

$$\Delta Z = \frac{Z}{\frac{S}{\Delta\theta Z} - 1}, \tag{3.5}$$

where S is physical size and $\Delta\Theta$ is acuity, by which the change of object size is detected. Thus, there is a clear trade-off between the detectability of relative size as a depth cue and loss in screen space. Equation 3.5 holds for the monoscopic condition. In the stereoscopic condition, the effect of relative size cue can be expected to be diminished as the binocular disparity exists as a relative depth cue between augmentations.

Rolland et al. (2002) did not find an effect of size on a decreased depth threshold, but they conducted the comparison using two different objects and sizes at a time, and thus, the relative size between objects was not available. Witha larger size, the shape of the augmented object is more easily detected.

3.4.2 Shape of Augmented Objects

It can be assumed that 3D augmentations give more information about depth than 2D objects because their shapes and structures are available for interpretation. The shading of 3D augmentations creates depth cues that can facilitate depth perception. Rolland et al. (2002) did not find a significant effect of 3D augmentation shapes on the depth threshold using cubes, octahedrons and cylinders. In addition, Rolland et al. (2002) studied the effect of shading on the depth threshold by comparing smooth and faceted cylinders and found that the structures in a faceted cylinder did not decrease the depth threshold.

3.4.3 Surface Properties of Augmented Objects

Transparency
The use of semi-transparency is an important property in AR, where the augmentations are overlaid on the real world. To keep the CT-G-ratio (see Section 3.4.1) as high as possible, it is important to understand the effect of the transparency of the augmentations on depth judgements. The augmentations should occlude the real world as little as possible so the observer is able to interact with the surrounding environment. Livingston et al. (2003) studied the effect of semi-transparency in the case of X-ray visualization. They found that increasing transparency improved depth judgments under wire frame and fill visualization conditions. They also discovered that with a constant semi-transparency of the planes, stereoscopic viewing improved depth judgments.

Color
To the author's knowledge, the effect of the color of augmentations on depth judgments has not been studied, even though it can be expected that it will have influence based on Section 2.1. Gabbard et al. (2010) studied the discernibility of computer graphics against typical background materials in an optical see-through system. The effect of background color on see-through color was measured; however no depth perception studies were conducted. In this dissertation, similar colors were used for the augmentations to avoid influences from the chromostereopsis effect (see Section 2.1).

Texture
Hou and Milgram (2000) found that the texture density of real world objects has an influence on the depth threshold in a pointing task. A denser texture yields more accurate results. However, studies evaluating the effect of the texture of augmentations have not been found.

3.4.4 Anchoring Augmented Objects to the Real Environment

Anchoring the augmentations to real-world structures can be performed using occlusion or cues related to lighting.

Handling occlusion between augmentations and the real world requires a 3D reconstruction of the real world. Dynamic occlusion handling requires active measurement of the depth between the observer and the augmentation, which is difficult to perform accurately (see Section 3.3.3).

Cast shadows have been shown to be efficient tools for visualizing the spatial relationships of objects in computer graphics (Wanger et al., 1992; Hubona et al., 1999) and in AR (Sugano et al., 2003; Jurgens et al., 2006). Cast shadows have been shown to be important in the positioning task (Sugano et al., 2003; Jurgens et al., 2006), but for veridical perception, they require that the lighting conditions be consistent with the real world. An augmented light source position that is different from that in the real world can be expected to distract depth perception (Kruijff et al., 2010).

With objects floating in the air, the position of an object can be deduced based on the height at which the shadow hits the ground (Hubona et al., 1999). Casting shadows on the ground plane are very sensitive to the correct measurement of the ground level. For example, if the ground level is measured as being lower than it really is, objects are perceived to be closer than they really are based on the position where the shadow hits the actual ground level.

3.4.5 Synthetic Depth Cues

Synthetic depth cues are different from natural depth cues in the sense that they require adding depth cues to the scene that are unrelated to the natural properties of scenes. Such depth cues are needed especially in X-ray visualization.

X-ray Visualization
Depth interpretation in challenging AR visualization cases has been extensively investigated (Kalkofen et al., 2011). One example of a demanding visualization case is showing the observer occluded information that would not be possible in the real world. This scenario is called ”X-ray vision" (Feiner et al., 1995). Seeing through walls represents a new concept in visualization that makes the interpretation the depths difficult. Different approaches have been developed to aid in the interpretation of depths of virtual objects (Livingston et al., 2013). For example, Furmanski et al. (2002) studied the effect of motion parallax and partial occlusion on visualizing obscured information. Bane and Hollerer (2004) created interactive tools for visualizing obscured information with tunnel cut-out visualization, as illustrated in Figure 3.3a. Tsuda et al. (2005) used five different techniques to visualize obscured information, including ground grids. Avery et al. (2008) built a system to evaluate see-through vision outdoors. In an improved version, Avery et al. (2009) used multiple viewpoints to enhance depth interpretation using an edge overlay (see Figure 3.3b). In addition, Dey et al. (2010) compared ”Melt”-visualization (Figure 3.3c) with the edge information of an occluding object and showed that Melt-visualization outperformed this alternative. More recently, Sandor et al. (2010) did not find an advantage of saliency information in addition to the edges of an occluding surface for visualization with the X-ray approach.

(a) **(b)** **(c)**

Figure 3.3. Examples of different approaches for X-ray visualization. A tunnel cut out (a) (Bane and Hollerer, 2004), edge overlay (b) (Avery et al., 2009) and melt volume (c) (Dey et al., 2010)

Perspective Cues
In addition, synthetic depth cues can be used to visualize objects that are not behind a wall. Livingston, Zhuming and Decker (2009) used a linear perspective cue by overlaying tram lines (similar to the lines shown in Figure 3.3a) on the ground plane to facilitate depth judgments. Wither and

Hollerer (2005) used grid planes to provide perspective cues for linear perspective and foreshortening. Using an artificial grid plane rather than the real ground surface most likely reduced the dominance of shadows in the depth perception task (Wither and Hollerer, 2005).

The viewing point can also be changed in AR applications. Using a radar-like presentation for the locations of augmented objects simultaneously while viewing the scene is a popular technique in the vista space (Wither and Hollerer, 2005).

Aerial Perspective Cues
Artificial depth cues that mimic the aerial perspective are based on the decreased intensity and saturation of objects. An example of the aerial perspective cue is proximity luminance, which is an artificial depth cue that is based on a decreased intensity of edges according to distance (Dosher et al., 1986). The closer the edge is to observer, the more intense the object appears to the observer. Livingston et al. (2003) revealed that decreasing intensity is an efficient ordinal depth cue within the vista space. However, studies on the effect of aerial perspective cues within the action space are lacking. Most likely, these cues can also be used to visualize ordinal depths within the action space.

In addition to decreased intensity, the aerial perspective reduces color saturation. Wither and Hollerer (2005) studied the effect of color labelling and color saturation as a synthetic depth cue and found it useful for visualizing relative depths in the vista space.

3.5 Viewing

3.5.1 Disparity Range

The disparity range is the length between the minimum and maximum disparities on the retina. In addition to the scene, the viewing geometry affects the disparity range. The disparity range depends on the screen size, viewing distance and the IPD of the observer. When the screen size increases and the viewing distance decreases, the disparity range increases. The same disparity range on a display perceived with a shorter IPD causes greater disparities in perception. Thus, children perceive longer disparity ranges than adults. The effect of viewing geometry on the perception of stereoscopic displays was simulated in a study by Woods et al. (1993). Livingston, Zhuming and Decker (2009) showed experimentally that the perceived disparity is a linear function of IPD.

However, the disparity range and depth magnitude do not follow the clear relationship expressed in Equation 2.1. The estimated egocentric depth to the objects affects the transformation of the disparity to an exocentric depth magnitude (see Section 2.2.1). Monocular cues are used for egocentric depth estimation; they have an effect on how much the perceived depth magnitudes depend on disparity. In virtual reality, for example, Willemsen et al. (2008) found no effect of matching virtual camera separation to the IPD of the observers on egocentric depth magnitude when the objects were on the ground plane.

3.5.2 Viewing Comfort

In stereoscopic applications, the two images should be aligned carefully horizontally and vertically. Limiting the horizontal disparity range is essential for visual comfort (Lambooij et al., 2009; Shibata et al., 2011) to avoid an accommodation-convergence-mismatch. In stereoscopic displays, the accommodation is on the screen plane, but the convergence varies according to the depth of the object, i.e., the gaze point, as illustrated in Figure 3.2e). Different approaches to solving accommodation-convergence mismatch have been presented, such as liquid lenses that allow variable focal distances (see a review by Urey et al. (2011)).

The comfortable disparity range varies between humans, but for most cases, a horizontal disparity selected according to the one degree rule can be expected to be small enough for comfortable viewing (Lambooij et al., 2009; Shibata et al., 2011). In diopters, the mentioned accommodation convergence mismatch rule is ± 0.3D (Masaoka et al., 2006), which equals $\pm\ 1.1^\circ$ when the IPD is 6.5 cm. In this dissertation, the disparity ranges were mainly within $\pm 1^\circ$ (see Table 4.1), with exceptions in PV, in which the disparity of 1° was exceeded. Excessive disparity means disparities that cannot be fused without visual discomfort. The length of the comfort depth range on a display is called the depth budget, as illustrated in Figure 3.2e).

Another cause of reduced viewing comfort is vertical disparity. Vertical disparity results from misalignments of cameras, inaccurate geometric camera calibration and misalignments of display components. Limiting vertical disparity avoids the fusion problem and increases viewing comfort. Ogle (1952) showed that the relative depths of two objects could be detected with a vertical offset of 25 arcmin and that the limit increases with larger displays (Stevenson and Schor, 1997). For comfortable viewing, the limit for the vertical offset is from 3 arcmin to 6 arcmin (Melzer et al., 2009). In this dissertation, the vertical disparity of the images was removed as much as possible by visually aligning the images. The vertical disparities were not measured.

3.5.3 Image Quality

In the context of stereoscopic AR systems, image quality is investigated with low-level attributes that depend on the imaging sensors, display components and rendering properties of the augmentations, and they have an effect on depth perception. The rendering of virtual objects is usually nonphotorealistic. Under restricted conditions, however, the rendering of augmented objects can be close to the quality of the representation of the surrounding real environment (Kán and Kaufmann, 2012). The discrepancy between the image quality of the augmentations and the surrounding real-world environment affects the perceived depth, and it is especially an issue with optical see-through displays (Jerome and Witmer, 2005; Grechkin et al., 2010). With video see-through displays, the quality differences between augmented objects and the surrounding world can be reduced by filtering the real world capturing to have a similar appearance as the augmentations (Fischer et al., 2006; Klein and Murray, 2008). The low-level image quality attributes presented here are sharpness, brightness, contrast, noise and colors.

Offering the same image quality for both views in stereoscopic systems is also an important issue. The quality differences between two views cause binocular rivalry. This is a problem especially with video see-through systems, where differences between the left and right camera views are present when there are differences in the capturing conditions. Different illuminations (color temperature, luminance) in the left and right cameras cause differences in the color and brightness of the images.

Sharpness

An increase in the angular resolution of the camera or display increases the stereo acuity until theoretical human stereo acuity is achieved. The depth threshold decreases as a function of stereo acuity based on Equation 3.1, but the exact relation between angular resolution and the depth threshold is still unknown in stereoscopic systems. When perception of the scene is based on binocular disparity and monocular cues are lacking, then the sharpness has a greater effect on the depth threshold (Utsumi et al., 1994). When enough monocular cues are present, the resolution does not influence the depth perception of the egocentric depth magnitude (Thompson et al., 2004).

Brightness

Egusa (1981) showed that brightness has an effect on depth perception. The brightness of the display is especially an issue with optical see-through displays when used outdoors (Livingston et al., 2003; Livingston, Zhuming and Decker, 2009), which may be one explanation for the differences in depth judgments conducted indoors and outdoors (Livingston, Zhuming and Decker, 2009).

Contrast

Contrast is defined as the ratio between foreground and background luminance. Contrast improves stereo acuity, but its effect after 21 dB is saturated (Halpern and Blake, 1988). It has been shown that an area that has higher contrast than the background appears to be closer (O'Shea et al., 1994) and low contrast can lead to overestimations of depths (Drascic and Milgram, 1991). In this dissertation, the effect of contrast on depth judgments cannot be evaluated; we are unable to measure the contrast of HMD displays.

Color

The effect of color on depth perception is presented among depth cues (see Section 2.1). Egusa (1981) showed that saturation affects depth perception. Highly saturated images are difficult to achieve especially with optical see-through displays (Kruijff et al., 2010). The reduced saturation may cause the object to appear to be farther away due to the human ability to use atmospheric haze as a depth cue.

Noise

Noise has been shown to influence the viewing experience and naturalness of stereoscopic images (Seuntiens et al., 2005). In addition, noise has been added to random dot stereograms to study its effect on stereoscopic fusion (Ditzinger et al., 2000). However, no studies have been found where the effect of visual noise on depth judgments would have been evaluated. Presumably, noise affects perceived resolution, which affects the perception of texture density.

High-level image quality attributes

High-level image quality is a complex issue, and its attributes cannot be measured from technical low-level image quality attributes. For example, naturalness can be considered a high-level image quality attribute, as it depends on the subjective preference of the color of grass or the sky (de Ridder et al., 1995). There are numerous image quality models for 2D images (e.g., Engeldrum, 1999), but the attributes for quantifying the viewing experience with S3D images are more subjective and abstract than for 2D images.

Seuntiens (2006) has developed a viewing experience model for stereoscopic content. In Seuntiens' model, the stereoscopic depth is expressed as one attribute for naturalness. The strength of depth sensation is one attribute for depth, and it has been shown that it correlates well with the extent of the disparity range. There are limits for disparity ranges, but more detailed information about subjective quality of stereoscopic depth is still lacking. The roundness factor has been suggested for describing of the flattening of objects in S3D images (Yamanoue et al., 2006).

However, it is well-known that the viewing experience of stereoscopic media is not captured by geometric factors alone, which makes its definition and research challenging.

The Seuntiens (2006) model is interesting because it raises the "naturalness" attribute above "image quality." In 2D image quality, naturalness is considered to be one attribute of image quality (De Ridder and Endrikhovski, 2002), not the other way around. The naturalness attribute can be expected to be important in S3D images, because unnatural phenomena (such as cardboard and puppet-theater effects) that do not appear in 2D images can occur.

Naturalness may be linked to the “life-like” experience, which is commonly mentioned when describing S3D images (Häkkinen et al., 2008).

In this dissertation, the high-level attributes related to stereoscopic depth are discussed within the context of photography.

3.5.4 Field of view (FOV)

FOV has an effect on the magnitude of perceived depth. The FOV of HMDs is usually less than 50° in the vertical direction, which limits the ground perception. Different results have been obtained regarding the effect of FOV on egocentric depth judgments depending on the viewing conditions. Wu et al. (2004) found that with limited FOV, the visibility of the ground is lost at close distances, which affects depth judgments. The underestimation of exocentric and egocentric depths was evident when the head was kept still. However, Knapp and Loomis (2003) did not discover this effect on egocentric depth when participants were allowed to move their head freely. It has been noted that when participants are able to move the head from down to up (from near to far) a limited FOV does not affect depth judgments (Wu et al., 2004; Creem-Regehr et al., 2005). This result emphasizes the importance of directing the gaze to the ground when the FOV is limited.

In this dissertation, the magnitude of exocentric depth was evaluated in PIII, where the vertical FOV was 18°. This restricted FOV has most likely influenced depth judgments, as in the study by Wu et al. (2004), who found decreased egocentric depth judgments when the vertical FOV was below 21°.

3.5.5 HMD Specific Issues

Display Type

The display type influences the mass and inertia of the display. This is relevant from the depth perception point of view, as mass and inertia have been shown to decrease egocentric depth judgments (Willemsen et al., 2004, 2009). The effect of mass and inertia on depth judgments is not studied in this dissertation, although it would be interesting to examine whether if the mechanics also affect exocentric depth judgments.

Focal Distance

The focal distance to the virtual display is the level at which the eye is accommodated, as illustrated in Figure 3.2e. With video see-through displays, the focal distance is on the screen for both real world imagery and augmented computer graphics. With optical see-through systems, the focal distance varies between the real world imagery and the screen focal distance at which the augmentations are located. In stereoscopic displays with a constant focal distance, the accommodation-convergence mismatch occurs. Its effect on distance perception is not well understood (Willemsen et al., 2008). Focusing cues from accommodation (blurring and muscle activity) affect slant estimates at short distances below 1 m (Watt et al., 2005), but the effect within the action space can be expected to be diminished (Nagata, 1991; Cutting, 1997).

4. Studies on Depth Perception within the Action Space

4.1 Introduction

The previous chapter makes it evident that the perceived depth is affected by multiple sources when viewed through a stereoscopic system. According to published studies, the differences in depth perception between monoscopic and stereoscopic viewing appear to be minor when enough pictorial depth cues are available. However, in situations where pictorial cues are limited, stereoscopic viewing can be expected to improve performance, as measured with the depth threshold, depth magnitude and detection accuracy. The performance time may also be decreased in tasks due to decreased uncertainty.

This chapter gives an overview of the experimental research of the dissertation. The structure of this chapter follows the framework shown in Figure 1.3. For the experimental studies, Table 4.1 extends Table 1.1 and lists the publications and objectives and summarizes the scene, imaging and viewing geometries of the experiments conducted as part of the studies reported in PI - PV. More detailed information of the independent and dependent variables in the experiments can be found in the respective sub-sections.

Table 4.1. Summary of the experiments.

Publication	Depth range of target(s) in a scene (m)	Camera separation (cm)	Camera convergence distance (m)	Stimuli disparity range (arcmin)	Camera device (Resolution per eye)	Display device (Resolution per eye)
I	2 - 6	7 - 21	2 - 6	-55 - 0	Self-mounted stereo camera (1280 × 1024)	Non-see-through HMD (800 × 600)
II	6 - 10	6	6	≈-5 - 22	Native stereo camera (568 × 424)	Video see-through HMD (800 × 600)
III	1.7 - 3.3	6	2.5	-34 - 20	Self-mounted stereo camera (377 × 283)	Video see-through HMD (320 × 480)
IV	≈ 4 - 50	7.7	≈ 6	≈ -42 - 36	Point-and-shoot stereo camera (3648 × 2736)	Non-see-through HMD (800 × 600)
V	0.7 - 5	2 - 10	1.2 - 1.5	-156 - 108	SLRs in a self-built stereo rig (5616 × 3744)	Autostereoscopic (960 × 1200)

4.2 Method for Comparing the Depth Threshold of Humans and Stereo Camera Measurements (PI)

4.2.1 Background and Objectives

In AR, stereoscopic cameras can be used for measuring the depths of a real world scene, and augmentations can then be attached to the measured depths (Ferrari et al., 2009). In this scenario, the depth thresholds requirements for measurements arise from the perceived depth thresholds through a stereoscopic system. The depth threshold of a stereo camera measurement should be smaller than the perceived depth threshold to utilize the entire potential of a stereoscopic system.

The first subobjective of this study was to develop a method that enables determination of the depth threshold of the stereo camera measurement and human stereoscopic perception (presented as Objective 1 in Section 1.4). Existing methods for measuring depth threshold were introduced in Section 1.3.2.

The second subobjective was to determine the effect of camera separation on depth thresholds through stereoscopic systems (presented as Objective 2 in Section 1.4). The theoretical perceived depth threshold through a stereoscopic system can be computed with Equation 3.1 and the theoretical depth threshold of stereo camera measurement can be computed with Equation 3.4. In practice, however, the stereoscopic measurement is also affected by calibration (see Section 3.3.1) and the finding of correspondence points (see Section 3.3.3). The human depth threshold is affected by the depth cues of the scene (see Section 2.1) and image quality (see Section 3.5.3).

4.2.2 Test Target

We designed a test target for depth threshold measurements. The requirements for the test target were derived from characteristics of both stereo camera measurement and human stereoscopic perception through stereoscopic systems (summarized in Table 4.2). The test target can be a part of the development process of stereoscopic AR systems, as depicted in Figure 1.5.

From the perceptual point of view, perceived depth threshold tasks commonly involve pairwise comparisons (see Section 1.3.2). An example of a device that allows a pairwise comparison is the Howard-Dolman device (Howard, 1919), which can be used to evaluate the perceived depth threshold in video see-through systems, but it is unsuitable for testing the depth threshold of stereo camera measurement due to the small number of possible measuring points of the movable rods. From the measurement standpoint, depth thresholds have been determined using a wide variety of targets, such as cars and heads.

To enable exocentric depth measurements, the test target has to have objects that are related to each other at different depths. Another option is to incrementally change the distance from a planar test target and to record the response of the stereo camera. However, the latter procedure is much slower because measurements need to be made at many depths to obtain the same data as with a test target with one distance between the camera and target.

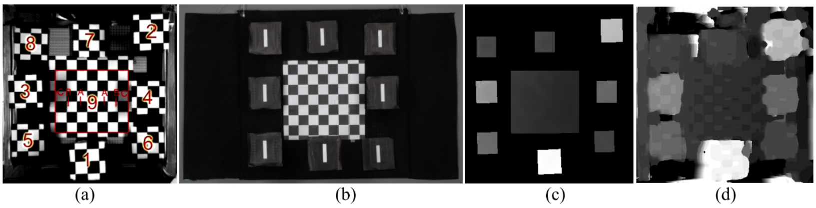
The test target (see Figures 4.1a and 4.1b) for this study was built of pieces with different depth levels mounted on a planar surface (500 mm × 500 mm). The depths of the levels were 206, 150, 93, 75, 54, 35, 17 and 6 mm (measured with measuring tape from the zero level, error ± 1 mm). The depths of the levels can be adjusted based on the depth threshold requirements of the applications of interest. The more accurate the stereo camera or the closer the working range, the smaller the depth differences can be between the levels. The constructed test target allows the mounting of pieces at eight depths. The depth range of the levels for this study was chosen to cover the

Table 4.2. Requirements for the test target arising from stereo camera depth measurement and stereoscopic vision.

Requirements from stereo camera measurement	Requirements from stereoscopic vision
• Levels at different depths • No occlusions: all depth measuring points have to be seen by both cameras • Geometrically well-known and simple: this enables comparison of stereo correspondence algorithms. Result of one correspondence algorithm is shown in Figure 4.1d and ground truth depth map is shown in Figure 4.1c). • The effect of different textures on measuring accuracy should be possible • The relative movements and rotations of the cameras have to be observable from the measurements: there has to be measuring points in both sides of the stereo rig's x- and y-axis to detect errors in calibration and alignment. • Accurate depth measurements should be possible: possibility to achieve subpixel disparity accuracy • High precision: enough measuring points	• Levels at different depths • Minimal or no monocular depth cues • Random ordering of the levels of the test target should be possible • Modular test target: the distances between levels should be adjustable • The levels of the test target can be mixed randomly to test depth threshold

theoretical perceived depth threshold range with the system computed according to Equation 3.1, using a disparity accuracy of 0.5 pixel. The results are plotted in Figure 4.2.

Figure 4.1a shows the target to be used for stereo camera measurements, and Figure 4.1b depicts the test target to be used in the perceptual evaluations. They are otherwise similar, but the patches on top of the levels in the latter case are white vertical bars. For measurements, a checkerboard pattern was chosen to enable an accurate locating of correspondences. For perceptual evaluations, white vertical bars in front of a black background were chosen because the disparity can be easily detected from vertical lines with high contrast. The depths of the target can be mixed randomly, and test participants can be asked to arrange them in the correct order according to depth.



(a) (b) (c) (d)

Figure 4.1. The test target for stereo stereo camera measurements (a) and perceptual tests (b). "Near" ground truth depth map from corner extraction (c) and a depth map from a correspondence algorithm (d). Levels 1-8, shown in (a), are in descending order from the zero level, marked with 9. The order of levels is the same in (a)-(d). In this case, $N = 8$, which enables 28 depth comparisons with one stimulus.

The area of the patches was 80 mm × 80 mm, the size of each checkerboard square was 30 mm and the length of each vertical bar was 60 mm. For perceptual evaluations, the depth cues from cast shadows should be minimized. To achieve this, the test target was uniformly illuminated with diffusion lighting, and it was rounded using a black curtain. Shadows were not present, as seen from the background in Figure 4.1b. The test target can be used within the action space, and the pictorial depth cues are restricted beyond a distance of 2 m. The relative size difference between bars is below

3 arcmin.

One of the benefits of a test target compared with a planar surface is its independence from a highly accurate ground truth distance between the stereo camera and target. For example, it does not matter if the measuring distance to the test target is 1999 mm or 2000 mm because the measurements are exocentric and relative to the zero level.

4.2.3 Measures

The measure for the depth threshold in this study was derived from the rank order method suggested by Thurstone (1931). He developed the method and found that very similar results can be achieved with the paired comparison and rank order methods. Applying the rank order method in acuity tests allows is a much faster way to measure depth threshold, as the number of depth comparisons with N depth is $N(N-1)/2$ with one trial. Pairwise comparisons are more laborious to conduct, as only one comparison is performed with one trial. The analysis of ranking data is conducted using JND. The 75 % level is selected for a limit of one JND. Depths that are distinguished at a lower correct rate than JND are interpreted to be at the same depth. The smallest correct ordering of depths is used as the measure of the depth threshold. We performed this task by computing the largest incorrectly ranked depth, and the depth of the next least correct rank was used as a value for the depth threshold.

In PI, the depth threshold of the stereo camera measurement was based on largest erroneous classification of depths using an average value, which is not actually a valid measure for the depth threshold because then the depth threshold is actually beyond the erroneous classified levels. Thus, we changed the measure to the smallest correct classification of levels. The depth threshold is the limit at which two depths can be correctly resolved.

Typically, the precision of stereo camera measurements is evaluated with the standard deviation calculated over all measurements. According to our definition for the depth threshold, using the standard deviation of all measurements is an insufficient measure. Our definition of the depth threshold is the least difference in depth between objects that can be correctly resolved. Calculating one standard deviation of all measurements does not take into account this issue. Stereo camera measurements are affected by calibration errors that have an influence on accuracy. Stereo cameras viewed by humans should fulfil the requirements for precision and accuracy arising from observers' visual capabilities. Thus, we compared the means and standard deviations of each depth measurement. We computed the largest incorrectly ranked depth, and the depth of the next least correct rank was used as a value for the depth threshold.

4.2.4 Materials and Methods

In the experiment, the depth threshold of the stereo camera measurement was compared with perceived depth threshold by varying the camera separation and distance. First, we measured the effect of camera separation on the depth threshold of stereo camera measurement. Second, the effect of increased camera separation on the human depth threshold was measured. The variables of the experiment are summarized in Table 4.3.

Each stereo camera was built of two board cameras. The cameras are small (approximately 30 mm for every dimension) and can be worn on the head. The board cameras were mounted on a bar, which allowed three camera separations: 7, 14 and 21 cm. The camera separations were derived from the anthropometry of the head.

The 7 cm camera separation represents the interpupillary distance (slightly above the average value, which is 6.3 cm (Dodgson, 2004)). The 14 cm camera separation represents the average width of the head (Poston, 2000, p. 73) and 21 cm was chosen to find whether it improves depth measurement accuracy.

The images for the stereo camera measurements were made at different distances, ranging from 0.7 m to 5.8 m. The distances within the personal space (below 2 m) were included to study the effect of distance on depth threshold of the stereo camera measurement.

The illumination was 1000 lx. This illumination level was chosen to reduce the noise and the effect of low contrast on depth judgments. The cameras were controlled by a laptop computer so that the exposure time and gain were kept constant. The corners of the checkerboard pattern were extracted using a Harris-corner finder (Harris and Stephens, 1988).

The images for the perceptual evaluations were taken with the same board cameras and under the same lighting conditions as with the stereo camera measurements. The images were taken at 2 - 6 m distances with a one meter interval. Between the distances, the order of the depths was mixed using a Latin square, so that their order was different at every distance and so each depth occurred in one location not more than once.

Table 4.3. The independent and dependent variables of the study for PI.

Independent variables	**N**	**Description**
Participant	8	Random variable (one female). The age span was from 22 to 26 years.
Camera separation	3	7, 14, 21 (in centimeters)
Images at different distances	≈11	The images for the stereo camera measurements were between 0.7 m - 5.6 m, using an approx. 0.4 m interval, resulting in approximately 11 images. The number of distances varied slightly between camera separations.
Images at different distances	5	For perceptual evaluations the distances were from 2 m to 6 m with a one meter interval, resulting in five images.
Dependent variables	**N**	**Description**
Depth threshold of stereo camera measurement	≈33	In millimeters. Measured as the smallest correct ordering of levels. Standard deviation was used as the threshold criterion.
Perceived depth threshold	15	In millimeters. Measured as the smallest correct ordering of levels. A correct rate of 75 % was used as the threshold criterion.

The task for the participants was to sort the levels according to descending order of depth. The participants verbally reported the vertical and horizontal locations of the level (for example up-left, middle-right, down-middle) to the experimenter, who wrote down the judgment. After all the depths were ranked, a new image was shown to the participant. The images were viewed in order of distance. Stereoscopic images were viewed using a stereoscopic HMD (horizontal FOV of 32° and 800 × 600 resolution) in a typical office room.

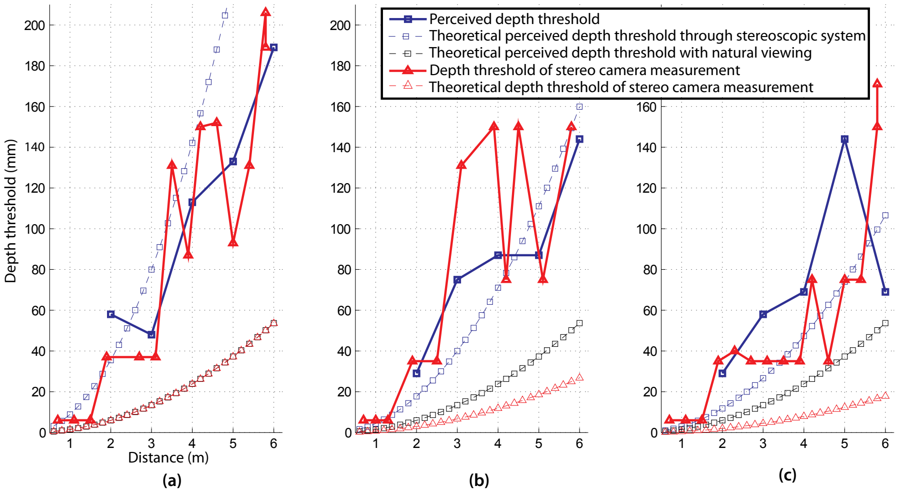
4.2.5 Results

The results in Figure 4.2 show that increasing the camera separation improves the depth threshold on average. However, the camera separation affected the depth threshold of the stereo camera measurements less than predicted by Equation 3.4. This indicates that the simple geometric theory underlying the equation is insufficient to predict the depth threshold of practical stereo camera depth measurements. Neither the theoretical depth threshold of the stereo camera measurements nor the theoretical depth threshold of human with natural viewing were achieved within the action space (i.e., longer than a two meter distance).

In practice, however, the perceived threshold through stereoscopic systems sets the requirement for stereo camera measurement. In this case, the depth thresholds of the stereo camera measurement are generally equal to the depth thresholds through stereoscopic systems. In this case, therefore, the depth thresholds of stereo camera measurements are small enough for stereoscopic systems.

In this experiment, the perceived depth threshold was increased by a factor of approximately 3 compared with theoretical natural viewing. The theoretical depth threshold computed with a disparity accuracy of 0.5 pixels is close to the perceived depth thresholds. With a 7 cm baseline, the perceived thresholds are mainly below the theoretical perceived thresholds, and with a 21 cm baseline the perceived thresholds are mainly above the theoretical perceived thresholds. With a 14 cm baseline, they are generally equal.

In PII and PIII the depth threshold was increased by a factor of approximately 7 compared with natural viewing when measured with the Howard-Dolman apparatus. These depth thresholds correspond to disparity accuracy of approximately one pixel. Achieving such a depth threshold of stereo camera measurement can be expected to be achieved with accurate calibration and correspondence matching.



(a) (b) (c)

Figure 4.2. Camera and human depth thresholds with different camera separations a) b = 7 cm, b) b = 14 cm and c) b = 21 cm. The theoretical depth threshold of stereo camera measurement was computed from Equation 3.4 using a 0.1 pixel disparity accuracy. The theoretical human depth threshold was computed from Equation 2.1 using a 20 arcsec disparity accuracy, and the theoretical depth threshold through the video see-through display was computed from Equation 3.1 using a 0.5 pixel disparity accuracy.

4.3 Auxiliary Augmentations (AAs) (PII and PIII)

4.3.1 Background and Objectives

The previous section dealt with the threshold of stereoscopic depth measurement and the perceived depth threshold through stereoscopic imaging and display components. This section focuses on depth perception in the case of augmented objects. The objective was to conceptualize a visualization approach for finding a solution to impairment of depth perception in AR. The objective is releted to Objective 3 presented in Section 1.4.

Wither and Hollerer (2005) showed that depth judgment errors between the user and augmented objects can be reduced if the scene contains multiple augmented objects. They used relative size as the depth cue at distances ranging from 20 to 60 m. Here, we study the effect of multiple augmentations using a more systematic approach. In addition to relative size, the effect of stereoscopic perception on depth judgments is studied.

4.3.2 Principle of Auxiliary Augmentations

In addition to the augmented object of interest (AOI), the conceived visualization approach involves the use of auxiliary augmentations (AAs). The AOI, such as an information label or an arrow in a wayfinding application, is the main object of interest. AAs are reference objects that resemble the AOI.

The principle behind the visualization approach is that the real world is overlaid with AAs. With the proposed approach, the user evaluates the position of the AOI by comparing it with the AAs. The principle is depicted in Figure 4.3. The AAs are anchored to the real world to achieve unambiguous perception. AAs increase the interaction with the physical world and offer relative depth cues for the AOI. Based on anchored AAs, the position of the AOI can be deduced using relative depth cues between the AAs and the AOI. The most accurate relative depth cues within the personal and action spaces (illustrated in Figure 2.1), namely, occlusion, motion parallax, binocular disparity, relative size and height in visual field (expressed as angular declination in PIII), are utilized for depth cueing. Based on accumulative depth cue integration theory, it can be expected that an increase in depth cues improves the depth judgments, as discussed in Section 2.1. Moreover, Holway and Boring (1941) showed that relative size coupled with binocular disparity improved depth judgments at distance range of 3 m to 37 m. Thus, a depth cue combination of binocular disparity and relative size is expected to provide accurate depth judgments within the action space.

In non-photorealistic rendering, augmented objects, both AOI and AAs, are easily seen as separate from the real world. Gestalt laws are used to explain visual grouping. They are a set of principles used to describe which visual properties prompt humans to combine visual elements into groups (Wertheimer, 1923). Gestalt laws have been shown to be efficient in connecting text labels and objects in virtual environments (Polys et al., 2011), and the laws can be used to provide connections between AOI and AAs. The gestalt laws of closure, similarity and proximity can at least be used.

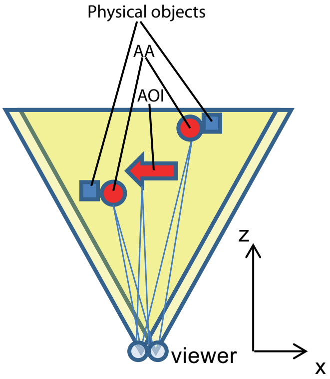


Figure 4.3. An example of AAs (circles) added to the scene to improving the depth localization of thei AOI (arrow). The squares represent real world objects.

Other relative depth cues between AAs and AOIs can be used, such as brightness and texture density. However, contrast functions as a relative depth cue only if the display can accurately reproduce the brightness differences. Especially with optical see-through displays, a lack of brightness is a common a problem making it a difficult relative depth cue in AR. The perception of texture density is dependent on the resolution of the display, which is still limited in many HMDs.

The AAs are located spatially near real-world objects; thus, the depth judgment cannot be biased along a short spatial distance between the AAs and the real-world objects. The depth between AAs and AOI should be small to reduce errors in judgments of depth magnitudes, as discussed in Section 1.3.2.

The human ability to transform disparity to depth magnitude depends on the viewing distance (see Section 2.2.1). Observers can use AAs to scale the viewing distance to the object and transform the disparity between AA and AOI to the depth magnitude closer to vedical. Thus, the use of AAs can be expected to be beneficial especially with stereoscopic perception.

Drascic and Milgram (1991) found that users locate the position of a stereoscopic virtual pointer by applying the method of limits approach. Users typically move the virtual pointer to the far and near limits before finding the correct location between the limits. This indicates that there must be at least two AAs, one in front of the AOI and one behind the AOI, as illustrated in Figure 4.3. However, a limitation is that this scenario is only possible if there are real-world objects in front of and behind the AOI. If this construct is the case, then the comparison task can be narrowed to a certain depth range, which is limited by AAs at the near and far depths. Using AAs is potentially useful in situations where anchoring the AOI itself is not possible. An example of such a case is X-Ray visualization (presented in Section 3.4.5), where the AOI cannot interact with the neighboring physical environment because it is not visible. Another case arises when scenes are viewed from perspectives that hide the ground plane.

4.3.3 Positioning the AAs

Optimal Position
Optimally, the AOI and AAs are within the foveal area (within a 1° visual angle of the visual axis), as within this area the objects are perceived as sharp and the stereo acuity is at the highest within this area (Schor and Wood, 1983). These constraints are very strict, and adding AAs according to these constraints can be difficult. Thus, we give more permissive guidelines for situations when optimal positioning is not possible.

Depth Position
The depth positions of AAs are limited by the comfortable viewing range (see Section 3.5.2) and Panum's fusional range (see Section 2.2.2). The comfortable viewing range for the mismatch is limited to approximately ±1° around the convergence angle of the screen plane (see Section 3.5.2).

For static and moderately large objects, Panum's fusional range is approximately ±20 arcmin for uncrossed and crossed disparities. The allowed disparity range from Panum's fusional range is denoted by α in Figure 4.4.

Horizontal Frontal Position
The guidelines for limiting the horizontal frontal position arise from eye and head movements. Eye movements that are within $5 - 10^\circ$ usually require only saccades, and eye movements greater than 10° require larger saccades with head movements (Arthur, 2000). Unnecessary head movements should be avoided in AR, as inaccuracies in head tracking may influence depth perception. As a result, we set the condition for the horizontal angle (β) to be within 10°, as illustrated in Figure 4.4.

In accurate depth judgment tasks, the limitation for the horizontal location arises from a horopter that deviates from a straight plane. The errors in depth judgments are approx. 1 arcmin at 10° from the middle axis, as discussed in Section 2.2.2.

Vertical frontal position
Depth perception is not based solely on commonly known depth cues but is also based on ground perception and understanding the relative arrangements of parts of the terrain. Gibson (1950) emphasized the importance of

ground perception in space perception. Hence, perceiving the overall layout that is related to the ground, can be considered a more integral part of distance perception than that related to the ceiling. In fact, more accurate depth judgments have been made with floors than with ceilings (Bian and Andersen, 2011). This may be because ceiling heights vary by place (Thompson et al., 2007). In addition, most objects in natural scenes located within the action space are below the eye level (Yang and Purves, 2003). Thus, for visual cueing, it is important to direct the user's gaze below the horizon, and thus, the positions of AAs should be below the eye height but above the ground level. The lower panel of Figure 4.4 illustrates this.

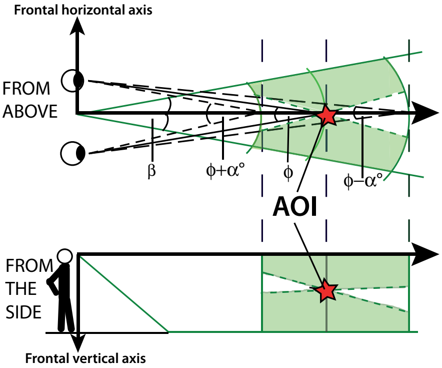


Figure 4.4. Possible positions for AAs are shown as the green area (gray in the black-and-white version). The upper panel shows the visualization space from above and the constraints for the depth and horizontal frontal position. With Panum's fusional limits, the depth positions for AAs are constrained by the convergence angle, denoted by ϕ, and the disparity, denoted by α. The horizontal position is constrained by the horizontal angle, denoted by β. The lower panel shows the visualization space from the side and the constraint for the vertical frontal position.

4.3.4 Anchoring the AAs to the Real Environment

The AAs need to be anchored to real world structures to enable the relative depth cues between real world structures and augmentations. If the AAs are perceived to be in a different position than where they are aligned, then the relative depth cues between the AOI and AAs do not facilitate the depth judgment.

Occlusion, as discussed in Sections 2.1 and 3.4.4, is the most dominant depth cue and is very efficient for cueing ordinal depth. Handling occlusion between augmentations and the real world requires a 3D reconstruction of the real world. Dynamic occlusion handling requires active measurement of the depth between the user and the augmentation, which is difficult to accomplish accurately.

However, the dominance of occlusions suggests that, rendering occlusions of AAs by static real-world objects can be an efficient approach for anchoring the AAs to the scene. Shadows, as discussed in Sections 2.1 and 3.4.4, have been shown to be an efficient way to visualize the spatial relationships of objects in AR. Thus, shadows should be used when a physical surface on which to cast them is available.

4.4 Augmentations above the Ground Plane (PII)

4.4.1 Background and Objectives

The first subobjective of this study was to implement the visualization approach of AAs to the objects above the ground plane. This arises from Objective 3 in Section 1.4. A relative size depth cue was created by adding AAs to the scene. The properties and positions of AAs were selected according to the constraints given in Section 4.3. These constraints were formed based on results from distance perception studies, and thus, it is expected that following the given guidelines, the accuracy of depth judgments is improved. The positions of the AAs were selected according to the constraints for position 4.3.3, anchoring 4.3.4 and screen space 3.4.1.

The second subobjective of this study was to establish the visual performance of perception through stereoscopic systems within the action space. It is still unknown how much stereoscopic perception decreases depth thresholds of objects within the action space that do not have contact with the ground (i.e., floating objects). The effect of stereoscopic viewing and relative depth on depth perception can also be significant within the action space in AR. This objective is related to Objective 4 in Section 1.4.

4.4.2 Materials and Methods

We studied the effect of AAs and stereoscopic and monoscopic viewing conditions on the depth threshold with the method of adjustments. Methods for measuring depth thresholds are discussed in Section 1.3.2. The task in the experiment consisted of asking the participants to align the depth position of a physical pointer (red ball) to match the position of the AOI. The AOI did not have a shadow underneath, as is evident in Figure 4.5, because we wanted to test a case in which the visual interaction between a virtual object and the physical environment is not possible. Participants were asked to stand and they were allowed to move freely on an exercise mat. The dimensions of the mat are marked in Figure 4.6.

An example of a stimulus is shown in Figure 4.5. The AAs and the AOI were a red cone with a semi-transparency of 0.6. It has been shown that similarity in color greatly influences the weight of relative size as a depth cue Sousa et al. (2012). Thus, using the same color for AAs and AOIs yields the most predictable results. The height and width of the AOI was 2.3°. The participant was told that if there are multiple augmented objects in the same scene, their size is equal. This allowed us to study the effect of relative size on depth judgments.

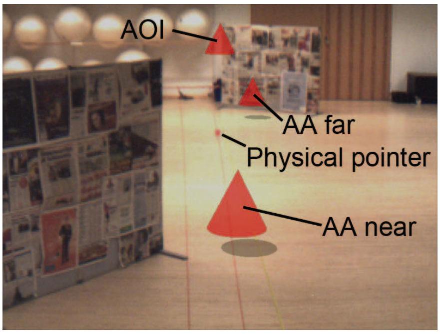


Figure 4.5. An example stimulus with the *Mono with AAs* condition. The distance to the AOI is 10 m, and the height position is 1 m.

The variables of the experiment are listed in Table 4.4. The *Stereo condition* was switched between *off* and *on*. With the *off*, condition the image from the left camera was shown to both eyes. With the *on* condition, the images from both cameras were shown to the eyes and the scene was rendered separately for the left and right eyes. With the *AA Condition*, the scene was shown with or without AAs. With the *Height position* condition, the height position of the AOI was varied between 0.5 m and 1.0 m. The distance of the AOI was varied from 6 m to 10 m. The distance to the near AA was 5 m, and

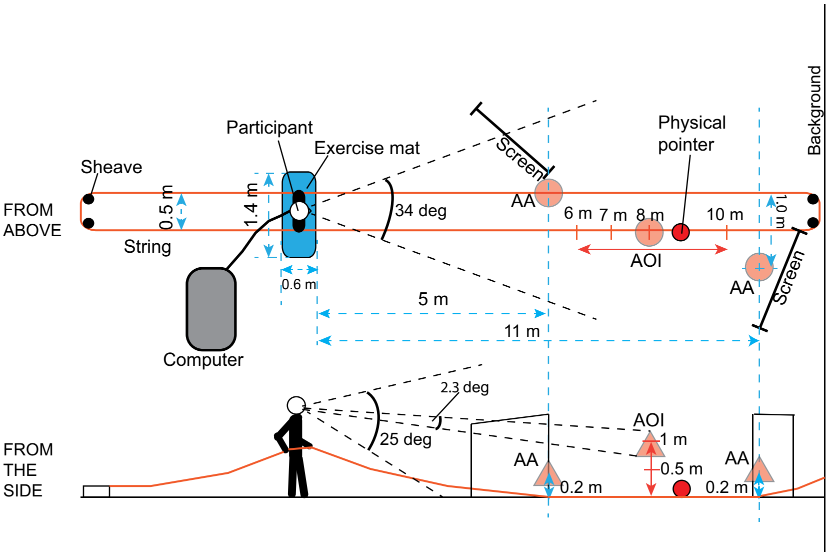


Figure 4.6. Dimensions of the environment used in the study related to PII. The dimensions of the space were 18 m × 18 m.

the distance to the far AA was 11 m from the user. The dimensions of the study are shown in Figure 4.6.

Table 4.4. The independent and dependent variables of the study related to PII.

Independent variables	**N**	**Description**
Participant	19	Random variable (nine females). The age span was from 21 to 50 years, with a mean age of 26.
Stereo condition	2	Off, on
AA condition	2	Without AA, with two AAs
Height position	2	0.5 and 1.0 (in meters)
AOI Distance	4	6, 7, 8, 10 (in meters)
Repetition	2	Each combination of the other independent variables was shown two times, with no identical stimulus in succession.
Dependent variables	**N**	**Description**
Depth judgment	1152	In centimeters
Depth threshold	1152	In meters, the smallest correct ordering of distances of AOI. Standard deviation was used as the criterion for the depth threshold.
Confidence of evaluations	1152	The rate of confidence on a scale (1 to 5) for judged depth: 1 represents ”Very unconfident” and 5 ”Very confident.”

4.4.3 Results

The results in Figure 4.7 and Table 4.5 show that the We fitted lines to the judgments using linear regression. The slopes can be observed in Figure 4.7, and the values are shown in Table 4.5.

Statistical analysis was conducted using a repeated-measures ANOVA for mean values of repetitions. The Stereo variable had a statistically significant effect on the signed [F(1,16) = 15.783, p = .001, η_p^2 = 49.7%] and absolute error [F(1,16) = 14.996, p = .001, η_p^2 = 48.4%]. The signed and absolute errors were significantly higher for the Mono condition than the Stereo condition. In addition, the AA variable had a statistically significant effect on the signed [F(1,16) = 30.641, p < .001, η_p^2 = 65.7%] and absolute error [F(1,16) = 51.711,

$p < .001$, $\eta_p^2 = 76.4\%$]. The errors were significantly smaller for the AAs condition. The signed error was mostly negative, which indicates underestimation of distance (i.e., the judged depths were smaller than the veridical values).

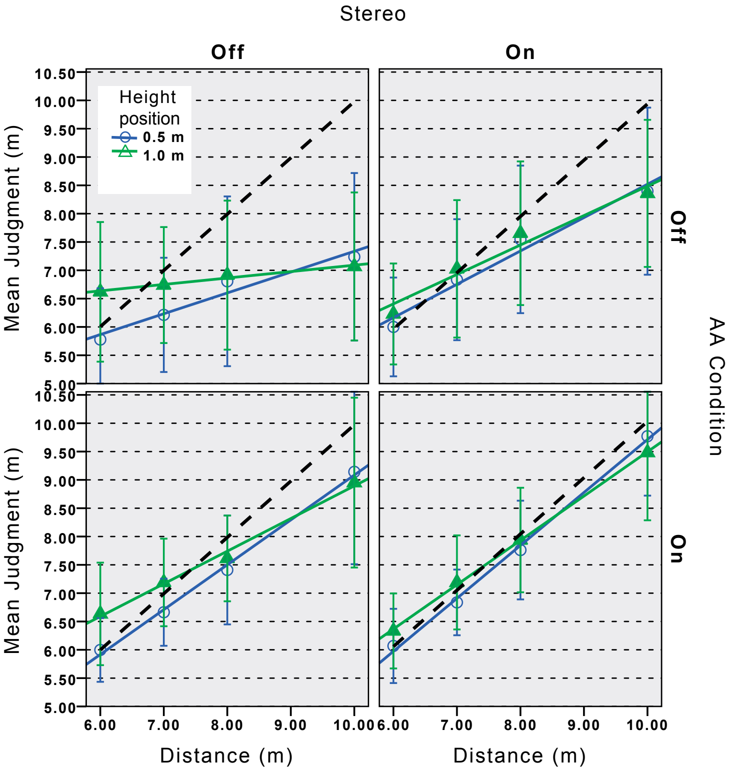


Figure 4.7. The mean judgments as a function of distance to AOI. Both the visualization condition and height position had a significant impact on judgments. The dashed lines represent veridical judgments. The error bars represent standard deviations.

The Height position did not have a significant effect on the absolute error [F(1,16) = .639, p = .639, η_p^2 = 1.4%]. However, it had a significant effect on the signed error [F(1,16) = 10.887, p = .005, η_p^2 = 40.5%]. The higher the object is the farther away it is perceived. The *height position* has an effect on slopes with visualization conditions. When the object is higher, the slope is reduced with every visualization condition. From Figure 4.7, it can be observed that the *height position* has an effect as a function of distance. At closer distances (6 m - 8 m), the judgments are overestimated, especially with the *Mono* condition, where the *height position* of the object is 1 m. When the *height position* of the AOI is 1 m, the AOI is viewed against the background. When the *height position* of the AOI is 0.5 m it is viewed against the floor at distances of 6 m to 8 m. At a 10 m distance, the all stimuli are perceived against the background, and the effect is not present. With the *Stereo* condition, the *height position* does not have an effect on depth judgments past a 6 m distance.

Depth thresholds (measured as the smallest depth difference between objects that is correctly resolved) were determined using standard deviations of depth judgments. For example, with the *Stereo* condition the standard deviation at 6 m distance exceeds the average value at 7 m, but not at 8 m. Thus, depths at 6 m and 7 m are interpreted to be on the same depth level. The

Table 4.5. The slopes and R^2 from the linear regression of mean judgment vs. distance to AOI for different visualization conditions and object height positions. (* the regression model is not statistically significant.)

Visualization condition	Height position (m)	Slope	R^2
Mono	0.5	0.335	0.13
	1	0.093	0.013*
Stereo	0.5	0.569	0.352
	1	0.511	0.337
Mono with AAs	0.5	0.797	0.542
	1	0.603	0.413
Stereo with AAs	0.5	0.946	0.737
	1	0.773	0.557

depth thresholds for other conditions were determined in a similar manner. The depth thresholds are shown in Table 4.6.

Table 4.6. The depth thresholds (in m) for different visualization conditions and object height positions. (* the depth threshold is below the minimum value of the study; ** the depth threshold is above the maximum value of the study)

Visualization condition	Depth threshold	
	Height position 0.5 m	Height position 1.0 m
Mono	4	4 **
Stereo	3	3
Mono with AAs	1*	3
Stereo with AAs	1*	2

The results provide guidelines for depth cueing augmentations in AR, as observed from the slopes in Table 4.7. The results support additive cue theory, as the slopes are increased toward veridical perception as a function of the number of depth cues. In addition, the higher the AOI was shown, the more stereoscopic perception influenced depth judgments.

When the AA was present and the AOI was 1.0 m above ground, the AOI judgments compared with veridical depth increased from 9 % to 60 % with the monoscopic viewing condition and from 51 % to 95 % with the stereoscopic viewing condition.

4.5 Augmentations in X-ray Visualization (PIII)

4.5.1 Background and Objectives

In this study, we investigated the effect of AAs and viewing conditions (stereoscopic and monoscopic) on the depth threshold for overriding occlusion in X-ray visualization.

The first subobjective is to reduce depth judgment errors with the visualization approach of AAs (presented as Objective 3 in Section 1.4). In the X-ray visualization case, there are no real-world objects visible behind the wall; hence, the AA is placed only at the wall distance.

Occlusion as the overriding and dominant depth cue must be taken into account when designing depth cueing in X-ray vision. The AA is designed to provide the occlusion cue. An example is shown in Figure 4.8. The circle (AA) is seen behind the gray wall because of occlusion and the completion of a known form (closure as the Gestalt law) (Ware, 2004). In other words, an occlusion depth cue can be created with an AA to enhance the ordinal depth interpretation, and the binocular disparity can be used for depth magnitude estimation between the wall and an AA. This approach thus combines the most dominant depth cue (occlusion) and the most accurate depth cue (binocular disparity) within the personal space and within the low end of the action space, as Figure 4.8 illustrates. In our technique, the AOI is observed through a rendered "virtual window", which represents a rectangular hole in the wall. It is similar to the object used by Sielhorst et al. (2006).

The proposed implementation is expected to make depth perception less ambiguous. Height in the visual field is used to enhance the impression of seeing through the walls by positioning the AA below the AOI, as illustrated in Figure 4.8.

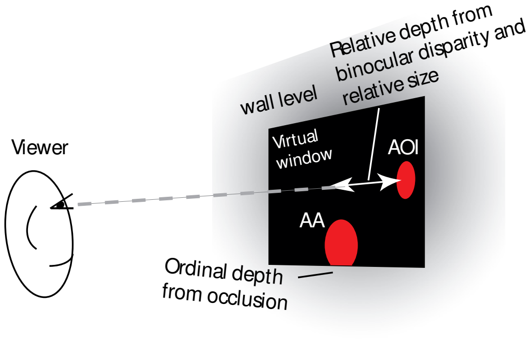


Figure 4.8. An illustration of auxiliary augmentation (AA) in X-ray visualization. In this scenario, the AA is used to show the augmented object of interest (AOI) behind the wall. The AOI (full circle) is visualized behind the AA (partially occluded circle) using virtual window and relative depth cues.

The effect of occlusion on depth thresholds in AR is still unknown. The effect of occlusions on stereoscopic depth thresholds has been studied by Ellis and Menges (1998) and Swan II et al. (2006). Ellis and Menges (1998) studied the effect of a partially occluding real-world object on depth judgments of a virtual object at close distances ($\leq$ 2 m). The stereoscopic viewing improved perceptual matching accuracy when the virtual object was occluded. They also found that an occluding surface could affect distance judgments through a measured change in convergence caused by an occluding object. Swan II et al. (2006) studied the effect of a binocular X-ray viewing condition by adding an occluding surface in a perceptual matching task at distances of 5 to 45 m and found that the occluding surface increased the depth judgment errors.

However, it is still unclear what the depth threshold is for binocular disparity at which the occlusion is reliably overridden. The second subobjective was to determine the effect of stereoscopic viewing on depth thresholds and is related to Objective 4 presented in Section 1.4. In addition, it can be expected that the evaluation times are will decrease with the AA. If the depth interpretation is facilitated, then it should decrease the performance time.

4.5.2 Materials and Methods

The task for the participants ($N = 14$) was to judge the depth of the AOI compared with the position of the wall in the X-ray visualization case. The effect of the AA is examined with two depth judgments. In the ordinal depth judgment, the effect of the AA for suppressing occlusion is investigated and the participants are asked to select the ordinal position (front, at or behind). The depth threshold is defined to correspond to the level at which 75 % of the evaluations are correct, as discussed in Section 1.3.2. In addition, we use a similar metric to that used by Livingston et al. (2003) to evaluate the ordinal error. The ordinal error level e_o is an absolute value of difference between judged ordinal position j_o and actual ordinal position a_o, and it is defined as $e_o = |j_o - a_o|$, in which judged values and actual positions have values from 1 to 3. For an ordinal error level of 2, for example, the AOI is positioned behind a wall, but the participant perceives it as in front of the wall.

In the depth magnitude judgment, discussed in Section 1.3.2, the metrical accuracy of depth cues (relative size and disparity) is studied and the participants are asked to judge the depth magnitude (in centimeters) of the AOI compared with the wall distance. The depth magnitude is the depth between the wall and the target, and the depth magnitude error e_i is computed as an absolute value of difference between judged magnitude j_i and actual magnitude a_i and it can be written as $e_i = |j_i - a_i|$.

Table 4.7. The independent and dependent variables.

Independent variables	**N**	**Description**
Participant	14	Random variable (two female). The age span was from 20 to 42 years, with a mean age of 27.
Viewing condition	2	Stereoscopic, non-stereoscopic
AA condition	2	With AA, without AA
Relative depth	7	Three in front of the wall, one at the wall and three behind the wall. Negative values correspond to values in front of the wall, and positive values refer to positions behind the wall.
Repetition	3	1, 2, 3
Dependent variables	**N**	**Description**
Depth threshold	1176	In centimeters. Measured as the smallest depth difference that is correctly resolved (in front of the wall, at the wall, behind the wall).
Depth magnitude	1176	In centimeters. Negative values for in front of and positive values for behind the wall.
Time	1176	The time of task completion was recorded with millisecond accuracy
Confidence of evaluations	1176	The rate of confidence on a scale of (1 to 5) for both ordinal position and judged depth: 1 represented "Very unconfident" and 5 "Very confident."

4.5.3 Results

Depth Threshold
The use of AA facilitated the assessment significantly. From Figure 4.9 it can been observed that with only the *Stereo with AA* visualization condition the correct rate is above 75 % (i.e., the error rate is below 25 %) at all depths from the wall. With *Mono* and *Stereo* conditions the error rate is above 25 % at all the depths behind the wall. With the *Mono with AA* condition, the correct rate of 75 % is exceeded at 80 cm behind the wall. With the stereoscopic condition, the depth threshold was decreased from above 20 arcmin to below 6 arcmin by using the AA.

When the AA was present and the augmented object of interest was behind the wall, the error rate in the ordinal task was reduced from 65 % to 33 % with the monoscopic viewing condition and from 42 % to 10 % with the stereoscopic viewing condition. With the *Mono with AA* condition the limit for JND was exceeded (error rate is less than 25 %) when the AOI was 80 cm behind the wall.

Statistical tests were conducted using repeated-measures ANOVA design, which differs from statistical test design used in PIII. Overall, the visualization condition had a statistically significant main effect on the ordinal error rate [F(3,30) = 7.928, p < .001]. With AA, the ordinal error was statistically significantly lower with both, monoscopic and stereoscopic, conditions as shown in Table 4.8.

Depth Magnitude
The data presented in Figure 4.10) show that the participants estimated all of the depth magnitudes to be shorter than the actual depth from the wall (position of AOI = 0 cm). The depth scale used appears to be considerably

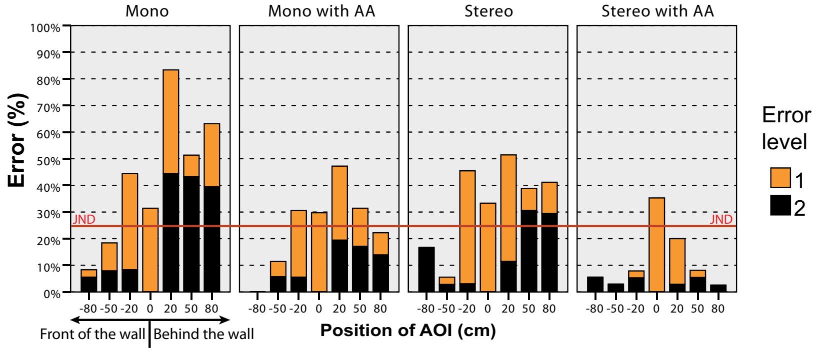


Figure 4.9. The percentage of errors in ordinal evaluations as a function of the actual depths with different visualization conditions. Error level 1 means an error of one step, for example, when a participant judges the AOI to lie at the wall, but the actual position is been behind the wall. Error level 2 means an error of two steps, for example, when a participant judges the AOI to lie in front of wall, but the actual position is behind the wall.

Table 4.8. Significance values from post hoc test for ordinal errors.

	Mono with AA	Stereo	Stereo with AA
Mono	p = .001	p = .079	p < .001
Mono with AA		p = .512	p = .060
Stereo			p < .001

flattened, as has been observed also with real world stimuli (Loomis et al., 1992; Allison et al., 2009) and in purely virtual environments (Hubona et al., 1999; Willemsen et al., 2008). Furthermore, verbal reporting has been shown to cause underestimation of distances compared with walking and reaching responses (Philbeck and Loomis, 1997; Napieralski et al., 2011). Verbal reporting may also have impacted our results.

The depth magnitude errors between the actual values and the depth judgments were calculated as the absolute difference between the actual and judged depths. There was a statistically significant main effect of the visualization condition [F(3,30) = 12.189, p < .001] on the depth magnitude error. Post hoc test (Table 4.9) revealed that the presence of the AA reduced the error under both of the viewing conditions. There is a statistically significant difference (p < .05) between the "Stereo with AA" and "Mono with AA" conditions. This result indicates that the presence of AA is especially useful in stereoscopic systems. Without AA, there was no statistically significant difference (p = .059) in the error between the "Mono" and "Stereo" conditions.

Table 4.9. Post hoc test for depth magnitude errors. Stereoscopic viewing with AA reduced the errors significantly under both viewing conditions.

	Mono with AA	Stereo	Stereo with AA
Mono	p = .002	p = .059	p = .001
Mono with AA		p = .717	p = .008
Stereo			p = .006

Performance Time

There was no statistically significant main effect of the visualization condition on time [F(3,30) = .032, p = .992)]. However, when the AA was present, there was a statistically significant main effect of the position of the AOI on response time [F(6,60) = 3.800, p = .003], as observed from V-shaped curve in Figure 4.11.

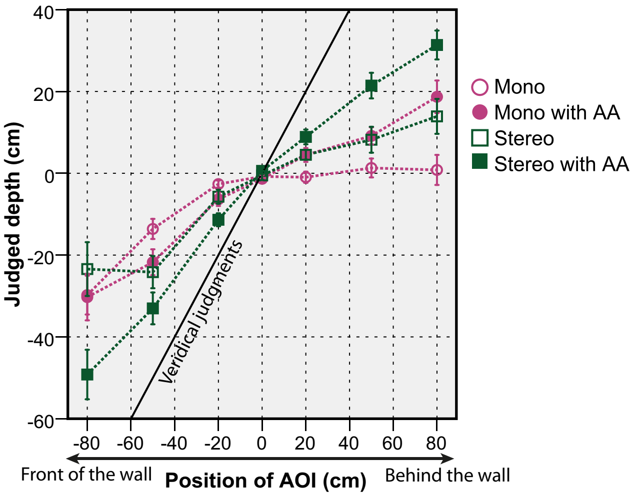


Figure 4.10. The mean depth magnitude judgments as a function of actual depth magnitudes. Negative values correspond to the front of the wall (toward the participant), and positive values correspond to behind the wall. Error bar represents the standard error of the mean.

When the AA is present, the mean response time increases monotonically as a function of distance from the wall under both viewing conditions. This result suggests that the depth difference between the AA and the AOI should be minimized. Without the AA, there was no statistically significant main effect of position of the AOI on response times [F(6,60) = .416, p = .866].

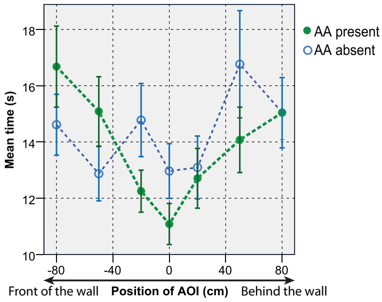


Figure 4.11. Mean times as a function of actual position of the AOI under different AA conditions. Error bar represents the standard error of the mean.

4.6 Imaging of Crowded Scenes (PIV)

4.6.1 Background and Objectives

The visual space can be very crowded in urban environments. In cities ,there are crowds, traffic, buildings, advertisements and shops in the views. The observation of crowds is a common task in airports, malls and at events because of increasing security concerns. The growing importance of crowd analysis is apparent in computer vision research (Zhan et al., 2008), a field in which some studies have utilized stereoscopic imagery (Huang et al., 2004; Yahiaoui et al., 2010). However, perceptual studies with human participants appear to be missing.

Partially occluded bodies are very common in crowded scenes (Leibe et al., 2005). The objective of the study relates to establishing the visual performance through stereoscopic systems within the action space (presented as Objective 4 in Section 1.4). Stereoscopic perception can be expected to improve the detection accuracy of complex and partially occluded areas as suggested by the literature (see Section 1.3.2 and Figure 4.12). To determine whether this is the case with crowds, a study was conducted to compare the segregation ability using two distinct viewing methods: monoscopic and stereoscopic.

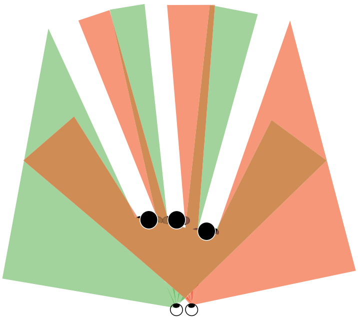


Figure 4.12. A schematic illustration of stereoscopic detection of half-occluded regions in crowd perception. The view seen by the left eye is shown as green, and the view seen by the right eye is shown as red. The overlapping views are shown as brown. The totally occluded areas are shown as white.

4.6.2 Materials and Methods

Images of crowded urban environments were taken using a stereo camera. The depth range of the crowded scenes and the geometry of stereoscopic photography are depicted in Table 4.1. The images were taken at eye level in situations that would be difficult for computer vision analysis to process and that would require human visual inspection because of the many largely occluded faces and bodies. We used still images instead of videos to enable more controlled stimuli. Images have previously been used to study crowd perception. For example, Ennis et al. (2011) used still images to investigate the realism of virtual crowds.

The head count was chosen for the study to evaluate the capabilities of viewing methods to distinguish people from each other. The head count of the participants shows how well they performed that task. The performance time to complete the task was measured. It was expected that the time is reduced in the stereoscopic condition because of the increased ability to segregate people. In addition, the participants were asked to evaluate the ease of completing the task using a 7-point Likert scale: “The task was completed. . . (1 = with difficulty, 7 = very easily).” The question was used to quantify the ease of recognizing individual persons from the image.

In addition to head counting, the participants judged the ease of evaluating depths between persons using a 7-point Likert scale: “The estimation of depths between persons was. . . (1= very difficult, 7= very easy).” This question was asked to determine how the overall structure of the crowd is perceived.

After the experiments, the participants were asked to decide which viewing method they preferred and why. The variables are shown in 4.10.

4.6.3 Results

The counting task showed differences in accuracy between the viewing methods, as summarized in Figure 4.13. Stereoscopic viewing was significantly more accurate (average error = 11.9 %, sd = 9.4%) than non-stereoscopic viewing (average error = 15.1%, sd = 7.3%; [t(144) = –2.446, p = .016)].

Table 4.10. The independent and dependent variables of the study related to PIV.

Independent variables		
Variable	Number	Description
Participant	19	Random variable (5 female). The age span of the participants was 21 to 43 years.
Viewing condition	2	Stereoscopic, non-stereoscopic
Images	8	Images of crowds
Dependent variables		
Detection accuracy	152	The accuracy of the counting task was measured with a correct rate
Time	152	The time of task completion was recorded within an accuracy of a tenth of second
Ease of the task	152	The task was completed. . . (1 = with difficulty, 7 = very easily)
Ease in estimating distances	152	The estimation of distances between persons was. . . (1= very difficult, 7= very easy)

There was no notable difference between performance times. The counting time with stereoscopic viewing (mean time 19.3 sec, sd = 6.8) was longer than with non- stereoscopic viewing (mean time 18.7 sec, sd =6.4 sec), but the difference was not statistically significant [t(144) = 0.60, p = .55)].

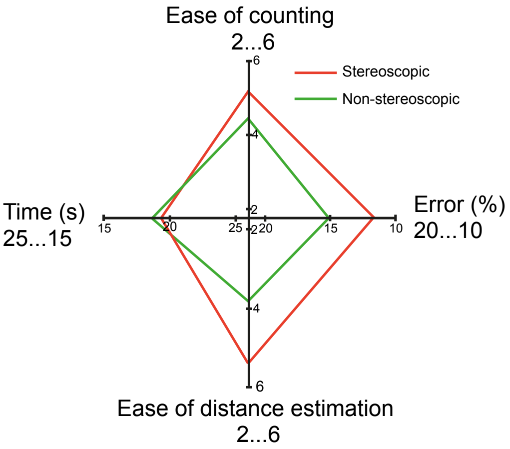


Figure 4.13. Summary of results from the study related to PIV. The scales of the axes are aligned so that the performance increases as a function of distance from the origo.

The counting task and the estimation of distances between people were determined to be easier with stereoscopic viewing. Overall, the majority of the participants (67 %) preferred stereoscopic viewing, whereas a minority of the participants (11 %) preferred non-stereoscopic viewing. The rest had no preference. The viewing condition did not have a statistically significant effect on performance time.

As mentioned, the images used in this test were taken at eye level, whereas surveillance cameras are usually placed higher above the crowd. A higher angle is likely to make the given task easier because of fewer occlusions. However, in large areas, such as plazas, the relative angle decreases at greater distances, so more occlusions occur. In future studies, it would be important to vary the angle and test the range of angles across which stereoscopic viewing becomes beneficial.

4.7 Imaging of Indoor Scenes with People (PV)

4.7.1 Background and Objectives

In addition to performance measures, the viewing experience is an important issue when considering stereoscopic systems for human use. In the development cycle of new systems, the emphasis is first on performance factors and progresses toward experiential factors with advances in performance, as illustrated in Figure 1.5.

Although models exist for visual experience, as discussed in Section 1.3.2, there is no well formulated model of the effect of stereoscopic depth on the viewing experience. The challenge in S3D imaging compared with traditional photography emerges from the imaging geometry because the perceived depth must be controlled to ensure a good level of visual experience.

To do this, there is a need to know which depth magnitudes are preferred, what the viewing conditions are, which camera parameters are used, and which depths occur in the natural scenes. The key component for controlling perceived depth is camera separation. The objective of the study was to discover the effect of camera separation on the viewing experience (presented as Objective 5 in Section 1.4.

4.7.2 Materials and Methods

Four typical indoor imaging scenes (shown in a upper row in Figure 4.14), most of which depicted people in natural settings, were imaged with camera separations from 2 to 10 cm. Goldmann et al. (2010) investigated the effect of camera separation on the viewing experience in outdoor conditions with longer camera separations than 10 cm. We did not choose to replicate that study and focused on smaller camera separations in an indoor environment. The depth range of the scenes and the geometry of the stereoscopic photography are depicted in Table 4.1. The participants evaluated the strength and naturalness of depth sensation and the overall viewing experience with at a scale of 1 to 7, where 7 represented the highest score. The variables are listed in Table 4.11. The statistical analysis, conducted using a repeated-measures ANOVA design and a post hoc test is used to analyze which camera separations differ statistically. This statistical test design differs from the one used in PV.

Figure 4.14. Scenes in the following order: “Bar”, ”Wine”, “Game” and “Composition.” The first row shows the S3D scenes used in this study. The bottom row shows the depth maps. The white values correspond to the nearest depth and the darkest values to the farthest depth. The depth maps were computed separately for each scene and should not be compared.

In addition, the participants were asked to draw an ellipse on the image to indicate the area from which they evaluated the naturalness of depth sensation. If the effect of the marked area was positive, the area was marked with a green transparent ellipse; if the effect was negative, the area was marked with a red transparent ellipse. This Recall Attention Map (RAM) approach is described in more detail in a study by Hakala et al. (2011). Possible effects of visual discomfort were noted by asking the participants about headaches and eye strain before and after the experiment.

The cardboard effect was estimated with a roundness factor, which was computed using the geometry of the stereoscopic pipeline, disparity maps and participant selections for important areas using the RAM approach.

Table 4.11. The independent and dependent variables.

Independent variables		
Variable	Number	Description
Participant	12	Random variable (three female). The age span was from 22 to 34 years.
Scene	4	Bar, Wine, Game, Composition
Camera separation	5 + (1)	2, 4, 6, 8, 10 (in centimeters), with the addition of 0 cm in the “Wine” scene
Repetition	2	Each stimulus was repeated
Dependent variables		
Strength of depth sensation	528	Scale of 1 to 7, where 7 represented the highest score.
Naturalness of depth sensation	528	Scale of 1 to 7, where 7 represented the highest score.
Viewing experience	528	Scale of 1 to 7, where 7 represented the highest score.

4.7.3 Results

Mean opinion scores

The mean opinion scores (MOS) of the three subjective experience factors are shown in Figure 4.15. Camera separation had a statistically significant (tested with a two-way repeated-measures ANOVA) effect on the strength of depth sensation [F(1.65,18.16) = 13.37, p < .001], naturalness of depth sensation [F(2.01,22.12) = 11.58, p < .001], and viewing experience [F(1.69,18.602) = 5.01, p = .022]. The degrees of freedoms were corrected using Huyhn-Feldt estimates of sphericity, as the assumptions of the sphericity were violated.

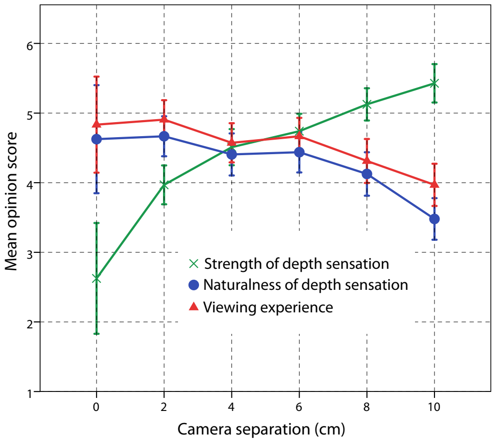


Figure 4.15. The MOS-values as a function of camera separation. Error bars are for the 95 % confidence level.

The strength of depth sensation increased with camera separation. Post hoc test revealed that the strengths of depth sensation differed statistically (p < .05) when the difference in camera separation was 4 cm or more. The result is in line with another study (Seuntiens, 2006), where the participants were able to sense the increased depth, but the differences in camera separations were higher. It has been shown that humans adapt to different depth scales quite easily if a reference depth scale is available (Milgram and Krueger, 1992).

The naturalness of depth sensation and viewing experience decreased at camera separations above 6 cm. Post hoc test revealed that the naturalness of depth sensation and viewing experience with a camera separation of 10 cm differed statistically ($p < 0.05$) from other levels of separation. The same trend, where the naturalness of depth decreases as a function of camera separation, has been found in other studies (Ijsselsteijn, 2000; Häkkinen et al., 2011).

Evaluations between and within Scenes

The results within scenes are shown in Figure 4.16. No clear differences between the scenes were evident, as the two-way repeated-measures ANOVA did not reveal any statistically significant differences in the strength of depth sensation [F(1.65,18.11) = 0.58, p = .539], the naturalness of depth sensation [F(3,33) = 0.48, p = .69], or the viewing experience [F(3,33) = 0.25, p = .99].

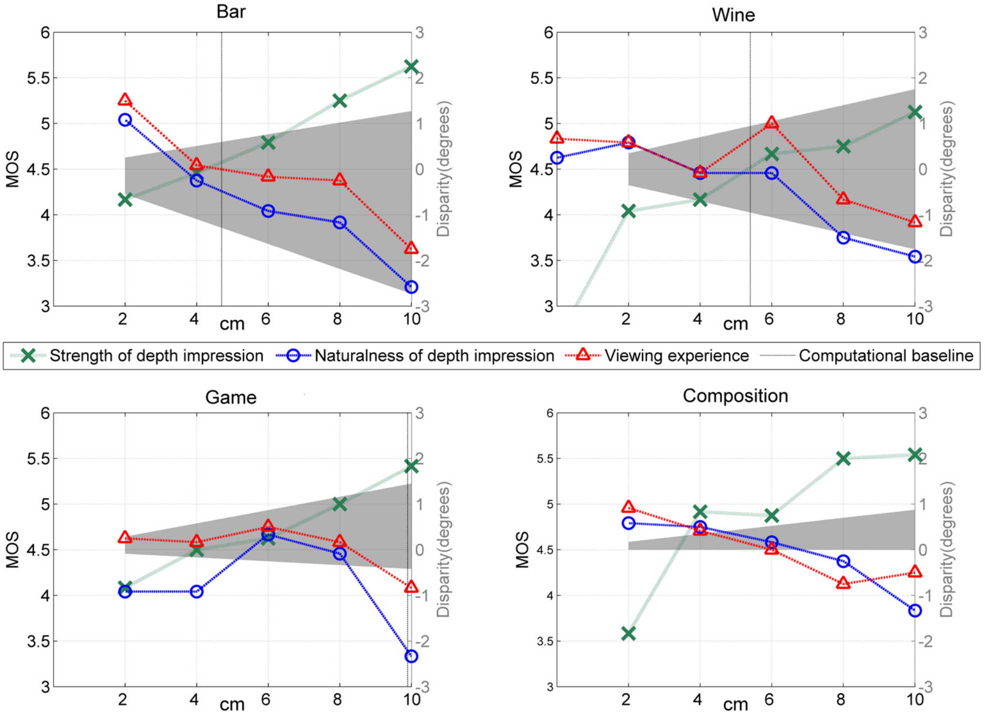


Figure 4.16. The MOS-values as a function of camera separation and disparity ranges. The disparity range (scale on right y-axis) was calculated according to Jones et al. (2001) and is shown as gray areas. The MOS-values are presented without error bars for clarity; the standard deviations are close to one for every attribute.

The interaction effect of camera separation and scene not statistically significant for any of the three variables: strength of depth sensation [F(5.16,56.73) = 1.51, p = .199], naturalness of depth sensation [F(5.52,60.73) = 1.69, p = .146], and viewing experience [F(5.39,59.25) = 1.64, p = .159].

Interestingly, the strength of depth sensation seemed to be strong even though the S3D image was shown only behind the screen, and the disparity range was short. This phenomenon was seen by comparing Bar and Composition. The strength of depth sensation in Composition (only a positive disparity from 0.2 ° to 1 °) was evaluated to be equally as high as Bar, even though its disparity range was shorter than in Bar (both positive and negative disparities).

In Wine, the hidden reference, namely, the 2-D image, was detected and perceived with the lightest depth sensation. This result indicates that the depth scale was used in the same way for the contents even though the disparity ranges of the contents are distinct.

Within scenes, however, there were statistically significant differences with respect to camera separation. In the scene with narrow depth variation in Bar, the naturalness of the depth sensation decreases with an increase in the length of the disparity range. In Game, there is greater depth variation in the scene, and the highest naturalness was achieved with a longer camera separation than in Bar. The viewing experience also changes according to camera separation, but the impact is smaller than with other attributes. For example, in Game the viewing experience is quite constant. The viewing experience behaves the same way as the naturalness of depth sensation. The limits for comfortable viewing are exceeded the most in Bar, and the near disparity is nearly 3° with the longest camera separation. This effect can also be observed in the RAMs, shown in Figure 4.17, where the amount of red at close depths increases as a function of camera separation.

(a) (b) (c)

Figure 4.17. The RAMs presented according to camera separation: (a) 2 cm, (b) 6 cm, and (c) 10 cm. The RAMs clearly show that with greater camera separation, the negative evaluations increase in the front part of the image where disparity has exceeded the 1 deg disparity limit. Green areas indicate a positive evaluation, red areas indicate a negative evaluation, and yellow areas indicate the evaluations are mixed. The RAMs are combined from all participants' evaluations.

Effect of the roundness factor

The selected roundness factors are computed by taking into account the disparity maps (Figure 4.14, bottom row) and the RAMs of the scenes (see the example in Figure 4.17). The depths for selected roundness factors are computed with the weighted average of the selected regions' depths. Finally, these selected roundness factors are compared with the naturalness of depth sensation. Figure 4.18 shows the naturalness of depth sensation with different camera separations as a function of selected roundness factors.

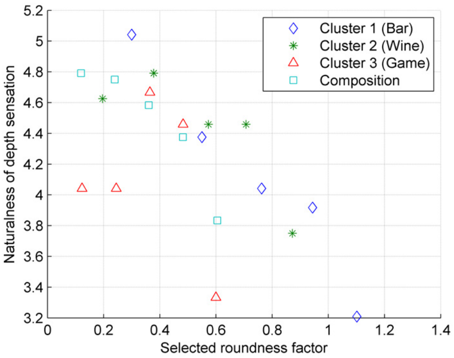


Figure 4.18. The naturalness of depth sensation as a function of the roundness factor.

It is worth noting that there was no positive correlation between naturalness of depth and the roundness factor with these imaging and viewing parameters. Contrary to expectations, the correlation is clearly negative ($r = -0.67$, $p < .01$). This result indicates that the effect of the roundness factor on the naturalness of depth sensation is not the critical factor under these viewing conditions. In Composition, the depth was perceived to be natural even at low levels of the roundness factors (0.1 to 0.3). When the roundness factor was between 0.3 and 1.1, the naturalness of depth sensation decreased as a function of the roundness factor in all content. The cardboard effect emerging from short camera separations is unlikely to be a problem with the desktop sized display used in this study. Interestingly, the naturalness of depth sensation was evaluated as high even though the roundness factor, previously thought to influence the naturalness evaluation, was low.

5. Discussion

This dissertation contributes to the knowledge of visual performance and subjective experience in stereoscopic systems and its conversion to spatial guidelines. This dissertation fulfills its mission by in terms of introducing stereoscopic systems to new application area of crowd perception and by providing guidelines for stereoscopic systems in AR within action space.

The contributions are discussed below with reference to the objectives presented in Section 1.4.

O1: Conceptualize and develop a method for measuring the depth threshold of human and stereoscopic measurements

The developed method can be used for comparing the depth threshold of stereo cameras and the human eye. The method is applicable when the same stereo camera is used for measuring the scene and displaying it to the observer. In such systems, the depth threshold of the stereoscopic measurement requirements should be derived from the perceived depth threshold.

The developed method can be used for selecting camera parameters for stereoscopic systems and for validating the geometric calibration of stereo cameras. The method enables measuring the depth thresholds with a test target using the rank ordering method. The rank ordering method is faster than pairwise comparison and was found to be applicable for depth threshold tasks.

The test target is simple and easy to assemble. However, it could be improved further, as with the current version there was the possibility to use relative size as a depth cue between the depth levels of the target. Different patch sizes would have removed the possibility of using relative size as a depth cue.

O2: Find the effect of camera separation on depth thresholds within the action space

The effects of camera separation on the depth threshold of stereoscopic measurement was evaluated with the new test target. Measurements with the test target showed differences between camera separations. However, the stereoscopic measurements were not improved as much as was theoretically expected.

In addition, an increase in camera separation did not significantly decrease the human depth threshold, as was expected from the geometry. This result is in a line with the studies of Rosenberg (1993) and Merritt et al. (2005). Rosenberg (1993) found no improvement in depth judgments with a camera separation beyond 3 cm. Merritt et al. (2005) used camera separations similar those of our study and demonstrated that increased camera separation did not increase performance in the detection of drop-offs in an off-road driving task.

Contributions related to this objective can be considered indicative, as the measurements were not repeated. Each distance and camera separation case was shown once to the participant. To achieve more reliable results, the measurement should be repeated so that multiple measurements are made at one distance and the target levels are mixed between shots.

O3: Design and implement a method to improve depth judgments concerning augmented objects within the action space

The reference augmentations, called auxiliary augmentations (AA), added relative size cue to the scenes. With relative size stereoscopic viewing improved the depth judgments for both of the studied scene types (augmented objects above the ground plane and X-ray visualization).

The results from the study with objects above the ground plane suggest that adding an AA to the scene is an efficient visualization approach within the action space. The results can be used to design computer aided navigators, where scale is within the action space, stereoscopically.

The results from the study with X-ray visualization suggest that the presence of auxiliary augmentation helps in depth judgment in ambiguous scenes. X-ray visualization scene is an example of a situation that is unfamiliar to observers. This unfamiliarity may cause ambiguous perception, which, according to the results, can be made unambiguous by using auxiliary augmentations. In X-ray visualization, the auxiliary augmentations were useful for visualizing objects behind a wall. For example, if water pipes are visualized behind a wall, then showing a stub of a pipe at the depth of the virtual window as an auxiliary augmentation can facilitate the interpretation.

Our study with monoscopic viewing showed an error rate of 63 % without AA and 23 % with AA when the AOI was 80 cm behind the wall. These results are comparable with a study by Furmanski et al. (2002) where AOI was visualized 1 m behind the wall using motion parallax and partial occlusion, which found approximate error rates of 90 % with motion parallax and 40 % with a partial occlusion cue. In the study by Furmanski et al. (2002), the AOI itself was partially occluded, whereas in our study, the AA was partially occluded. The lower error rate with a partially occluded AA compared to a partially occluded AOI underlines the efficiency of the AA approach.

Other depth cues than binocular disparity and relative size could be utilized between the AOI and AAs. Brightness and texture density in particular could improve the depth perception because, according to Nagata (1991), brightness and texture perception are more sensitive to depth differences than relative size. The effect of the texture gradient was not investigated in this dissertation. Hou (2001) observed that texture density affected the alignment accuracy of a stereoscopic pointer. A more dense texture gave more accurate results. Thus, highly textured AAs should be used in tasks that require high accuracy. However, the texture density is limited by the resolution of the display and thus the perception of fine texture decreases at distances within the action space. In addition, shading and aerial perspective were not used in this dissertation. By also accounting for these constant and unused depth cues, the proposed approach can be expected to be even more efficient.

In stereoscopic systems, the AAs can be dynamically and adaptively added to a scene by searching strong correspondences between the left and right views around the AOI. Furthermore, if the scene is modeled with sufficient detail in the neighborhood of correspondences, the AAs can be added with correct occlusions. Occlusion is the most dominant depth cue (Cutting and Vishton, 1995); thus, detecting physical objects that occlude virtual objects is a very efficient way of anchoring the virtual objects to the scene. With stereoscopic systems, the measurement accuracy is theoretically decreased proportionally to the squared distance (see Equation 3.4). Thus, the measurement of physical objects is more accurate at close distances, which has been utilized in AR applications by Zhu et al. (e.g., 2010). This property of stereoscopic systems encourages the addition of AAs at close distances. Adding AAs at close distances and using relative size as a depth cue utilizes the stereoscopic measurement and human depth perception in an optimal way. The measurement accuracy at close distances is high, and the human ability to detect depth differences accurately based on relative size remains high with longer distances (Cutting and Vishton, 1995).

Spatial guidelines were investigated for the positions of AAs. In the X-ray visualization scene type, it was experimentally found that the AAs should be as close as possible to the AOIs to reduce the performance time of depth judgments. The other spatial guidelines for adding AAs were defined based on analyzing the literature. The following spatial guidelines were found. The AAs should be be added to the scene:

- by anchoring them to the physical world with shadows or occlusion,
- by using the same size as the AOI,
- by placing them in front of and behind of the AOI,
- by placing them within Panum's fusional range (if this is not possible, the depth position should be selected according to the limits for comfortable viewing),
- by directing the gaze of the observer to the ground level, and
- by making them as horizontally close as possible to the AOI to avoid unnecessary eye or head movements.

Generalizing the results of the effect of AAs on depth judgments with other AR systems is problematic. The components of AR systems, such as cameras, tracking, computer graphics and displays, vary between studies and all influence depth perception, as described in Chapter 3. To enable comparisons between different AR visualization methods, a common test environment should be developed. However, using only one system as a platform may bias the development of systems and visualization techniques. Our studies in PII and PIII were conducted using two different systems, environments and methods. The results from both studies showed that the auxiliary augmentations decreased the depth thresholds and yielded more accurate depth judgments.

The use of the AA approach to reduce depth errors is expected to be applicable to handheld systems that typically suffer from inaccurate depth judgments. The error rates are too high for applications such as indoor wayfinding. In addition to problems in visualization techniques, the current sensors in mobile phones (e.g., GPS, gyroscope, accelerometer, compass) are not adequate for accurate tracking of the user, and thus the augmentations are not accurately aligned to the field of view of the observer. The AA approach relies on accurately anchoring at least one augmentation to the scene. By overlaying a virtual reference object on the scene whose position the observer perceives correctly, the position of other virtual objects can be determined. Depth sensors have become available for handheld devices that enable measuring the depths of physical objects at near distances. This ability allows the alignment of AAs at close depths with correct occlusions, which has been shown to be an efficient approach for anchoring the AA (PIII).

The auxiliary augmentation approach is also expected to be useful in virtual reality, where incorrect depth perception has also been a problem (e.g., Hu et al., 2002; Willemsen et al., 2008, 2009; Grechkin et al., 2010). In a virtual environment, the space is computationally known, and thus the AAs can be added to the scene with full control of occlusions and shadows.

In this dissertation, the performances of psychophysical tasks were measured with verbal and motoric responses. To extend the response domain, eye tracking could be included in future studies. This approach would provide more information for defining the spatial guidelines for stereoscopic representation. It would be interesting to see how the AAs effect eye movements when judging distances. With eye tracking it would be possible to research where people look at while judging distances. This would provide more information for where to add AAs. In this dissertation, the positions of the AAs were pre-defined. Adding them dynamically to the scene with correct occlusions is a computer vision problem, in which the correspondences between left and right views should be found and the edges of occluding objects defined. Developing such a system is a challenge for future studies.

O4: Establish the visual performance of perception through stereoscopic systems within the action space

Stereoscopic viewing has mainly been utilized in applications in the personal space, but this dissertation showed improvement in depth judgments when augmented objects were incorporated within the action space. The performance potential of the S3D system was investigated in three scene types: augmented objects above the ground plane, X-ray visualization and urban crowds. The visual performance with stereoscopic system was increased compared with a non-stereoscopic system in the case of all studied scene types as follows (summarized in Table 5.1):

– Augmented objects above the ground plane

The depth perception of objects that are not connected to the ground level seems to be markedly improved with stereoscopic viewing. The depth judgments were less affected by height in the visual field with stereoscopic viewing. Stereoscopic viewing improved the depth perception with and without AAs. Previous studies within the action space have used objects on the ground level (e.g., Willemsen et al., 2009; Grechkin et al., 2010). In virtual reality, in the study by Willemsen et al. (2009), the slopes of the judgments with stereoscopic viewing ranged from 0.46 to 0.65 depending on the judgment protocol. With stereoscopic perception, the slopes of our judgments were 0.56 at a height of 0.5 m and 0.51 at a height of 1 m.

Beyond the selected distance range of our study (over 10 m), the effect of stereoscopic perception on depth judgments is more likely to be reduced, as the slope for stereoscopic perception seems to decline at an 8 m distance. Grechkin et al. (2010) found no declination of slope (0.70) at distances of 6 m to 18 m when objects were aligned at ground level for virtual and augmented reality conditions. A similar effect can be achieved for objects above the ground plane with AAs. When the AAs were present in the scene, the declination of the slope was not observed. These results agree with the case of X-ray visualization, in which the AA had a similar influence. Livingston, Zhuming, Swan II and Smallman (2009) studied depth perception indoors and outdoors in a matching task. The physical object was placed on the ground, and the virtual object was matched to the corresponding distance. The results for the indoor condition were similar to ours. The average signed error at 10 m distance was approx. -1.5 m, which is the same result as our result for the stereo without AA condition. In our experiment, for the AA condition, the error at 10 m was reduced to less than 0.5 m. It is expected that the AA approach will also decrease errors with optical see-through displays. The relative size seems to help in scaling the egocentric distance, which facilitates the perception of disparity. Allison et al. (2009) found a similar influence with natural viewing within the action space.

As for autostereoscopic displays on mobile devices, Kerber et al. (2013) observed that the relative size cue dominated the depth threshold, and binocular disparity did not decrease the depth threshold, contradicting our results from studies in PII and PIII. The reasons stereoscopic perception did not decrease the depth threshold in the study by Kerber et al. (2013) are most likely the unnaturally large viewing distance (1 m) for mobile phones and the quality of the autostereoscopic display. Typically, autostereoscopic displays are sensitive to viewing position, as crosstalk increases when moving away from the intended viewing distance.

– X-ray visualization

In the ambiguous X-ray visualization case, the stereoscopic perception improved depth judgments with the relative size cue compared with using relative size alone. Without AAs, stereoscopic viewing did not statistically improve the depth judgments.

It has been shown that stereoscopic depth perception is dependent on the evaluated egocentric viewing distance (Johnston, 1991). If the egocentric viewing distance is misevaluated, then the relative depth interpretation is biased. The use of AA could help in calibrating the egocentric viewing distance, in which case the relative depth would be more accurately perceived. Ellis and Menges (1998) speculated that convergence distance would influence depth judgments when a graphical object is occluded by a physical object. It can be expected that the presence of AA helps the observer converge to the correct distance, thus making the relative depth judgment more accurate. Additionally, a change in absolute disparity does not offer a detectable change of depth with continuous movement; therefore, the absolute disparity cannot be considered an effective depth cue alone (Erkelens and Collewijn, 1985).

- Typical outdoor scenes with crowds

  The detection accuracy of stereoscopic viewing was investigated in a novel scene type: urban crowds. The perception of crowds was significantly improved with stereoscopic viewing. This improved ability can be explained with the increased ability to segregate partially occluded areas with stereoscopic viewing. Previous experiments on the stereoscopic detection of half-occluded objects were conducted with random dot stereograms, but the current study showed an increased performance in natural scenes with crowd perception. In addition, the perceived ease of the task and the perceived ease in judging distances between humans were increased. These results suggest that video-based surveillance could benefit from stereoscopic systems.

  In crowd perception, it would be worth studying how the search strategies differ between stereoscopic and non-stereoscopic presentations. Participants should also be asked to judge the actual distance between humans. The perceived ease of perceiving distances between humans might not correlate with the actual distance judgments.

Table 5.1. The summary of the increased visual performance of stereoscopic systems compared with non-stereoscopic systems.

Publication	Scene type	Statistically significant effects of stereoscopic viewing compared with monoscopic viewing
II	AR: Objects above ground plane	Stereoscopic viewing improved the depth judgments (depth threshold) with and without AAs. The confidence of judgments was also increased.
III	AR: X-ray visualization	Stereoscopic viewing improved the depth judgments (depth threshold and depth magnitude) with and AAs. The confidence of the judgments was also increased.
IV	Urban crowds	The counting task was performed more accurately with stereoscopic viewing. The task was also perceived as easier with stereoscopic viewing.

O5: Discover the effect of camera separation on the viewing experience of stereoscopic photographs

The typical imaging scenes were imaged with camera separations from 2 to 10 cm, and the influences of geometric factors that depend on camera separation were explored. The results from the subjective tests indicated that the strength of depth sensation increased as a function of camera separation, which was expected based on the geometry of the stereoscopic imaging. The participants were able to perceive the change of depth scale even though the images were shown in random order without a reference depth scale. The depth was perceived equally strongly between contents even though the length and position of the disparity ranges varied. The most natural depth perception and best viewing experience was achieved with camera separations of 2 cm to 6 cm. With these camera separations, the disparity range was primarily below 1°.

The results showed that the roundness factor did not predict the naturalness of depth sensation. The roundness factor limits for stereoscopic cinema production (Mendiburu, 2009) do not apply for smaller display sizes. Yamanoue et al. (2006) found a relationship between the cardboard effect and the roundness factor when participants were explicitly asked to evaluate the thickness of a particular object in a S3D image. However, the role the roundness factor plays in the naturalness of depth sensation remains an open question. Further studies are needed to determine how much the roundness factor affects evaluations of the naturalness of depth sensation and how the different display sizes and viewing distances affect the emergence of the cardboard effect. In this study, desktop-sized displays were used but with smaller displays the cardboard effect is more likely to occur because of the higher non-linearity between perceived and scene depths.

The results are promising for stereoscopic content production in mobile conditions. A positive viewing experience and naturalness of depth were achieved even at short camera separations (2 to 6 cm). With these camera separations, visual discomfort from a mismatch between accommodation and convergence is unlikely to be a problem, and the cardboard effect (quantified with roundness factor) will not negatively affect the naturalness of depth sensation. It seems that to be perceived as natural, the depth sensation must originate from the scene's depth variation itself, not from wide camera separation. This result is in the line with findings from recent study by Hakala et al. (2014), and the result can be used when designing stereoscopic systems.

The display may have an effect due to crosstalk biasing preferences for lower disparities. Quantifying and reporting crosstalk with the results would have made them more generalizable. Unfortunately, such measuring devices were not available for the studies. However, all the equipment used is documented and available for further evaluation.

Beyond the practical implications discussed in the list above, the dissertation has theoretical implications for cue theory. Overall, the results confirm that the more depth cues available, the closer to veridical the depth judgments can be made. This is in line with the weak observer theory, in which depth perception is supported by depth cues with different weights (Landy et al., 1995) (discussed in Section 2.1). To be more specific, the results support accumulative depth cue integration. Accumulative depth cue integration occurs if the influence of the depth cue with the highest accuracy is increased by adding another depth cue. It has been previously observed that in the presence of a limited number of pictorial depth cues, the accumulative depth cue integration is the most probable (Bruno and Cutting, 1988). The results related to PII and PIII of this dissertation support these observations: binocular disparity and relative size showed accumulative depth cue integration. This result is in a line with a study by Holway and Boring (1941), where depth judgments relying on relative size were improved with binocular disparity. In addition, relative size has been found to interact additively with height in the visual field and motion parallax (Bruno and Cutting, 1988), which makes it a highly applicable depth cue for AR scenes.

The contributions are related to higher level design approaches and guidelines, and the dissertation offers insights into computational models for designing stereoscopic systems. However, constructing a computational model for perceived depth would require the characterization of each component of the system, as defined in Figure 3.1. The number of design parameters of stereoscopic systems affecting depth perception is large, and thus, evaluating their effect on performance requires substantially more work than was feasible in the context of this dissertation. Most likely there are undiscovered interaction mechanisms in the pictorial, mechanical and scene parameters that interact with stereoscopic perception. More research is needed to understand these relations more deeply.

6. Conclusions

Depth perception through stereoscopic systems is affected by numerous sources of errors, although stereoscopic systems are expected to be superior compared with non-stereoscopic systems. The aim of the dissertation was to determine to what extent this is the case and, moreover, and to find how the stereoscopic systems should be designed. Consistent with the mission to extend the application of stereoscopy to new domains, we focused on stereoscopic systems in the cases of augmented and natural scenes within the action space.

In general, it can be concluded that not applying the potential of stereoscopic perception in applications within the action space is a waste of human visual capabilities. Binocular disparity can convey the depth even without pictorial depth cues, allowing interaction to be reproduced in a more accurate and constant manner. Overall, in tasks in which a visually complex scene needs to be interpreted without ambiguity, the benefits of stereoscopic viewing are apparent. However, existing AR systems do not utilize the possibilities of stereoscopy widely, nor does photography. The cost of stereoscopic system components has decreased, and low cost stereo cameras have become available. Such cameras coupled with head-mounted or hand-held displays can be used for measuring and showing scenes to observers. This development enables the construction of stereoscopic systems that combine the different application domains studied in the dissertation.

The studies in the dissertation showed that the depth judgment errors can be reduced markedly with appropriate design. Based on the observations on perceived and measured depth, the dissertation presented design guidelines for stereoscopic systems. These were given for selecting stereoscopic imaging parameters, visualizing the depth of augmentations and using stereoscopic imaging to perceive crowds.

It is difficult to fulfill the need for stereoscopic systems with the functionalities of displaying the stereoscopic image with a high-quality viewing experience and measuring the depths in the scene accurately. This study demonstrated methods and results for designing such systems. As a key factor in controlling depth through stereoscopic systems, the camera separation was varied to find its effect on stereoscopic measurement and the viewing experience. The highest subjective evaluations were obtained using short camera separations. From the viewing experience point of view this result is promising, as viewing comfort is not an issue with short camera separations, and short separations are easy to implement in hand-held devices. From the measurement point of view, this result is less encouraging, as the measurement accuracy decreases for short camera separations. However, the results showed that the effect of camera separation on the measurement accuracy was less than theoretically expected.

In AR, the observer interacts with the surrounding 3D world and thus the utilization of stereoscopic perception through a stereoscopic AR system is considered important. Stereoscopic perception enhances the depth perception in AR within the action space. The major finding was that the depth judgments were substantially improved with the use of reference virtual objects, called auxiliary augmentations. They provided interaction between the augmentations and the physical world as well as a relative size cue among augmentations. Both theoretical and practical reasons support the generalization of the auxiliary augmentation approach to VR and handheld systems.

The stereoscopic system improved the perception of crowds compared with the non-stereoscopic condition due to the increased ability to segregate partially occluded objects. This showed that stereoscopic systems could be used in surveillance applications, where crowd perception is an important topic. Stereoscopic systems have origins from the 19th century and the usage of systems have varied from endoscopic surgery to interpreting aerial images. Other new application areas will likely to emerge for stereoscopic systems as the developments in head-mounted and autostereoscopic display technologies will continue.

Bibliography

Allison, R., Gillam, B. and Vecellio, E. (2009), 'Binocular depth discrimination and estimation beyond interaction space', *Journal of Vision* **9**(1), 10:1–14.

Arthur, K. W. (2000), Effects of Field of View on Performance with Head-Mounted Displays, PhD thesis, University of North Carolina.

Avery, B., Sandor, C. and Thomas, B. H. (2009), Improving Spatial Perception for Augmented Reality X-Ray Vision, *in* 'Virtual Reality Conference', IEEE, Lafayette, LA, pp. 79–82.

Avery, B., Thomas, B. H. and Piekarski, W. (2008), User evaluation of see-through vision for mobile outdoor augmented reality, *in* 'Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality', ISMAR '08, pp. 69–72.

Azuma, R. (1997), 'A survey of augmented reality', *Presence-Teleoperators and Virtual Environments* **4**(August), 355–385.

Badcock, D. R. and Schor, C. M. (1985), 'Depth-increment detection function for individual spatial channels.', *Journal of the Optical Society of America. A, Optics and image science* **2**(7), 1211–6.

Bane, R. and Hollerer, T. (2004), Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality, *in* 'IEEE/ACM International Symposium on Mixed and Augmented Reality', ISMAR '04, Washington DC, pp. 231–239.

Barfield, W., Hendrix, C. and Bystrom, K.-E. (1999), 'Effects of Stereopsis and Head Tracking on Performance Using Desktop', *Presence* **8**(2), 237–240.

Barr, A. and Stroud, W. (1924), 'Stereoscopic range finder, US Patent 1,514,948, 4p.'.

Bian, Z. and Andersen, G. J. (2011), 'Environmental surfaces and the compression of perceived visual space', *Journal of vision* **11**(7), 4:1–14.

Blakemore, C. (1970), 'The range and scope of binocular depth discrimination in man', *The Journal of Physiology* **211**, 599–622.

Bowersox, J. C., Cordts, P. R. and LaPorta, a. J. (1998), 'Use of an intuitive telemanipulator system for remote trauma surgery: an experimental study.', *Journal of the American College of Surgeons* **186**(6), 615–21.

Brewster, D. (1856), *The Stereoscope; its History, Theory, and Construction, with its Application to the fine and useful Arts and to Education: With fifty wood Engravings*, John Murray, London.

Bruno, N. and Cutting, J. E. (1988), 'Minimodularity and the perception of layout.', *Journal of experimental psychology* **117**(2), 161–70.

Bülthoff, H. H. and Mallot, H. a. (1988), 'Integration of depth modules: stereo and shading.', *Journal of the Optical Society of America. A, Optics and image science* **5**(10), 1749–58.

Burt, P. and Julesz, B. (1980), 'A disparity gradient limit for binocular fusion.', *Science* **208**(4444), 615–617.

Cakmakci, O. and Rolland, J. (2006), 'Head-Worn Displays: A Review', *Journal of Display Technology* **2**(3), 199–216.

Chan, H.-P., Goodsitt, M. M., Helvie, M. A., Hadjiiski, L. M., Lydick, J. T., Roubidoux, M. A., Bailey, J. E., Nees, A., Blane, C. E. and Sahiner, B. (2005), 'ROC study of the effect of stereoscopic imaging on assessment of breast lesions', *Medical Physics* **32**(4), 1001–1009.

Changizi, M. A. and Shimojo, S. (2008), '"X-ray vision" and the evolution of forward-facing eyes.', *Journal of theoretical biology* **254**(4), 756–67.

Chen, W.-C., Hsiao, F.-J. and Lin, C.-W. (2010), An automatic parallax adjustment method for stereoscopic augmented reality systems, *in* 'IEEE International Symposium on Mixed and Augmented Reality', IEEE, pp. 215–216.

Collett, T. S. (1985), 'Extrapolating and Interpolating Surfaces in Depth', *Proceedings of the Royal Society of London. Series B. Biological Sciences* **224**(1234), 43 –56.

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A. and Thompson, W. B. (2005), 'The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Environments', *Perception* **34**(3), 191–204.

Cutting, J. E. (1997), 'How the eye measures reality and virtual reality', *Behavior Research Methods Instruments and Computers* **29**(1), 27–36.

Cutting, J. E. and Vishton, P. M. (1995), Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth, *in* W. Epstein and S. Rogers, eds, 'Handbook of perception and cognition', Vol. 5, Academic press, San Diego, CA., pp. 69–117.

de Ridder, H., Blommaert, F. J. J. and Fedorovskaya, E. A. (1995), Naturalness and image quality: chroma and hue variation in color images of natural scenes, *in* B. E. Rogowitz and J. P. Allebach, eds, 'IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology', International Society for Optics and Photonics, pp. 51–61.

De Ridder, H. and Endrikhovski, S. (2002), 'Invited Paper: Image Quality is FUN: Reflections on Fidelity, Usefulness and Naturalness', *SID Symposium Digest of Technical Papers* **33**(1), 986–989.

Dey, A., Cunningham, A. and Sandor, C. (2010), Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments, *in* 'ACM Symposium on Virtual Reality Software and Technology', VRST '10, pp. 211–218.

Dey, A., Jarvis, G., Sandor, C. and Reitmayr, G. (2012), Tablet versus Phone: Depth Perception in Handheld Augmented Reality, *in* 'IEEE International Symposium on Mixed and Augmented Reality', pp. 187–196.

Ditzinger, T., Stadler, M., Strüber, D. and Kelso, J. a. (2000), 'Noise improves three-dimensional perception: stochastic resonance and other impacts of noise to the perception of autostereograms.', *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics* **62**(2 Pt B), 2566–75.

Dodgson, N. A. (2004), Variation and extrema of human interpupillary distance, *in* A. J. Woods, J. O. Merritt, S. A. Benton and M. T. Bolas, eds, 'Proc SPIE Vol. 5291', International Society for Optics and Photonics, pp. 36–46.

Dodgson, N. A. (2013), '3D without the glasses', *Nature* **495**(7441), 316–317.

Dosher, B. A., Sperling, G. and Wurst, S. A. (1986), 'Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure', *Vision Research* **26**(6), 973–990.

Drascic, D. (1991), Skill Acquisition and Task Performance in Teleoperation Using Monoscopic and Stereoscopic Video Remote Viewing, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting vol. 35', number 19, SAGE Publications, pp. 1367–1371.

Drascic, D. and Grodski, J. J. (1993), Using Stereoscopic Video for Defence Teleoperation, *in* 'SPIE Stereoscopic Displays and Applications', Vol. 1915, pp. 58–69.

Drascic, D. and Milgram, P. (1991), Pointing accuracy of a virtual stereographic pointer in a real stereoscopic video world, *in* 'SPIE Stereoscopic Displays and Applications II', Vol. 1457, pp. 623–626.

Drascic, D. and Milgram, P. (1996), Perceptual Issues in Augmented Reality Reality, *in* 'Proc. SPIE Stereoscopic Displays and Applications', Vol. 2653, pp. 123–134.

Durgin, F. H., College, S., Ave, C., Leonard-solis, K., Masters, O. and Schmelz, B. (2012), 'Expert performance by athletes in the verbal estimation of spatial extents does not alter their perceptual metric of space', *i-Perception* **3**(5), 357–367.

Eby, D. W. and Loomis, J. M. (1987), 'A study of visually directed throwing in the presence of multiple distance cues.', *Perception & psychophysics* **41**(4), 308–12.

Egusa, H. (1981), 'Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field', *Perception* **12**(2), 167–175.

Ellis, S. R. and Menges, B. M. (1998), 'Localization of virtual objects in the near visual field', *Human factors* **40**(3), 415–431.

Engeldrum, P. (1999), Image quality modeling: Where are we?, *in* 'IS&T ANNUAL CONFERENCE', pp. 251–255.

Ennis, C., Peters, C. and O'Sullivan, C. (2011), 'Perceptual effects of scene context and viewpoint for virtual pedestrian crowds', *ACM Transactions on Applied Perception* **8**(2), 1–22.

Erkelens, C. J. and Collewijn, H. (1985), 'Motion perception during dichoptic viewing of moving random-dot stereograms.', *Vision research* **25**(4), 583–588.

Feiner, S., Webster, A., Krueger, T., MacIntyre, B. and Keller, E. (1995), 'Architectural Anatomy', *Presence: Teleoperators and Virtual Environments* **4**(3), 318–325.

Ferrari, V., Megali, G., Troia, E., Pietrabissa, A. and Mosca, F. (2009), 'A 3-D mixed-reality system for stereoscopic visualization of medical dataset.', *IEEE transactions on bio-medical engineering* **56**(11), 2627–33.

Fischer, J., Cunningham, D., Bartz, D., Wallraven, C., Bulthoff, H. and Strasser, W. (2006), Measuring the Discernability of Virtual Objects in Conventional and Stylized Augmented Reality, *in* 'Eurographics Symposium on Virtual Environments', pp. 53–61.

Foley, J. M. (1980), 'Binocular distance perception', *Psychological Review* **87**, 411–434.

Foley, J. M., Ribeiro-Filho, N. P. and Da Silva, J. a. (2004), 'Visual perception of extent and the geometry of visual space', *Vision Research* **44**(2), 147–156.

Fukusima, S. S., Loomis, J. M. and Da Silva, J. a. (1997), 'Visual perception of egocentric distance as assessed by triangulation.', *Journal of experimental psychology. Human perception and performance* **23**(1), 86–100.

Furmanski, C., Azuma, R. and Daily, M. (2002), Augmented-Reality Visualizations Guided by Cognition: Perceptual Heuristics for Combining Visible and Obscured Information, *in* 'International Symposium on Mixed and Augmented Reality', IEEE Computer Society, Darmstadt, pp. 215–225.

Gabbard, J. L., Tech, V., Zedlitz, J., Ii, J. E. S. and Iii, W. W. W. (2010), More Than Meets the Eye : An Engineering Study to Empirically Examine the Blending of Real and Virtual Color Spaces, *in* 'IEEE Virtual Reality Conference', pp. 79–86.

Geisler, W. S. (2008), 'Visual perception and the statistical properties of natural scenes.', *Annual review of psychology* **59**, 167–92.

Gescheider, G. A. (1997), *Psychophysics, The Fundamentals*, 3. edn, New Jersey.

Getty, D., D'Orsi, C. and Pickett, R. (2008), 'Stereoscopic digital mammography: Improved accuracy of lesion detection in breast cancer screening', *Digital Mammography. Lecture Notes in Computer Science* **5116**, 74–79.

Gibson, J. (1950), *The perception of the visual world*, Houghton Mifflin, Oxford.

Gilinsky, A. S. (1951), 'Perceived size and distance in visual space', *Psychological Review* **58**, 460–482.

Gogel, W. C. (1976), 'An indirect method of measuring', *Perception & psychophysics* **20**(6), 419–429.

Goldmann, L., De Simone, F. and Ebrahimi, T. (2010), Impact of acquisition distortions on the quality of stereoscopic images, *in* 'International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)'.

Goodsitt, M. M., Chan, H.-P. and Hadjiiski, L. (2000), 'Stereomammography: Evaluation of depth perception using a virtual 3D cursor', *Medical Physics* **27**(6), 1305.

Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F. and Kearney, J. K. (2010), 'How does presentation method and measurement protocol affect distance estimation in real and virtual environments?', *ACM Transactions on Applied Perception* **7**(4), 1–18.

Greenberg, S., Marquardt, N., Diaz-marino, R. and Wang, M. (2011), 'Proxemic Interactions: The New Ubicomp?', *ACM Interactions* **18**(1), 42–50.

Hakala, J., Nuutinen, M. and Oittinen, P. (2011), Interestingness of stereoscopic images, *in* 'SPIE, Stereoscopic Displays and Applications', San Francisco, California, USA, pp. 78631S–78631S–11.

Hakala, J., Westman, S., Salmimaa, M., Pölönen, M., Järvenpää, T. and Häkkinen, J. (2014), 'Why 3D Cameras are Not Popular: A Qualitative User Study on Stereoscopic Photography Acceptance', *3D Research* **5**(1), 4:1–18.

Häkkinen, J., Hakala, J., Hannuksela, M. and Oittinen, P. (2011), Effect of image scaling on stereoscopic movie experience, *in* 'Stereoscopic Displays and Applications, Proc. SPIE vol. 7863'.

Häkkinen, J., Kawai, T., Takatalo, J., Leisti, T., Radun, J., Hirsaho, A. and Nyman, G. (2008), Measuring stereoscopic image quality experience with interpretation based quality methodology, *in* 'Proc. SPIE', Vol. 6808, pp. 68081B–68081B–12.

Halpern, D. L. and Blake, R. B. (1988), 'How contrast affects stereoacuity', *Perception* **17**(4), 483 – 495.

Hanna, G. B., Shimi, S. M. and Cuschieri, A. (1998), 'Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy.', *Lancet* **351**(9098), 248–51.

Harris, C. and Stephens, M. (1988), A combined corner and edge detector, *in* 'Alvey vision conference', Vol. 15, pp. 147–151.

Harris, J. M. and Wilcox, L. M. (2009), 'The role of monocularly visible regions in depth and surface perception.', *Vision research* **49**(22), 2666–85.

Hayhoe, M., Gillam, B., Chajka, K. and Vecellio, E. (2009), 'The role of binocular vision in walking', *Visual Neuroscience* **26**(01), 73–80.

Heike, C. L., Upson, K., Stuhaug, E. and Weinberg, S. M. (2010), '3D digital stereophotogrammetry: a practical guide to facial image acquisition.', *Head & Face Medicine* **6**(1), 18:1–11.

Helmholz, H. (1867), *Handbuch der physiologischen Optik*, 1 edn.

Hendrix, C. and Barfield, W. (1995), 'Relationship between monocular and binocular depth cues for judgements of spatial information and spatial instrument design', *Displays* **16**(3), 103–113.

Henson, D. B. and Williams, D. E. (1980), 'Depth perception in strabismus.', *The British journal of ophthalmology* **64**(5), 349–53.

Holway, A. H. . and Boring, E. G. . (1941), 'Determinants of Apparent Visual Size with Distance Variant Source', *The American Journal of Psychology* **54**(1), 21–37.

Hou, M. (2001), User experience with alignment of real and virtual objects in a stereoscopic augmented reality interface, *in* 'Centre for Advanced Studies on Collaborative research', IBM Press.

Hou, M. and Milgram, P. (2000), Effect of Surface Characteristics on Alignment of Graphic and Real Objects in a Stereoscopic Augmented Reality Environment, *in* 'Human Factors and Ergonomics Society Annual Meeting Proceedings', Vol. 44, pp. 476–479.

Howard, H. J. (1919), 'A test for the judgment of distance', *Transactions of the American Ophthalmological Society* **17**, 195–235.

Howard, I. P. (2012*a*), *Perceiving in depth Volume 1: Basic Mechanisms*, Oxford university press, New York.

Howard, I. P. (2012*b*), *Perceiving in depth Volume 3: Other mechanisms of depth perception*.

Howard, I. P. and Rogers, B. J. (2012), *Perceiving in depth Volume 2: Stereoscopic vision*, Oxford university press, New York.

Hsu, J., Babbs, C. F., Pizlo, Z. and Delp, E. J. (1993), A Study of the Effectiveness of Stereo Imaging with Applications in Mammography, *in* 'Proc. SPIE 1913, Human Vision, Visual Processing, and Digital Display IV', Vol. 1913, pp. 154–165.

Hu, B. and Knill, D. C. (2011), 'Binocular and monocular depth cues in online feedback control of 3D pointing movement.', *Journal of vision* **11**(7), 23:1–13.

Hu, H. H., Gooch, A. A., Creem-Regehr, S. H. and Thompson, W. B. (2002), 'Visual Cues for Perceiving Distances from Objects to Surfaces', *Presence: Teleoperators and Virtual Environments* **11**(6), 652–664.

Huang, X., Liyuan Li and Sim, T. (2004), Stereo-based human head detection from crowd scenes, *in* 'International Conference on Image Processing', Vol. 2, pp. 1353–1356.

Hubona, G. S., Wheeler, P. N., Shirah, G. W. and Brandt, M. (1999), 'The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization', *ACM Transactions on Computer-Human Interaction (TOCHI)* **6**(3), 214–242.

Ijsselsteijn, W. (2000), 'Subjective Evaluations of Stereoscopic Images: Effects of Camera Parameters and Display Duration', *IEEE Transactions on Circuits and Systems For Video Technology* **10**(2), 225–233.

IJsselsteijn, W., de Ridder, H., Huib, J., Avons, S. E. and Bouwhuis, D. (2001), 'Effects of Stereoscopic Presentation, Image Motion, and Screen Size on Subjective and Objective Corroborative', *Presence* **10**(3), 298–311.

Jameson, D. and Hurvich, L. M. (1959), 'Note on Factors Influencing the Relation between Stereoscopic Acuity and Observation Distance', *Journal of the Optical Society of America* **49**(6), 639–639.

Jennings, J. and Charman, N. (1994), 'Depth resolution in stereoscopic systems', *Applied optics* **33**(22), 5192–5196.

Jerome, C. J. and Witmer, B. G. (2005), The Perception and Estimation of Egocentric Distance in Real and Augmented Reality Environments, *in* 'Proceedings of the Human Factors and Ergonomics Society Annual Meeting', pp. 2249–2252.

Johnston, E. B. (1991), 'Systematic distortions of shape from stereopsis', *Vision Research* **31**(7–8), 1351–1360.

Jones, G., Lee, D., Holliman, N. and Ezra, D. (2001), Controlling perceived depth in stereoscopic images, *in* 'Proc. SPIE 4297', Vol. 4297, pp. 42–53.

Jones, J. A., Swan, I. J., Singh, G., Kolstad, E. and Ellis, S. R. (2008), The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception, *in* 'Symposium on Applied perception in graphics and visualization', number 212, pp. 9–14.

Jones, J. A., Swan II, J. E. and Ellis, S. R. (2011), Peripheral Visual Information and Its Effect on Distance Judgments in Virtual and Augmented Environments, *in* 'ACM SIGGRAPH Applied Perception in Graphics and Visualization (APGV)', pp. 29–35.

Jones, R. K. and Lee, D. N. (1981), 'Why two eyes are better than one: the two views of binocular vision.', *Journal of experimental psychology. Human perception and performance* **7**(1), 30–40.

Julesz, B. (1960), 'Binocular depth perception of computer generated patterns', *Bell Systems Technical Journal* **39**, 1125–1162.

Julesz, B. (1971), *Foundations of cyclopean perception*, Chigago Press, Oxford, England.

Jurgens, V., Cockburn, A. and Billinghurst, M. (2006), Depth cues for augmented reality stakeout, *in* 'Proceedings of the 7th ACM SIGCHI New Zealand chapter's international conference on Computer-human interaction: design centered HCI', CHINZ '06, pp. 117–124.

Kalkofen, D., Sandor, C., White, S. and Schmalstieg, D. (2011), Visualization Techniques for Augmented Reality, *in* B. Furht, ed., 'Handbook of Augmented Reality', Springer New York, New York, NY, pp. 65–98.

Kán, P. and Kaufmann, H. (2012), Physically-Based Depth of Field in Augmented Reality, *in* 'Eurographics', pp. 89–92.

Kelly, J. (2011), 'How 3D glasses helped defeat Hitler', *BBC News Magazine* .

Kerber, F., Lessel, P., Mauderer, M., Daiber, F., Oulasvirta, A. and Kr, A. (2013), Is Autostereoscopy Useful for Handheld AR ?, *in* 'International Conference on Mobile and Ubiquitous Multimedia', pp. 2–5.

Klein, G. and Murray, D. (2007), Parallel Tracking and Mapping for Small AR Workspaces, *in* 'IEEE and ACM International Symposium on Mixed and Augmented Reality', IEEE, pp. 1–10.

Klein, G. and Murray, D. (2008), Compositing for small cameras, *in* '2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality', Ieee, pp. 57–60.

Knapp, J. M. and Loomis, J. M. (2003), 'Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual', *Presence: Teleoperators & Virtual Environments* **13**(5), 572–577.

Knill, D. C. (2005), 'Reaching for visual cues to depth : The brain combines depth cues differently for motor control and perception', *Journal of Vision* **5**(2), 103–115.

Kooi, F. and Toet, A. (2004), 'Visual comfort of binocular and 3D displays', *Displays* **25**(2-3), 99–108.

Kruijff, E., Swan, J. and Feiner, S. (2010), Perceptual issues in augmented reality revisited, *in* 'IEEE International Symposium on Mixed and Augmented Reality', Seoul, pp. 3–12.

Kuhl, S. a., Thompson, W. B. and Creem-Regehr, S. H. (2009), 'HMD calibration and its effects on distance judgments', *ACM Transactions on Applied Perception* **6**(3), 19:1–20.

Lambooij, M., IJsselsteijn, W., Fortuin, M. and Heynderickx, I. (2009), 'Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review', *Journal of Imaging Science and Technology* **53**(3), 030201–14.

Landy, M., Banks, M. and Knill, D. (2011), Ideal-Observer Models of Cue Integration, *in* 'Sensory cue integration', Oxford university press, pp. 5–29.

Landy, M. S., Maloney, L. T., Johnston, E. B. and Young, M. (1995), 'Measurement and modeling of depth cue combination: in defense of weak fusion', *Vision Research* **35**(3), 389–412.

Leibe, B., Seemann, E. and Schiele, B. (2005), Pedestrian Detection in Crowded Scenes, *in* 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)', Vol. 1, IEEE, pp. 878–885.

Leske, D. A. and Holmes, J. M. (2004), 'Maximum angle of horizontal strabismus consistent with true stereopsis.', *Journal of AAPOS : the official publication of the American Association for Pediatric Ophthalmology and Strabismus / American Association for Pediatric Ophthalmology and Strabismus* **8**(1), 28–34.

Lipton, L. (2012), 'Brief history of electronic stereoscopic displays', *Optical Engineering* **51**(2), 021103.

Livingston, M. A. and Ai, Z. (2008), The effect of registration error on tracking distant augmented objects, *in* 'IEEE/ACM International Symposium on Mixed and Augmented Reality', IEEE, pp. 77–86.

Livingston, M. A., Dey, A., Sandor, C. and Thomas, B. H. (2013), Pursuit of "X-ray Vision" for Augmented Reality, *in* W. Huang, L. Alem and M. A. Livingston, eds, 'Human Factors in Augmented Reality Environments', Springer, chapter 4, pp. 67–107.

Livingston, M. A., Swan II, J. E., Gabbard, J. L., Höllerer, T. H., Hix, D., Julier, S. J., Baillot, Y. and Brown, D. (2003), Resolving Multiple Occluded Layers in Augmented Reality, *in* 'IEEE/ACM International Symposium on Mixed and Augmented Reality', IEEE Computer Society, Tokyo, pp. 56–65.

Livingston, M., Zhuming, A. and Decker, J. (2009), A user study towards understanding stereo perception in head-worn augmented reality displays, *in* 'IEEE International Symposium on Mixed and Augmented Reality', pp. 53–56.

Livingston, M., Zhuming, A., Swan II, J. and Smallman, H. (2009), Indoor vs. Outdoor Depth Perception for Mobile Augmented Reality, *in* 'IEEE Virtual Reality Conference', IEEE, pp. 55–62.

Loftus, A., Servos, P., Goodale, M. A., Mendarozqueta, N. and Mon-Williams, M. (2004), 'When two eyes are better than one in prehension: monocular viewing and end-point variance', *Experimental brain research.* **158**(3), 317–27.

Loomis, J. M., da Silva, J. A., Fujita, N. and Fukusima, S. S. (1992), 'Visual space perception and visually directed action', *Journal of Experimental Psychology: Human Perception and Performance* **18**(4), 906–921.

Loomis, J. M., Philbeck, J. W. and Zahorik, P. (2002), 'Dissociation between location and shape in visual space.', *Journal of Experimental Psychology: Human Perception and Performance* **28**(5), 1202–1212.

Loomis, J. and Philbeck, J. (2008), Measuring spatial perception with spatial updating and action, *in* R. L. Klatzky, M. Behrmann and B. MacWhinney, eds, 'Embodiment, ego-space, and action', Taylor & Francis, pp. 1–43.

Masaoka, K., Hanazato, A., Emoto, M., Yamanoue, H., Nojiri, Y. and Okano, F. (2006), 'Spatial distortion prediction system for stereoscopic images', *Journal of Electronic Imaging* **15**(1), 013002–12.

Mazyn, L. I. N., Lenoir, M., Montagne, G., Delaey, C. and Savelsbergh, G. J. P. (2007), 'Stereo vision enhances the learning of a catching skill.', *Experimental brain research* **179**(4), 723–6.

Mazyn, L. I. N., Lenoir, M., Montagne, G. and Savelsbergh, G. J. P. (2004), 'The contribution of stereo vision to one-handed catching.', *Experimental brain research* **157**(3), 383–90.

McCandless, J. W., Ellis, S. R. and Adelstein, B. D. (2000), 'Localization of a time-delayed, monocular virtual object superimposed on a real environment', *Presence (Cambridge, Mass.)* **9**(1), 15–24.

McClain, J., Cacioppo, A., Reising, J. and Koubek, R. (1990), 'The interaction of chromostereopsis and stereopsis in stereoscopic CRT displays', *Applied Ergonomics* **21**(2), 101–106.

McIntire, J. P., Havig, P. R. and Geiselman, E. E. (2012), What is 3D good for? A review of human performance on stereoscopic 3D displays, *in* 'Proc SPIE vol. 8383', pp. 83830X–83830X–13.

McKee, S. P. (1983), 'The spatial requirements for fine stereoacuity', *Vision Research* **23**(2), 191–198.

McKee, S. P. and Taylor, D. G. (2010), 'The precision of binocular and monocular depth judgments in natural settings', *Journal of Vision* **10**(10), 5:1–13.

McWhorter, S. W., Hodges, L. F. and Rodriguez, W. E. (1991), Evaluation of Display Parameters Affecting User Performance of an Interactive Task in a Virtual Environment, Technical report, Georgia Institute of Technology.

Melzer, J. E., Brozoski, F. T., Letowski, T. R., Harding, T. H. and Rash, C. E. (2009), Guidelines for HMD design, *in* 'Helmet-Mounted Displays: Sensation, Perception and Cognition Issues', pp. 805–848.

Mendiburu, B. (2009), *3D movie making: stereoscopic digital cinema from script to screen*, Focal Press.

Merritt, J. O., CuQlock-Knopp, V. G., Kregel, M., Smoot, J. and Monaco, W. (2005), Perception of terrain drop-offs as a function of L-R viewpoint separation in stereoscopic video, *in* C. E. Rash and C. E. Reese, eds, 'Proc. SPIE 5800, Helmet- and Head-Mounted Displays X: Technologies and Applications', Vol. 5800, pp. 169–176.

Milgram, P. and Kishino, F. (1994), 'A taxonomy of mixed reality visual displays', *IEICE TRANSACTIONS on Information and Systems* **E77-D**(12), 1321–1329.

Milgram, P. and Krueger, M. (1992), Adaptation effects in stereo due to on-line changes in camera configuration, *in* 'Proc SPIE vol. 1669', pp. 122–134.

Mohler, B. J., Creem-Regehr, S. H. and Thompson, W. B. (2006), The influence of feedback on egocentric distance judgments in real and virtual environments, *in* 'Proceedings of the 3rd symposium on Applied perception in graphics and visualization - APGV '06', ACM Press, New York, New York, USA, pp. 9–14.

Mojsilovic, A., Gomes, J. and Rogowitz, B. (2004), 'Semantic-Friendly Indexing and Quering of Images Based on the Extraction of the Objective Semantic Cues', *International Journal of Computer Vision* **56**((1/2)), 79–107.

Mulligan, J., Isler, V. and Daniilidis, K. (2001), Performance evaluation of stereo for tele-presence, *in* 'International Conference on Computer Vision', Vol. 2, pp. 558–565.

Nagata, S. (1991), How to reinforce perception of depth in single two-dimensional pictures, *in* S. Ellis, M. Kaiser and A. Grunwald, eds, 'Pictorial communication in virtual and real environments', Taylor & Francis, pp. 527–544.

Nakayama, K. (1996), 'Binocular visual surface perception', *Proceedings of the National Academy of Sciences* **93**(2), 634–639.

Nakayama, K. and Shimojo, S. (1992), 'Experiencing and perceiving visual surfaces', *Science* **257**(5075), 1357–1363.

Nakayama, K., Shimojo, S. and Silverman, G. H. (1989), 'Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects', *Perception* **18**(1), 55–68.

Napieralski, P. E., Altenhoff, B. M., Bertrand, J. W., Long, L. O., Babu, S. V., Pagano, C. C., Kern, J. and Davis, T. A. (2011), 'Near-field distance perception in real and virtual environments using both verbal and action responses', *ACM Transactions on Applied Perception* **8**(3), 1–19.

Nelson, T. R., Ji, E. K., Lee, J. H., Bailey, M. J. and Pretorius, D. H. (2008), 'Stereoscopic evaluation of fetal bony structures.', *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine* **27**(1), 15–24.

Ogle, K. N. (1952), 'On the limits of stereoscopic vision', *Journal of Experimental Psychology* **44**(4), 253–259.

Ogle, K. N. (1963), 'Stereoscopic Depth Perception and Exposure Delay Between Images To the Two Eyes.', *Journal of the Optical Society of America* **53**(11), 1296–304.

Ooi, T. L., Wu, B. and He, Z. J. (2001), 'Distance determined by the angular declination below the horizon', *Nature* **414**(6860), 197–200.

O'Shea, R. P., Blackburn, S. G. and Ono, H. (1994), 'Contrast as a depth cue', *Vision Research* **34**(12), 1595–1604.

Oskam, T., Hornung, A., Sumner, R. W. and Gross, M. (2012), Fast and Stable Color Balancing for Images and Augmented Reality, *in* 'International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission', IEEE, pp. 49–56.

Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S. and Harris, J. M. (2010), 'Stereoscopic perception of real depths at large distances', *Journal of vision* **10**(6), 19:1–16.

Passmore, P., Nielsen, C., Cosh, W. and Darzi, A. (2001), Effects of viewing and orientation on path following in a medical teleoperation environment, *in* 'Proceedings IEEE Virtual Reality 2001', IEEE Comput. Soc, pp. 209–215.

Peterson, S., Axholt, M. and Ellis, S. R. (2008), Managing Visual Clutter: A Generalized Technique for Label Segregation using Stereoscopic Disparity, *in* 'IEEE Virtual Reality', IEEE, pp. 169–176.

Peterson, S. D., Axholt, M. and Ellis, S. R. (2009), 'Objective and subjective assessment of stereoscopically separated labels in augmented reality', *Computers & Graphics* **33**(1), 23–33.

Philbeck, J. W. and Loomis, J. M. (1997), 'Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions.', *Journal of Experimental Psychology: Human Perception and Performance* **23**(1), 72–85.

Polys, N. F., Bowman, D. A. and North, C. (2011), 'The role of Depth and Gestalt cues in information-rich virtual environments', *International Journal of Human-Computer Studies* **69**(1-2), 30–51.

Poston, A. (2000), Human Engineering Design Data Digest, Report from Department of Defense Human Factors Engineering Technical Advisory Group Washington, DC, Technical report.

Prussog, A., Muhlbach, L. and Bocker, M. (1994), 'Telepresence in Videocommunications', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **38**(3), 180–184.

Rawlings, S. C. and Shipley, T. (1969), 'Stereoscopic acuity and horizontal angular distance from fixation.', *Journal of the Optical Society of America* **59**(8), 991–3.

Richards, W. (1970), 'Stereopsis and stereoblindness', *Experimental Brain Research* **10**(4), 380–388.

Rieser, J. J., Ashmead, D. H., Talor, C. R. and Youngquist, G. A. (1990), 'Visual perception and the guidance of locomotion without vision to previously seen targets', *Perception* **19**(5), 675–689.

Rolland, J., Biocca, F., Barlow, T. and Kancherla, A. (1995), Quantification of adaptation to virtual-eye location in see-thru head-mounted displays, *in* 'Proceedings Virtual Reality Annual International Symposium '95', IEEE Comput. Soc. Press, pp. 56–66.

Rolland, J., Cakmakci, O., Lee, K.-S., Fidopiastis, C., Hamza-Lup, F. and Santhanam, A. (2007), Beyond the desktop: emerging technologies for supporting collaborative teams in virtual environments, *in* Y. Wang, T. T. Tschudi, J. P. Rolland and K. Tatsuno, eds, 'Proc. SPIE 6834, Optical Design and Testing III', International Society for Optics and Photonics, pp. 68340G–68340G–8.

Rolland, J. P. and Fuchs, H. (2000), 'Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization', *Presence* **9**(3), 287–309.

Rolland, J. P., Meyer, C., Arthur, K., Hill, C. and Rinalducci, E. (2002), 'Method of Adjustments versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Head-Mounted Displays', *Presence* **11**(6), 610–625.

Rosenberg, L. (1993), The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks, *in* 'Proceedings of IEEE Virtual Reality Annual International Symposium', Ieee, pp. 27–32.

Sahabi, H. and Basu, A. (1996), 'Analysis of Error in Depth Perception with Vergence and Spatially Varying Sensing', *Computer vision and image understanding* **63**(3), 447–461.

Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B. and Willemsen, P. (2005), 'Throwing versus walking as indicators of distance perception in similar real and virtual environments', *ACM Transactions on Applied Perception* **2**(1), 35–45.

Sakano, Y. and Ando, H. (2010), 'Effects of head motion and stereo viewing on perceived glossiness', *Journal of vision* **10**(9), 15:1–14.

Sandor, C., Cunningham, A., Dey, A. and Mattila, V. V. (2010), An Augmented Reality X-Ray system based on visual saliency, *in* 'IEEE International Symposium on Mixed and Augmented Reality', IEEE, Seoul, pp. 27–36.

Sands, J., Lawson, S. and Benyon, D. (2004), Do we need stereoscopic displays for 3D augmented reality target selection tasks?, *in* 'International Conference on Information Visualisation', Ieee, pp. 633–638.

Schor, C. M. (1993), Spatial constraints of stereopsis in video displays, *in* S. R. Ellis, ed., 'Pictorial communication in virtual and real environments', pp. 546–557.

Schor, C. and Wood, I. (1983), 'Disparity range for local stereopsis as a function of luminance spatial frequency', *Vision Research* **23**(12), 1649–1654.

Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M. and Banks, M. S. (2008), 'The surface of the empirical horopter', *Journal of Vision* **8**(3), 1–20.

Servos, P., Goodale, M. A. and Jakobson, L. S. (1992), 'The role of binocular vision in prehension: a kinematic analysis', *Vision Research* **32**(8), 1513–1521.

Seuntiens, P. J. (2006), Visual experience of 3D TV, Doctoral dissertation, Technische Universiteit Eindhoven.

Seuntiens, P. J., Heynderickx, I. E. J., IJsselsteijn, W. A., van Den Avoort, P. M., Berentsen, J., Dalm, I., Lambooij, Marc T. M and Oosting, W. (2005), Viewing experience and naturalness of 3D images, *in* 'Proceedings of SPIE', Vol. 6016, SPIE, pp. 601605–601605–7.

Sherstyuk, A., Dey, A., Sandor, C. and State, A. (2012), Dynamic Eye Convergence for Head-mounted Displays Improves User Performance in Virtual Environments, *in* 'ACM SIGGRAPH Symposium on interactive 3D Graphics and Games', Vol. 1, pp. 23–30.

Shibata, T., Kim, J., Hoffman, D. M. and Banks, M. S. (2011), 'The zone of comfort : Predicting visual discomfort with stereo displays', *Journal of Vision* **11**(8), 11:1–29.

Shipley, T., Ogle, K. N. and Rawlings, S. (1970), 'The nonius horopter - I. History and theory', *Vision Research* **10**(11), 1225–1262.

Sielhorst, T., Bichlmeier, C., Heining, S. and Navab, N. (2006), Depth Perception - A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons, *in* 'Int. Conf. on Medical Image Computing and Computer Assisted Intervention,', pp. 364–372.

Sielhorst, T., Feuerstein, M. and Navab, N. (2008), 'Advanced Medical Displays: A Literature Review of Augmented Reality', *Journal of Display Technology* **4**(4), 451–467.

Singh, G., Swan II, J. E. and Jones, J. A. (2010), Depth Judgment Measures and Occluding Surfaces in Near-Field Augmented Reality, *in* 'Symposium on Applied Perception in Graphics and Visualization', Vol. 1, pp. 149–156.

Sousa, R., Smeets, J. B. J. and Brenner, E. (2012), 'The effect of variability in other objects' sizes on the extent to which people rely on retinal image size as a cue for judging distance', *Journal of Vision* **12**(10), 6:1–8.

Stevenson, S. B. and Schor, C. M. (1997), 'Human stereo matching is not restricted to epipolar lines', *Vision Research* **37**(19), 2717–2723.

Stöckel, K. (2009), *Berlin im olympischen Rausch: die Organisation der Olympischen Spiele 1936*, Diplomica Verlag.

Stringer, N. S. (2003), The Use of Enhanced Depth Information in Telepresence, Doctoral dissertation, University of Surrey.

Sugano, N., Kato, H. and Tachibana, K. (2003), The effects of shadow representation of virtual objects in augmented reality, *in* 'IEEE and ACM International Symposium on Mixed and Augmented Reality', IEEE, Tokyo, pp. 76– 83.

Sutherland, I. E. (1968), A head-mounted three dimensional display, *in* 'Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)', ACM Press, New York, New York, USA, pp. 757–764.

Swan II, J. E., Livingston, M. A., Smallman, H. S., Brown, D., Baillot, Y., Gabbard, J. L. and Hix, D. (2006), A perceptual matching technique for depth judgments in optical, see-through augmented reality, *in* 'Virtual Reality Conference', pp. 19–26.

Swan II, J., Jones, A. and Kolstad, E. (2007), 'Egocentric depth judgments in optical, see-through augmented reality', *IEEE Transactions on Visualization and Computer Graphics* **13**(3), 429–442.

Takagi, A., Yamazaki, S., Saito, Y. and Taniguchi, N. (2000), Development of a stereo video see-through HMD for AR systems, *in* 'IEEE and ACM International Symposium on Augmented Reality', pp. 68–77.

Takeichi, H., Watanabe, T. and Shimojo, S. (1992), 'Illusory occluding contours and surface formation by depth propagation', *Perception* **21**(2), 177–184.

Teather, R. J. and Stuerzlinger, W. (2011), Pointing at 3D targets in a stereo head-tracked virtual environment, *in* 'IEEE Symposium on 3D User Interfaces (3DUI)', IEEE, pp. 87–94.

Thompson, W. B., Dilda, V. and Creem-Regehr, S. H. (2007), 'Absolute distance perception to locations off the ground plane', *Perception* **36**(11), 1559–1571.

Thompson, W. B., Gooch, A. A., Creem-regehr, S. H., Loomis, J. M. and Beall, A. C. (2004), 'Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments ?', *Presence: Teleoperators & Virtual Environments* **13**(5), 560–571.

Thomson, J. a. (1983), 'Is continuous visual monitoring necessary in visually guided locomotion?', *Journal of experimental psychology. Human perception and performance* **9**(3), 427–43.

Thurstone, L. L. (1931), 'Rank Order as a Psychophysical Method', *Journal of Experimental Psychology* **14**, 187 – 201.

Tsirlin, I., Allison, R. S. and Wilcox, L. M. (2008), 'Stereoscopic transparency: Constraints on the perception of multiple surfaces', *Journal of Vision* **8**(5), 1–10.

Tsirlin, I., Wilcox, L. M. and Allison, R. S. (2011), 'The Effect of Crosstalk on the Perceived Depth From Disparity and Monocular Occlusions', *IEEE Transactions on Broadcasting* **57**(2), 445–453.

Tsuda, T., Yamamoto, H., Kameda, Y. and Ohta, Y. (2005), Visualization methods for outdoor see-through vision, *in* 'International conference on Augmented telexistence', ACM, Christchurch, New Zealand, pp. 62–69.

Tufte, E. R. (1983), *The visual display of quantitative information*, Graphics press.

Uratani, K., Machida, T., Kiyokawa, K. and Takemura, H. (2005), A study of depth visualization techniques for virtual annotations in augmented reality, *in* 'IEEE Proceedings Virtual Reality', Ieee, pp. 295–296.

Urey, H., Chellappan, K. V., Erden, E. and Surman, P. (2011), 'State of the Art in Stereoscopic and Autostereoscopic Displays', *Proceedings of the IEEE* **99**(4), 540–555.

Utsumi, A., Milgram, P., Takemura, H. and Kishino, F. (1994), 'Investigation of Errors in Perception of Stereoscopically Presented Virtual Object Locations in Real Display Space', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **38**(4), 250–254.

Van Beurden, M. H. P. H., Van Hoey, G., Hatzakis, H. and Ijsselsteijn, W. a. (2009), Stereoscopic displays in medical domains: a review of perception and performance effects, *in* 'Proceedings of SPIE vol. 7240', Spie, pp. 72400A–72400A–15.

Van den Berg, A. V. and Brenner, E. (1994), 'Why two eyes are better than one for judgements of heading', *Nature* **371**, 700 – 702.

Van der Willigen, R. F. (2011), 'Owls see in stereo much like humans do.', *Journal of vision* **11**(7), 10:1–27.

Wallach, H. and Zuckerman, C. (1963), 'Stereoscopic depth constancy', *The American Journal of Psychology* **76**(3), 404–412.

Wanger, L., Ferwerda, J. and Greenberg, D. (1992), 'Perceiving spatial relationships in computer-generated images', *IEEE Computer Graphics and Applications* **12**(3), 44–58.

Ware, C. (2004), *Information visualization: perception for design*, Morgan Kaufmann, San Francisco.

Ware, C. and Mitchell, P. (2005), Reevaluating stereo and motion cues for visualizing graphs in three dimensions, *in* 'Proceedings of the 2nd symposium on Appied perception in graphics and visualization - APGV '05', ACM Press, New York, New York, USA, pp. 51–58.

Watt, S. J., Akeley, K., Ernst, M. O. and Banks, M. S. (2005), 'Focus cues affect perceived depth', *Journal of Vision* **5**(10), 834–862.

Wertheimer, M. (1923), 'Untersuchungen zur Lehre von der Gestalt II', *Psycologische Forschung* **4**, 301–350.

Westheimer, G. and McKee, S. (1980), 'Stereoscopic acuity with defocused and spatially filtered retinal images', *Journal of the Optical Society of America* **70**(7), 772–778.

Westheimer, G. and Tanzman, I. J. (1956), 'Qualitative depth localization with diplopic images.', *Journal of the Optical Society of America* **46**(2), 116–7.

Wheatstone, C. (1838), 'Contributions to the Physiology of Vision.-Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision', *Philosophical Transactions of the Royal Society of London* **128**, 371–394.

Willemsen, P., Colton, M. B., Creem-regehr, S. H. and Thompson, W. B. (2004), The Effects of Head-Mounted Display Mechanics on Distance Judgments in Virtual Environments, *in* 'Symposium on Applied perception in graphics and visualization', pp. 35–38.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H. and Thompson, W. B. (2009), 'The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments', *ACM Transactions on Applied Perception* **6**(2), 8:1–14.

Willemsen, P., Gooch, A. A., Thompson, W. B. and Creem-Regehr, S. H. (2008), 'Effects of stereo viewing conditions on distance perception in virtual environments', *Presence: Teleoperators and Virtual Environments* **17**(1), 91–101.

Wither, J. and Hollerer, T. (2005), Pictorial depth cues for outdoor augmented reality, *in* 'IEEE International Symposium on Wearable Computers', Osaka, pp. 92–99.

Witmer, B. G. and Kline, P. B. (1998), 'Judging Perceived and Traversed Distance in Virtual Environments', *Presence* **7**(2), 144–167.

Wloka, M. M. and Anderson, B. G. (1995), Resolving occlusion in augmented reality, *in* 'Proceedings of the 1995 symposium on Interactive 3D graphics', ACM Press, New York, New York, USA, pp. 5–12.

Woods, A., Docherty, T. and Koch, R. (1993), Image Distortions in Stereoscopic Video Systems, *in* 'Stereoscopic Displays and Applications IV,', Vol. 1915, pp. 36–48.

Woods, A. J. (2011), How are Crosstalk and Ghosting defined in the Stereoscopic Literature?, *in* A. J. Woods, N. S. Holliman and N. A. Dodgson, eds, 'SPIE Stereoscopic Displays and Applications XXII, Vol. 7863', Vol. 7863, pp. 78630Z–78630Z–12.

Wu, B., Ooi, T. L. and Zijiang, J. H. (2004), 'Perceiving distance accurately by a directional process of integrating ground information', *Nature* **428**, 73–77.

Xueting, L. and Ogawa, T. (2013), A depth cue method based on blurring effect in augmented reality, *in* 'Augmented Human International Conference', ACM Press, New York, New York, USA, pp. 81–88.

Yahiaoui, T., Khoudour, L. and Meurie, C. (2010), 'Real-time passenger counting in buses using dense stereovision', *Journal of Electronic Imaging* **19**(3), 031202–1–031202–11.

Yamanoue, H., Okui, M. and Okano, F. (2006), ‘Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images’, *Circuits and Systems for Video Technology, IEEE Transactions on* **16**(6), 744–752.

Yang, Z. and Purves, D. (2003), ‘A statistical explanation of visual space’, *Nature Neuroscience* **6**(6), 632–640.

Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A. and Xu, L.-Q. (2008), ‘Crowd analysis: a survey’, *Machine Vision and Applications* **19**(5-6), 345–357.

Zhang, L., Member, S., Vázquez, C. and Knorr, S. (2011), ‘3D-TV Content Creation : Automatic 2D-to-3D Video Conversion’, *IEEE Transactions on Broadcasting* **57**(2), 372–383.

Zhao, W. and Nandhakumar, N. (1996), ‘Effects of camera alignment errors on stereoscopic depth estimates’, *Pattern Recognition* **29**(12), 2115–2126.

Zhu, J., Pan, Z., Sun, C. and Chen, W. (2010), ‘Handling occlusions in video-based augmented reality using depth’, *Computer Animation and Virtual Worlds* **21**(5), 509–521.

Errata

Publication I

Equation 4 and the line above: The sign $\leq$ should be $\geq$. The sign is correct after that in the rest of the publication.

Publication III

Statistical analysis should have been conducted using repeated-measures ANOVA design.

Publication V

Statistical analysis should have been conducted using repeated-measures ANOVA design.

Aalto-DD 25/2014

ISBN 978-952-60-5584-8
ISBN 978-952-60-5585-5 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Media Technology
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS