

Institut für Visualisierung und Interaktive Systeme

Universität Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

Bachelorarbeit

Evaluation der Blicktiefenschätzung mittels Eye-Tracking in Augmented Reality

Seyda Zarife Öney

Studiengang:	Medieninformatik
Prüfer/in:	Jung. Prof. Michael Sedlmair
Betreuer/in:	Dipl.-Inf. Nils Rodrigues, Michael Becher, M.Sc.
Beginn am:	3. März 2019
Beendet am:	17. September 2019

Kurzfassung

Die Einblendung von virtuellen Objekten in die reale Umgebung war damals nur aus Filmen bekannt. Mithilfe der heutzutage verfügbaren Augmented Reality (AR) Technologien kann dies nun auch in der Realität umgesetzt werden. Deshalb wird der Begriff AR in der heutigen Zeit öfter erwähnt. Mit zunehmendem Interesse steigt auch der Fokus auf die Weiterentwicklung von AR Technologien, insbesondere für eine bessere und realistischere Darstellung von virtuellen Objekten. Viele aktuelle AR Technologien basieren auf Head Mounted Displays (HMD). Darunter gehört auch die HoloLens, die in dieser Arbeit verwendet wird. Mit Hilfe von HMD Systemen lassen sich Objekte, in einer beliebigen Entfernung zum Betrachter, in der realen Welt platzieren. Jedoch stellt sich die Frage, ob unsere Augen die Entfernung zu virtuellen Objekten gleichermaßen wahrnehmen, wie die Entfernung zu realen Objekten. Daher soll in dieser Arbeit die Frage untersucht werden, ob AR Einfluss auf die Wahrnehmung der Blicktiefe hat, indem ein Eye-Tracker, der in die HoloLens eingebaut wurde, verwendet wird. In einer Studie sollen die Blicktiefen-Daten für reale und virtuelle Objekte erfasst, analysiert und anschließend ausgewertet werden.

Abstract

In the past, the displaying of virtual objects into the real environment was only seen in movies. Now this can be made possible in reality with the current available Augmented Reality (AR) technologies. Therefore, the term AR is mentioned more often today. With stronger interest the focus on the further development of AR technologies increases, especially for a better and more realistic representation of virtual objects. Many current AR technologies are based on Head Mounted Displays (HMD). This also includes the HoloLens, which is used in this work. With the help of HMD systems, objects can be placed at any distance from the viewer in the real world. However, the question arises if our eyes perceive the distance of virtual objects in the same way as the real ones. Therefore, this thesis will point out if AR has an influence on the perception of gaze depth by using an eye tracker built into the HoloLens. In a study, the depth of view data for real and virtual objects will be collected, analysed and finally evaluated.

Inhaltsverzeichnis

1. Einleitung	11
1.1. Gliederung	12
2. Grundlagen	13
2.1. Augmented/Mixed Reality	13
2.2. Das menschliche Auge	14
2.3. Tiefenwahrnehmung	15
2.4. Tiefenwahrnehmung in AR	19
2.5. Eye-Tracking	22
3. Verwandte Arbeiten	25
3.1. Subjektive Tiefeneinschätzung	25
3.2. Tiefeneinschätzung mittels Eye-Tracking	27
4. Aufgabenstellung	29
4.1. Aufgabenstellung	29
5. Konzept	31
5.1. Anforderungen an die Testszene	31
5.2. Gestaltung der Testszene	31
6. Verwendete Hard- und Software	35
6.1. Microsoft Hololens	35
6.2. Pupil Labs Binocular Add-On	36
6.3. Unity	36
6.4. Pupil Software	37
7. Realisierung der Testszene	41
7.1. Erstellung der Aufgaben	41
7.2. Implementierung - Vorgehen	41
7.3. Hololens – Eye-Tracker Kommunikation	42
7.4. Kalibrierung	42
7.5. Büchertransformation	42
7.6. Datenerfassung	44
8. Studie	45
8.1. Studiendesign	45
8.2. Studienablauf	46
8.3. Probanden	48
8.4. Auswertung	49

9. Zusammenfassung und Ausblick	61
9.1. Zusammenfassung	61
9.2. Ausblick	61
Literaturverzeichnis	63
A. Anhang	67
A.1. Testszene	68
A.2. Einverständniserklärung	70
A.3. Fragebogen	73
A.4. Sehtest	76
A.5. Histogramme der Daten	78

Abbildungsverzeichnis

2.1.	Das Reality-Virtuality Kontinuum. [MTK95]	14
2.2.	Die Anatomie des menschlichen Auges [SSTT12].	15
2.3.	[Dah06].	16
2.4.	Zu sehen ist der Horopter in horizontale Ebene[WJX+16].	19
2.5.	Vergenz-Akkommodations-Konflikt [Kra16]	20
2.6.	Hornhautreflexion und helle Pupille im Bild der Infrarotkamera [JK03b]	23
3.1.	Versuchsaufbau [SSE15]	26
3.2.	Darstellung der Optik eines Wheatstone Stereoskops [DPHW11]	27
5.1.	Zahlen- und Wörrätsel, die für die Studie verwendet wurden.	32
5.2.	Schematische Darstellung über die genaue Position der Bücher.	34
6.1.	Pupil Capture Haupt-Fenster	38
6.2.	Die Pupil Capture Fenster der einzelnen Eye-Tracker.	39
7.1.	Kalibrierungsszene der Eye-Trackers.	43
7.2.	AR Marker, das durch die HOloLens erkannt wurde.	44
8.1.	virtuelle Testszene	46
8.2.	Kalibrierungsszene der HoloLens	47
8.3.	Durchschnittlicher Erfolg beim Lösen der Aufgaben auf den Büchern	50
8.4.	Der Boxplot zeigt die Bewertung der Ähnlichkeit zwischen der virtuellen und realen Bücher. Dabei bedeutet 1, dass es überhaupt nicht zutrifft, 5 bedeutet, dass es vollkommen zutrifft.	51
8.5.	Die Daten der realen und virtuellen Szene, die als Boxplots dargestellt werden.	52
8.6.	Boxplot der einzelnen Bücher jeweils in den realen und virtuellen Szenen	53
8.7.	Balkendiagramm von den Durchschnittswerten der einzelnen Bücher in der realen und virtuellen Szene	53
8.8.	Histogramm von den gesamten Daten	54
8.9.	Histogramm der realen (8.9a) und virtuellen Szene (8.9b)	55
8.10.	Histogramme von den einzelnen Büchern	59
A.1.	Skizze der Testszene in Hochformat	68
A.2.	Erster Teil der Einverständniserklärung	70
A.3.	Zweiter Teil der Einverständniserklärung	71
A.4.	Erster Teil der Einverständniserklärung	73
A.5.	Zweiter Teil der Einverständniserklärung	74
A.6.	Zweiter Teil der Einverständniserklärung	75
A.7.	Snellentest [Wik19]	76

A.8. Histogramm von Buch1 der realen Szene.	78
A.9. Histogramm von Buch1 der virtuellen Szene.	78
A.10. Histogramm von Buch2 der realen Szene.	78
A.11. Histogramm von Buch2 der virtuellen Szene.	78
A.12. Histogramm von Buch3 der realen Szene.	78
A.13. Histogramm von Buch3 der virtuellen Szene.	78
A.14. Histogramm von Buch4 der realen Szene.	78
A.15. Histogramm von Buch4 der virtuellen Szene.	78
A.16. Histogramm von Buch5 der realen Szene.	78
A.17. Histogramm von Buch5 der virtuellen Szene.	78

Tabellenverzeichnis

5.1. Abstand der einzelnen Bücher zum Betrachter	33
8.1. Ergebnis des zweiseitigen Welch t-Tests von den einzelnen Büchern	56
8.2. Ergebnis des zweiseitigen Welch t-Tests von den einzelnen Studienteilnehmern . .	56
8.3. Ergebnis des Kruskal-Wallis Test von den einzelnen Büchern	57
8.4. Ergebnis des Kruskal-Wallis Test von den einzelnen Studienteilnehmern	57

1. Einleitung

Was würde passieren, wenn sich die Realität und Virtualität miteinander vermischen würden? AR ist der Begriff, der sich hinter diesem Prinzip versteckt. In den letzten Jahren hat sie an Bedeutung gewonnen und wird in vielen Bereichen wie beispielsweise in der Industrie und Spielebranche eingesetzt. Durch zunehmendes Interesse in diesem Gebiet werden AR-Technologien weiterentwickelt mit der Intention, möglichst realitätsnahe Darstellungen zu erhalten. So haben aktuelle AR-Technologien, wie die HoloLens, hochauflösende Displays und sind in der Lage, virtuelle Objekte beliebig zu platzieren. Durch Weiterentwicklung dieser und anderer Eigenschaften wird die Unterscheidung von realen und virtuellen Elementen in der Zukunft schwieriger werden.

Aktuell sind Displays fähig, virtuelle Objekte sowohl weit entfernt als auch direkt vor dem Betrachter darzustellen. Aber was nehmen unsere Augen wahr? Sind die Augen in der Lage zu unterscheiden, dass es sich um ein Objekt handelt, das eigentlich auf einem Bildschirm vor unserer Nase liegt und nicht im physischen Raum? Augen haben die Fähigkeit zur Wahrnehmung räumlicher Tiefe. In einer realen Umgebung wird die Entfernung zu einem Objekt durch verschiedene Tiefenhinweise geschätzt, unter anderem durch die Akkommodation und Konvergenz der Augen. Die Blicktiefenschätzung in aktuellen AR HMD Technologien kann nicht auf die gleiche Art und Weise erfolgen, da hier die fokale Distanz konstant bleibt und deshalb die Augen immer auf eine bestimmte Distanz akkommodieren. Um dieses Problem zu beheben müssten Lichtfeld-Anzeigen genutzt werden. Somit stellt sich die Frage, ob sich die Blicktiefenschätzung in AR von der echten unterscheidet. Im Rahmen der Bachelorarbeit soll diese Frage behandelt werden. Zur Beantwortung dieser Frage ist eine Technologie notwendig, um zu messen, wie weit bzw. in welche Entfernung man schaut. Hier kommt ein Eye-Tracker zum Einsatz, der dreidimensionale Blickdaten erfassen kann. Dabei kann er in gewissen AR-Headsets integriert werden, um einerseits die Augen als Interaktionsinstrument zu verwenden, und um andererseits Analysen mit den Blickpositionen durchzuführen. Für diese Bachelorarbeit wurde eine HoloLens mit einem eingebauten Eye-Tracker verwendet, welcher speziell für die HoloLens entwickelt wurde.

Eye-Tracker werden in unterschiedlichsten Bereichen genutzt, wie beispielsweise in der Psychologie, Neurowissenschaften und im Bereich User Experience, wobei in diesen Anwendungsfeldern die 2D-Blickdaten ausreichend sind. Für unseren Anwendungsfall ist die Extrahierung der 3D Blickdaten notwendig, da die Blicktiefe gemessen werden muss. Dafür werden in einer Studie zwei Arten von Szenen realisiert: Es existiert eine reale Szene, bei der Objekte in verschiedenen Entfernungen platziert werden. Diese Szene wird ebenso in AR realisiert, indem die realen Objekte durch virtuelle ersetzt werden. In beiden Szenen werden die Blickdaten extrahiert und die erfassten Blicktiefen-Daten miteinander verglichen. In der Studie wird folgende Haupthypothese untersucht: **H1:** „Die Wahrnehmung der Blicktiefe unterscheidet sich in AR im Vergleich zur Realität.“

1.1. Gliederung

Die Arbeit ist in folgender Weise gegliedert:

Kapitel 2 - Grundlagen. Hier werden die Grundlagen, die zum Verständnis der Aufgabenstellung benötigt werden, vorgestellt. Hierzu wird Augmented Reality definiert und erläutert. Anschließend wird die Funktion des Auges beschrieben und auf die visuelle Wahrnehmung eingegangen. Zum Schluss dieses Kapitels wird das Eye-Tracking vorgestellt.

Kapitel 3 - Verwandte Arbeiten. Kapitel 3 stellt wissenschaftliche Arbeiten vor, die einen ähnlichen thematischen Aspekt wie diese Arbeit aufweisen.

Kapitel 4 - Aufgabenstellung. Hier wird die Aufgabenstellung näher erläutert.

Kapitel 5 - Konzept In diesem Kapitel wird das Konzept der Testszene beschrieben.

Kapitel 6 - Verwendete Hard- und Software. Hier werden die verwendeten Hard- und Software zur Realisierung der Testszene genannt und genauer erläutert.

Kapitel 7 - Realisierung der Testszene. Die Realisierung des Konzepts wird in diesem Kapitel erläutert. Es wird insbesondere auf die Implementierung eingegangen.

Kapitel 8 - Studie. In diesem Kapitel wird das Design und der Ablauf der Studie erläutert und anschließend die Auswertung und Ergebnisse der Studie vorgestellt.

Kapitel 9 - Zusammenfassung und Ausblick fasst die Ergebnisse der Arbeit zusammen und stellt Anknüpfungspunkte vor.

2. Grundlagen

2.1. Augmented/Mixed Reality

Da die Arbeit eine Untersuchung in AR anstellt, werden im Folgenden die Begriffe AR, Virtual Reality (VR) und Mixed Reality (MR) erläutert und differenziert.

Augmented Reality

Unter Augmented Reality versteht man die Erweiterung der realen Welt um computergenerierte Zusatzobjekte, die auch als virtuelle Objekte bezeichnet werden [MTK95]. Dieser Begriff wird oft im Zusammenhang mit HMD-Geräten erwähnt. So wurde bereits in den 1960er Jahren die erste AR Technologie entwickelt, als Sutherland die Darstellung von 3D-Grafiken mit Hilfe eines durchsichtigen HMDs ermöglichte [Sut68]. Augmented Reality hat jedoch ab 1997 zunehmend an Bedeutung gewonnen, nachdem Azuma diesen Bereich genauer definiert sowie vorhandene Anforderungen und bisherige Forschungen in diesem Bereich aufgezeigt hat [Azu97].

Nach der Definition von Azuma beschränkt sich AR weder auf bestimmte Display-Technologien, noch auf den Sehsinn und kann somit auf alle Sinne übertragen werden [ABB+01]. Dabei definiert Azuma die folgenden drei Kriterien, die ein AR-System besitzen muss:

1. Kombination von realen und virtuellen Objekten in einer realen Umgebung
2. Genaue Registrierung von realen und virtuellen Objekten
3. Interaktion in 3D und in Echtzeit.

Reality-Virtuality-Kontinuum

Augmented Reality wird in der Literatur häufig unterschiedlich definiert und deshalb besteht kein Konsens über die Definition [MTK95]. Oft wird sie mit Virtual Reality in Relation gebracht. Unter Virtual Reality versteht man die vollständige Ersetzung der realen Umgebung durch eine virtuelle [ABB+01]. Die Beziehung der beiden Konzepte ist auf das von Milgram und Takemura definierte Reality-Virtuality-Kontinuum zurückzuführen (siehe Abbildung 2.1) [MTK95]. Das Kontinuum kann vereinfacht als eine Verlaufslinie zwischen der realen Umgebung und virtuellen Umgebung betrachtet werden. Dabei beschreibt die linke Seite des Kontinuums die Umgebung, die nur aus realen Objekten besteht. Hierbei kann die Wahrnehmung der realen Umgebung sowohl durch unmittelbare Betrachtung als auch über eine Anzeige erfolgen. Auf der rechten Seite des Kontinuums werden alle Umgebungen zugeordnet, die nur aus virtuellen Objekten bestehen. Dazu gehören beispielsweise Computergrafik-Simulationen, die entweder monitorbasiert oder immersiv

2. Grundlagen

sein können. Der Bereich zwischen diesen beiden Umgebungen wird als Mixed Reality definiert, sodass in dieser Umgebung eine Kombination aus realen und virtuellen Elementen dargestellt wird. Deshalb ist Augmented Reality nach Milgram und Kishino Teil von Mixed Reality. Augmented Virtuality ist eine weitere Umgebung, die Teil von Mixed Reality ist, in der die virtuelle Umgebung durch reale Objekte erweitert wird. Während bei Augmented Reality der reale Anteil überwiegt, ist dies bei Augmented Virtuality der virtuelle Anteil.

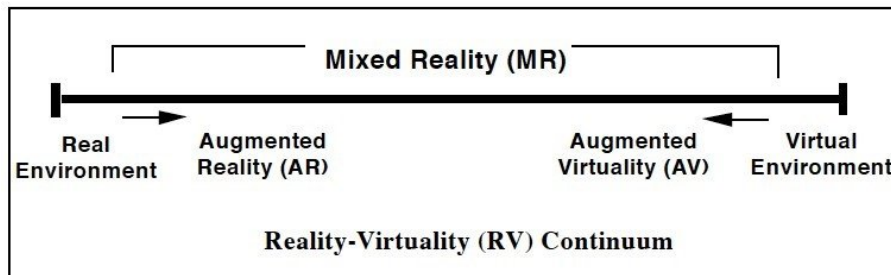


Abbildung 2.1.: Das Reality-Virtuality Kontinuum. [MTK95]

2.2. Das menschliche Auge

Der Primärsinn eines Menschen ist der Sehsinn, da 80 % der Informationen über das Auge aufgenommen und an das Gehirn weitergeleitet werden [Dah06]. Die Augen besitzen unterschiedliche Funktionen, darunter das Farbsehen, dynamische Sehen - die Wahrnehmung von Bewegungen - und insbesondere die Tiefeneinschätzung, die einen wichtigen Teil dieser Arbeit darstellt. Im Folgenden wird zuerst der Aufbau und die Funktion des Auges erläutert und dabei wird der Fokus auf die Blicktiefenschätzung gelegt.

In Abbildung 2.2 wird der Aufbau des Auges dargestellt. Dabei findet die visuelle Wahrnehmung von Informationen durch Lichteinfall von einem Objekt in das Auge statt, indem die Lichtstrahlen die einzelnen Bestandteile des Auges passieren um diese in der Netzhaut in Signale umzuwandeln und an das Gehirn weiterzuleiten [SSTT12]. Zunächst gelangt das Licht durch die Hornhaut zur Pupille. Die Hornhaut ist die äußerste Schicht des Auges und hat aufgrund ihrer gekrümmten Form die Funktion, das Licht zu brechen (Lichtfokussierung) [SSTT12]. Das gebrochene Licht geht als nächstes durch eine Öffnung, die als Pupille bezeichnet wird. Die Lichtmenge, die später auf die Netzhaut trifft, wird durch die Größe der Pupille reguliert. Dabei wird ihre Größe und Form durch die Iris definiert. Je größer die Pupille und Iris werden, desto mehr Licht gelangt in das Auge. Bei schlechten Lichtverhältnissen wird sie größer, um mehr Licht durchzulassen, bei hoher Lichtintensität kleiner um die Lichtmenge einzuschränken.

Als nächstes durchdringen Lichtstrahlen die Linse [SSTT12]. Da die Linse sich verformen lässt, kann sie Licht aus verschiedenen Tiefen, d.h. aus der Nähe und Weite, konzentrieren. Die Fokussierung auf Objekte in unterschiedlicher Distanz wird durch die Linse und den Ziliarmuskeln, von denen die Linse umgeben ist, erreicht. Beim Fokussieren eines nahen Objektes ziehen sich die Ziliarmuskeln zusammen und die Linse wird dicker, sodass sie mehr Brechkraft bekommt. Bei weit entfernten Objekten entspannen sich die Ziliarmuskeln und die Linse bekommt eine flache Form. Abschnitt 2.3 enthält nähere Informationen zur Tiefenwahrnehmung.

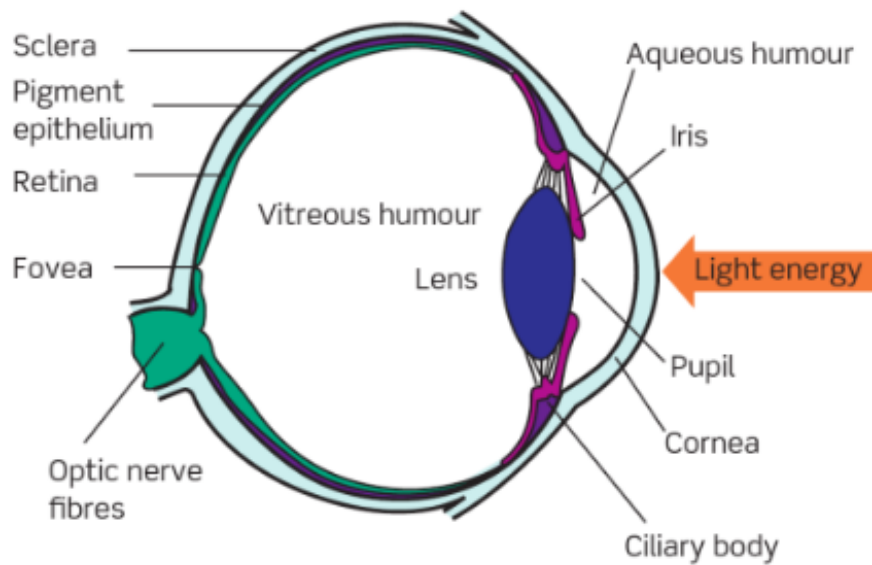


Abbildung 2.2.: Die Anatomie des menschlichen Auges [SSTT12].

Nachdem die Linse die Lichtstrahlen gebündelt hat, schickt sie sie auf die Innenseite des Auges auf die Netzhaut, damit die Lichtstrahlen dort in Nervenimpulse umgewandelt und über den Sehnerv ins Gehirn weitergeleitet werden [SSTT12]. Die Lichtstrahlen durchdringen die fünf Schichten der Netzhaut um schließlich die Photorezeptoren zu erreichen [GB09]. Es gibt zwei Arten von Photorezeptoren, Stäbchen und Zapfen. Es gibt 120 Millionen Stäbchen im Auge, die empfindlicher gegenüber Licht sind, wobei sie nicht farbempfindlich sind. Sie werden für die Wahrnehmung von Helligkeitsunterschieden (Schwarz-Weiß-Sehen) gebraucht. Zapfen sind weniger lichtempfindlich aber sie sind farbempfindlich, sodass sie für das Farbsehen des Auges bei (Tages)licht zuständig sind. Es gibt 6 Millionen Zapfen im Auge und drei Zapfentypen, die auf Licht in unterschiedlichen Wellenlängen empfindlich reagieren: S-Zapfen sind auf kurze Wellenlängen (blau) empfindlich, M-Zapfen auf mittlere Wellenlängen (grün) und L-Zapfen auf lange Wellenlängen (rot) [GB09]. Sie decken die jeweiligen Farbspektren ab. Die Photorezeptoren sind nicht überall gleich verteilt. In der Fovea, der Bereich mit bestem Sehvermögen, befinden sich nur Zapfen, während um die Fovea die Stäbchen dicht angesiedelt sind.

Die Stäbchen und Zapfen haben die Funktion, die Lichtimpulse in elektrische Impulse umzuwandeln. In den anderen Schichten der Netzhaut werden die Impulse weiterverarbeitet und anschließend über die Ganglienzellen (über die Axone) in den Sehnerv und somit ins Gehirn weitergeleitet. Die visuelle Aufnahme von Informationen wird schließlich nach Ausführung der einzelnen Schritte im Auge erreicht. [SSTT12]

2.3. Tiefenwahrnehmung

Die Fähigkeit der Tiefenwahrnehmung ist ein wichtiges Merkmal um im Alltag zurechtzukommen. Da unsere Umgebung dreidimensional ist und wir mit ihr interagieren, müssen unsere Augen den Abstand eines Objektes bestimmen können. Obwohl die Umgebung dreidimensional wahrgenommen wird, wird auf die Netzhaut im Auge ein zweidimensionales Bild abgebildet. Der Tiefeneindruck im Auge entsteht durch zwei Arten von Tiefenhinweisreizen: den okulomotorischen und visuellen

Hinweisreizen [TJ96]. Während okulomotorische Hinweise sich auf die Ausrichtung und Anpassung der Augen auf unterschiedliche Distanzen beziehen, werden visuelle Hinweise in monokulare und binokulare unterteilt. Durch Kombination dieser Hinweise nehmen wir die Tiefe wahr. Dabei lernen wir durch Erfahrungen mit unserer Umgebung, den Zusammenhang zwischen den Hinweisreizen und der Tiefe zu verstehen [GB09]. Im Folgenden werden beide Hinweisreize verdeutlicht.

2.3.1. Okulomotorische Tiefenhinweisreize

Okulomotorisch setzt sich aus „Okulus“, welches übersetzt aus dem Lateinischen Auge bedeutet, und „motorisch“, das die Bewegung verkörpern soll, zusammen [Dah06]. Somit sind die Bewegungen der Augen gemeint, da sie sich an die Entfernung des fokussierten Objektes anpassen. Um den Blick auf etwas zu fokussieren sind drei Arten von okulomotorischen Kriterien notwendig [RHFL10]. Diese sind die Akkommodation, Konvergenz und Pupillendynamik, wobei deren Kombination als okuläre Near-Triade bezeichnet wird. Dazu werden beim Betrachten eines Objektes für eine klare, binokulare Sicht die Augen um einen bestimmten Grad akkommodiert und konvergiert. Der Winkel der Augen zueinander ist dabei von der Entfernung eines Objektes abhängig.

Als Akkommodation wird der Prozess bezeichnet, bei dem die Augen ihre optische Brechkraft an die Distanz des fokussierten Objektes anpassen, sodass ein scharfes Bild auf der Netzhaut projiziert wird [RHFL10]. Dabei verändern die Ziliarmuskeln ihren Zustand abhängig von der Distanz des fokussierten Objektes, um die Linse entsprechend zu formen (siehe Abschnitt 2.2) [TJ96].

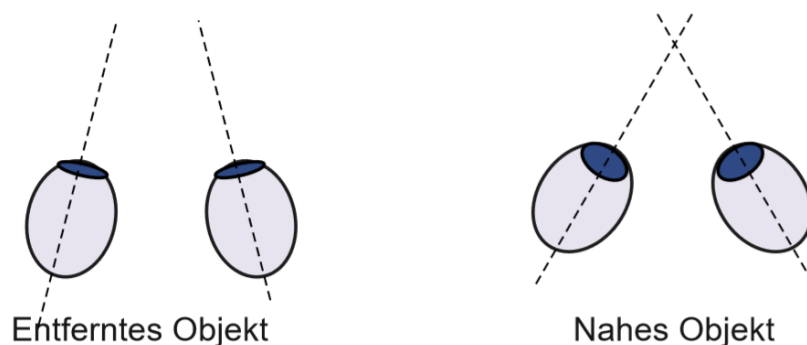


Abbildung 2.3.: [Dah06].

Die Akkommodation findet hauptsächlich gemeinsam mit der Konvergenz statt. Als Konvergenz wird die Bewegung der Augen in entgegengesetzte Richtungen bezeichnet. Bei Fokussierung eines nahen Objektes bewegen sich die Augen nach innen, um das Objekt jeweils im fovealen Bereich der Netzhäute abzubilden.

So hängen die beiden Prozesse eng miteinander zusammen, da beispielsweise beim Fokussieren eines nahen Objektes die Augen sich nach innen bewegen (konvergieren) und die Ziliarmuskeln sich zusammenziehen (akkommodieren), wodurch ein möglichst scharfes Bild des Objektes auf der Netzhaut abgebildet wird [TJ96]. Hinzu kommt die Anpassung der Pupillengröße, wobei eine nahe Konvergenz/Akkommodation zur Verkleinerung der Pupille führt während die weite Konvergenz/Akkommodation zur Vergrößerung der Pupille beiträgt [LMF+09]. Demzufolge spüren wir die Bewegungen der Augen und die verschiedenen Zustände in den Augen um die Tiefe wahrnehmen und einschätzen zu können.

Monokulare Tiefenhinweisreize

Bei monokularen Hinweisen handelt es sich um Hinweise für die Tiefenwahrnehmung, die nur mit einem Auge erkannt werden können [Dah06]. Monokulare Tiefenhinweise können in Bilder eingesetzt werden, um sie drei-dimensional wirken zu lassen. Die verschiedenen Arten werden nun im Einzelnen betrachtet.

Okklusion Die Okklusion beschreibt den Zustand, bei der ein Objekt ein anderes Objekt vollständig oder Teile davon überdeckt, damit das überdeckte Objekt weiter entfernt erscheint [GB09]. Die Okklusion stellt ordinale Informationen bereit, da sie ausschließlich Informationen über die Tiefenordnung gibt. Dabei können keine Rückschlüsse über den Abstand von zwei Objekten gezogen werden. Vorteil der Okklusion ist, dass die Wirksamkeit bezüglich der Tiefenwahrnehmung nicht mit der Entfernung abnimmt.

Relative Größe Dieser Hinweis beschreibt die Wahrnehmung der Entfernung von gleich großen Objekten, indem die Objekte, die weiter entfernt sind, im Blickfeld kleiner dargestellt werden, als diejenigen, die in der Nähe sind [Dah06; GB09]. Schließlich kann der Benutzer anhand der Größe des Bildes auf der Netzhaut beurteilen, wie weit das Objekt entfernt ist. Dazu ist für den Betrachter das nötige Wissen über die Größe der Objekte notwendig [GB09].

Relative Höhe Anhand dieses Hinweises werden Objekte, die sich unterhalb vom Horizont befinden und im Blickfeld eine höhere Basis haben, als weiter entfernt wahrgenommen. Das gleich gilt für Objekte, die sich über dem Horizont befinden, beispielsweise Wolken, aber im Blickfeld niedriger sind.

Perspektive Darunter wird die Veränderung des Erscheinungsbildes von Objekten oder Oberflächen mit zunehmender Entfernung verstanden. Dabei gibt es vier Formen [SSTT12], die Objekte annehmen können:

1. Lineare Perspektive beschreibt, wie parallele Linien mit zunehmender Entfernung konvergieren bzw. sich annähern. Dieser Effekt kann in Eisenbahnschienen beobachtet werden.
2. Texturgradienten lassen die Textur von Elementen, die in einer Szene gleichmäßig verteilt sind, mit zunehmender Entfernung dichter verpackt erscheinen. Dadurch können Texturgradienten Informationen über den Abstand und die Neigung von Oberflächen geben. Außerdem bewirken schnelle Veränderungen von Texturgradienten, dass Kanten und Ecken erkannt werden. [SSTT12]
3. Bei der Luftperspektive erscheinen Objekte, die sich weiter weg vom Betrachter befinden, weniger scharf als nähere Objekte [Dah06]. Dies wird durch die Streuung des Umgebungslichts aufgrund staubiger Luft oder Nebel verursacht [GB09].
4. Schatten sind eine wichtige Eigenschaft der Tiefenwahrnehmung. Objekte werfen abhängig von der Richtung der Lichtstrahlen Schatten und können dadurch Informationen über ihre Position geben. Außerdem wird ein räumlicher Eindruck vermittelt.

Bewegungsparallaxe Bei der Bewegungsparallaxe bekommt der Betrachter, der sich in eine bestimmte Richtung bewegt, den Eindruck, dass die Objekte sich in die entgegengesetzte Richtung bewegen [GB09]. Nahe Objekte scheinen sich schneller zu bewegen als diejenigen, die weiter entfernt sind. Dieser Effekt wird verursacht, indem das Bild eines nahen Objektes

eine größere Positionsänderung auf der Netzhaut hervorruft, während das Bild eines Objektes, das weiter entfernt vom Betrachter liegt, sich um einen kleineren Abstand bewegt, sodass der Eindruck der langsamen Fortbewegung des Objektes entsteht. Dieser Effekt kann beobachtet werden, wenn man während einer Zugfahrt aus dem Fenster schaut.

Binokulare Tiefenhinweisreize

Da wir zwei Augen besitzen, die ca. 6cm horizontal voneinander entfernt sind, nehmen wir unsere Umgebung aus verschiedenen Blickwinkeln wahr [GB09]. Durch den Unterschied der Blickwinkel der beiden Augen werden unterschiedliche Bilder erzeugt, die im Gehirn zu einem Bild verschmelzen. Dies wird auch als **binokulare Disparität** bezeichnet und trägt eine entscheidende Rolle bei der Tiefenwahrnehmung, welches durch Wheatstone, dem Erfinder des Stereoskops im 19. Jahrhundert), bewiesen wurde [Qia97].

Die binokulare Disparität wird erzeugt, indem sich die Augen auf ein bestimmtes Objekt fokussieren, sodass der Fixierpunkt des Objektes in den beiden Augen jeweils auf der Fovea (Mitte der Netzhaut) abgebildet wird [Dah06; SSTT12]. Der restliche Teil des Objektes und die Umgebung werden auf der Netzhaut der beiden Augen auf verschiedenen Stellen projiziert und werden deshalb als Disparität der Netzhautpunkte bezeichnet [Dah06]. Es gibt einen Bereich, in dem keine Netzhautdisparität vorliegt. Dieser Bereich wird Horopter genannt [LMF+09]. Der Horopter stellt die Gesamtheit der Punkte dar, die auf korrespondierende Punkte der Netzhaut fallen und verläuft durch den Fixierpunkt (siehe Abbildung 2.4). Korrespondierende Punkte sind Stellen auf der Netzhaut, die an beiden Augen in gleicher Entfernung und Richtung zur Fovea liegen. An diesen Stellen entsteht keine Netzhautdisparität. So fällt das fixierte Objekt auf die Foveas, während Teile der Umgebung, die sich auf dem Horopter befinden, auf die korrespondierenden Punkte der Netzhaut fallen. Punkte außerhalb des Horopters weisen Netzhautdisparitäten auf. Während Punkte, die sich vor dem Horopter befinden, eine gekreuzte Netzhautdisparität haben, haben Punkte hinter dem Horopter eine nichtgekreuzte Disparität [LMF+09]. Die Disparitäten werden dadurch erkannt, indem die Bilder, welche auf die Netzhaut projiziert werden, übereinander geschoben werden. Wenn sich die abgebildeten Punkte nicht überlappen, liegt eine Disparität vor. Der Abstand der beiden Punkte wird dabei als absoluter Disparitätswinkel bezeichnet [GB09]. Durch sie kann die Entfernung eines Objektes wahrgenommen werden. Der Betrag des Disparitätswinkels wird mit zunehmendem Abstand des Objektes zum Horopter größer. Außerdem verändert sich die absolute Disparität jedes Mal bei Blickwechsel. Wenn ein Mensch sich in einer Szene pro Sekunde drei mal auf etwas fokussiert, müssen jedes Mal ein neuer Horopter und die absoluten Disparitäten berechnet werden. Eine effizientere Möglichkeit Disparitätsinformationen zu erhalten, ist die relative Disparität, welches die Differenz der absoluten Disparitäten von zwei Objekten beschreibt. Weitere Informationen dazu können aus der Arbeit von Goldstein [GB09] entnommen werden.

Die Disparität liefert nur bis zu einem bestimmten Grad Tiefeninformationen. Es können nur Punkte, die innerhalb eines kleinen Bereichs – dem Panum – um den Horopter liegen, im Gehirn so verarbeitet werden, dass die beiden Bilder auf der Netzhaut

miteinander fusionieren bzw. verschmelzen und als ein dreidimensionales Bild wahrgenommen wird [LMF+09]. Punkte außerhalb des Panums verursachen Doppelbilder.

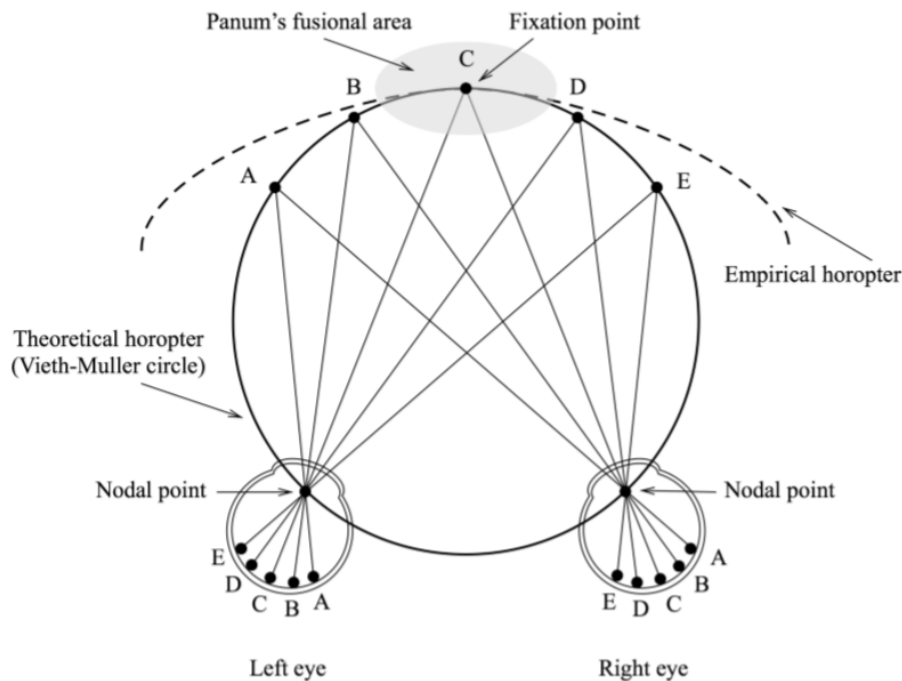


Abbildung 2.4.: Zu sehen ist der Horopter in horizontale Ebene[WJX+16].

2.4. Tiefenwahrnehmung in AR

Die Tiefenwahrnehmung spielt auch in AR eine wichtige Rolle, um AR-Anwendungen optimal nutzen zu können, sodass Benutzer die räumliche Beziehung zwischen realen und virtuellen Objekten möglichst genau verstehen. Deshalb muss sichergestellt werden, dass virtuelle Objekte an den richtigen Stellen in der realen Welt abgebildet und genau wahrgenommen werden.

Während in der realen Umgebung die Wahrnehmung der Tiefe durch natürlich vorkommende Tiefenhinweise, die fast immer in Übereinstimmung sind, gewährleistet wird, müssen in AR virtuelle Objekte auf einem zweidimensionalen Display abgebildet werden [DM96; DWSS17]. Dabei wird mit Hilfe von Grafiktechniken zum Rendern von virtuellen Objekten die Illusion erzeugt, dass Objekte in ihrer beabsichtigten räumlichen Position in der realen Umgebung erscheinen. So werden Stereo-Rendering, perspektivische Projektion und Beleuchtungsmodelle eingesetzt [DWSS17]. Damit können die in Abschnitt 2.3 genannten Tiefenhinweise umgesetzt werden.

Stereoskopische Visualisierung

Die Tiefenhinweise, die in AR-Geräten eingesetzt werden, sind zum Großteil monokular und können somit in zweidimensionalen Abbildungen erkannt werden [DWSS17]. In modernen AR-Technologien werden zusätzlich binokulare Tiefenhinweise bereitgestellt, um eine bessere dreidimensionale Wahrnehmung zu gewährleisten. Realisiert wird dies durch eine stereoskopische Darstellungsvariante und wird in binokularen HMDs eingesetzt.

2. Grundlagen

Die stereoskopische Visualisierung basiert auf dem Prinzip der binokularen Parallaxe [MPWL13]. Dazu werden zwei Bilder mit Parallaxeninformationen generiert, indem die Szene aus unterschiedlichen Perspektiven dargestellt wird. Anschließend erscheinen die beiden Bilder auf den Displays vor den Augen und werden ans Gehirn weitergeleitet um eine stereoskopische bzw. räumliche Wahrnehmung zu bewirken.

Probleme Die Nutzung von Stereoskopischen Displays ermöglicht eine bessere Tiefenwahrnehmung, kann aber Konflikte im Auge hervorrufen. Ein Problem entsteht bei zusätzlicher Verwendung von monokularen Hinweisen, wenn eine Inkonsistenz zwischen ihnen herrscht [DM96]. Denn bei widersprüchlichen Hinweisen oder bei Fehlen von Tiefeninformationen erscheint die Tiefe ungenau. Deshalb ist die Konsistenz zwischen den Tiefenhinweisen wichtig, um keine falschen Tiefeninformationen zu erhalten.

Es existieren weitere Probleme, die von Benutzer und von den Einschränkungen der aktuellen Technologie abhängig sind. Benutzerspezifische Probleme entstehen beispielsweise durch Unterschiede zwischen den realen und angenommenen Augenabständen, der Sehschärfe und Kalibrierungsprobleme [KSF10]. Der Nachteil bedingt durch die technologischen Einschränkungen ist das inkonsistente Verhalten zwischen Vergenz und Akkommodation: Beim Fokussieren auf ein virtuelles Objekt konvergieren die Augen auf den fokussierten Punkt, während die Linsen der Augen auf die Bildebene bzw. Fokalebene scharf stellen (akkomodieren). Der Abstand der Bildebene vom Betrachter in HMDs wird von der Optik gebildet und ist bei aktueller Technik konstant. Dabei findet die Konvergenz in unterschiedlichen Tiefen, außerhalb der Bildebene, statt, wodurch ein Konflikt entsteht. Dieses Problem wird als **Vergenz-Akkommodations-Konflikt** bezeichnet und kann Einfluss auf die Wahrnehmung der Tiefe haben (siehe Abbildung 2.5) [DM96]. Zudem kann dies zu Augenbelastungen führen.

Für eine optimale Tiefenwahrnehmung müssen sich Akkommodation und Konvergenz gegenseitig anpassen. Dieses Prinzip wurde bereits in Abschnitt 2.3.1 erläutert.

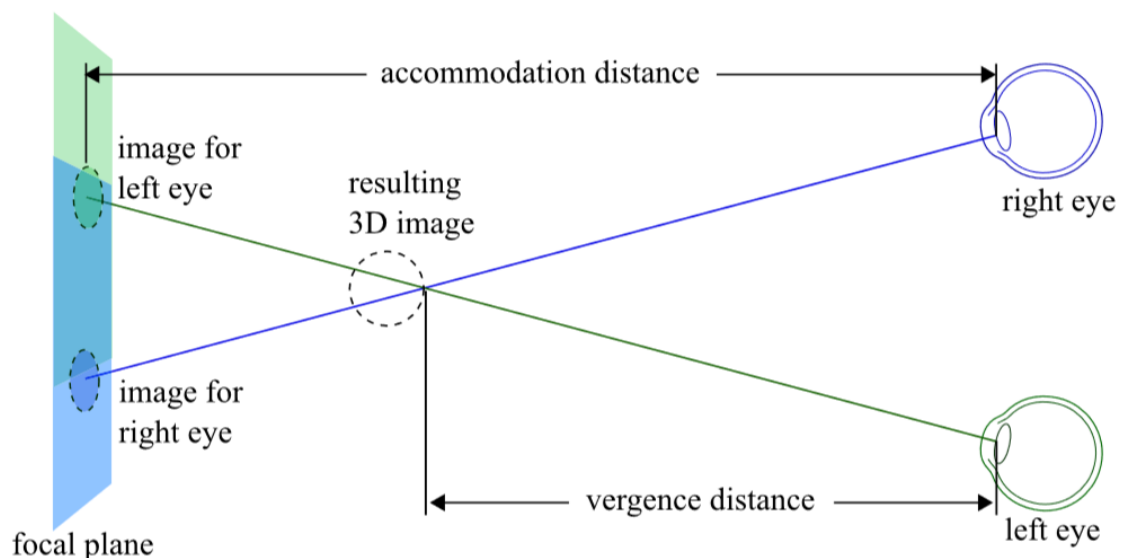


Abbildung 2.5.: Vergenz-Akkommodations-Konflikt [Kra16]

Einflüsse in AR

Die Erkennung der Tiefe in AR hängt von verschiedenen Faktoren ab. Kruijff et al. [KSF10] unterteilen die Faktoren in Umgebung, Capturing/Erfassung, Erweiterung, Display-Gerät und Benutzer.

Umgebung Die Wahrnehmung von virtuellen Objekten kann durch die Umgebung beeinflusst werden, die zu Problemen in der Tiefenwahrnehmung führen können [KSF10]. Die Struktur der Umgebung – die Anordnung der Objekte – ist relevant, da sie bei Verwendung von Tiefenhinweisen Unterstützung leistet. Ansonsten kann beispielsweise Unordnung in der Umgebung zu Okklusionsproblemen führen, das die deutliche Erkennung der Szene erschwert. Ein weiterer Einflussfaktor ist die Beleuchtung. Die Lichtverhältnisse in der Umgebung beeinflussen die Darstellung der virtuellen Objekte durch Veränderung deren Farben und Kontraste. Zudem verursachen sie Reflexionen auf den Displays (und Linsen), die die Wahrnehmung von virtuellen Objekten stören.

Capturing/Erfassung In Capturing werden alle Bedingungen betrachtet, die während der Umwandlung des optischen Bildes in ein digitales Signal durch eine Kamera eine Rolle spielen [KSF10]. Dazu gehört auch die Bildauflösung. Beispielsweise kann eine geringe Bildauflösung die Tiefenwahrnehmung beeinträchtigen, da die Möglichkeit besteht, dass die Objekte eine flachere Form annehmen und nicht voneinander unterschieden werden können. Da Kameras nur in bestimmten Beleuchtungsbedingungen qualitative Aufnahmen machen können, wird deren Funktionalität bei über- oder unterbelichteten Umgebungen gestört. Daraus folgt, dass Kontraste und Objektdetails schlechter dargestellt werden, was wiederum zu Verlust der Tiefeninformationen führt. Der Einfluss des Kontrastes liegt darin, dass bei niedrigem Kontrast Objekte, die in unterschiedlichen Tiefen liegen, den Eindruck hinterlassen, als würden sie in der Tiefe miteinander verschmelzen. Dadurch erscheinen sie unschärfer und werden deshalb als weiter entfernt wahrgenommen.

Erweiterung Unter diesem Begriff werden alle Einflüsse zusammengefasst, die im Zusammenhang mit der Registrierung von digitalen Inhalten auf Oberflächen entstehen [KSF10]. Der Fokus liegt hier auf dem Technologie-Aspekt. Eine wichtige Eigenschaft von Technologien ist die Unterstützung von Okklusion. In der realen Welt ist die Überdeckung von entfernten Objekten durch nahe liegende selbstverständlich. Wenn diese Eigenschaft in AR falsch umgesetzt wird, indem die Überdeckung der Objekte nicht der Wirklichkeit entspricht, kann dies eine falsche Tiefenwahrnehmung hervorrufen und den Eindruck vermitteln, dass die Objekte nicht Teil der Szene sind. Die Rendering-Qualität ist eine spezifische Eigenschaft für die Genauigkeit der Darstellung von digitalen Objekten, die abhängig von der eingesetzten Technologie ist. Es besteht kein direkter Zusammenhang zwischen der Rendering-Qualität und der Tiefeneinschätzung. Problematisch kann es werden, wenn zur Verbesserung der Genauigkeit Antialiasing-Methoden verwendet werden. So können Unterschiede in der Rendering-Qualität (Auflösung) und in Aliasing Einfluss auf die Wahrnehmung haben, indem die Akkommodation unterschiedlich erfolgt. Daraus folgt eine falsche stereoskopische Disparität.

Display-Gerät Da AR in verschiedenen Arten von Geräten genutzt wird, haben solche Geräte unterschiedliche Displays mit speziellen Darstellungsarten [KSF10]. Ein sehr bekanntes Beispiel ist für das stereoskopische Display ist die HoloLens, die ein wichtiger Bestandteil dieser Arbeit ist und deshalb in Tiefenwahrnehmung in AR/Stereoskopische Visualisierung vorgestellt wird.

Displays haben unterschiedliche Sichtfelder (FOV). Während Menschen ein horizontales FOV von über 180° haben, ist dies in Display-Geräten sehr eingeschränkt. Beispielsweise haben Video see-through HMDs ein horizontales FOV von 30-100°, während es in optischen see-through HMDs noch kleiner ist. Obwohl der Sichtbereich für die virtuellen Objekte eingeschränkt ist, kann die reale Umgebung bei manchen Display-Arten außerhalb des FOV-Bereichs gesehen werden. Bei einem kleinen FOV ist es möglich, dass die Anzeige von virtuellen Elementen nicht vollständig ist. Deshalb gilt: Je größer der Sichtbereich ist, desto genauer ist die Tiefenwahrnehmung [DM96].

Die Auflösung und Pixeldichte eines Displays bestimmen die scharfe Erkennung von Objekten. Die Tiefenwahrnehmung kann dadurch beeinflusst werden, da Objekte, die scharf erkennbar sind, näher erscheinen als sie es in Wirklichkeit sind. Dies wird durch die Perspektiven-Eigenschaft der monokularen Hinweise verursacht. (siehe Kapitel monokulare Hinweise/Perspektive).

Benutzer Die Fähigkeit des Benutzers, Entfernungen von verschiedenen Objekten zu erkennen, erfolgt mit Hilfe von Tiefenhinweisen, die in Kapitel Tiefenwahrnehmung beschrieben wurden. Manche dieser Hinweise werden gezielt eingesetzt. Dabei ist die Okklusion eine der wichtigsten Hinweise, die bei fehlerhaftem Einsatz zu falschen Tiefen-Informationen führt [DM96]. Die Auswirkungen werden bei unzureichender Verwendung von Tiefenhinweisen größer, da keine eindeutige Erkennung der Tiefen erfolgen kann. Ein weiteres Problem, das im Zusammenhang mit stereoskopischen Displays auftritt ist der Vergenz-Akkommodations-Konflikt, der vom Benutzer wahrgenommen werden kann. Dieser Konflikt wird in Abschnitt 2.4 erklärt.

2.5. Eye-Tracking

Eye-Tracking wird zur Erfassung von Augenbewegungen genutzt [Duc07]. Aus der Erfassung lassen sich mehrere Informationen herauslesen. Eine der wichtigsten Bewegungsarten des Auges, die bestimmt werden können, sind Fixationen, in denen die Augen auf ein Punkt fokussieren und somit relativ bewegungslos sind [JK03a]. Schnelle Augenbewegungen zwischen den Fixationspunkten werden als Sakkaden bezeichnet. Dabei wird als Gaze der Punkt bezeichnet, auf welches man schaut.

2.5.1. Eyetracker

Das zur Messung verwendete Gerät wird als Eye-Tracker bezeichnet. Für die Überwachung der Augenbewegungen gibt es zwei Arten von Techniken. Die erste Technik misst die Position der Augen in Bezug auf den Kopf [Duc07]. Bei der zweiten Möglichkeit wird die Ausrichtung der Augen im Raum gemessen, die auch als *point of regard* (POR) bezeichnet werden. Hierfür

gibt es mehrere Möglichkeiten, die Augenbewegungen zu messen: Die Elektrookulografische Technik, die in den 70er Jahren am weitesten verbreitet war [Duc07], basiert auf Elektroden, die auf die Haut um die Augenhöhle angebracht werden [JK03a]. Dabei werden die elektrischen Potentialunterschiede gemessen. Eine weitere veraltete Technik sind die scleralen Kontaktlinsen, die direkten Kontakt mit dem Augapfel benötigen. Dazu trägt man Kontaktlinsen, die die Hornhaut und Sklera (siehe Abbildung 2.2) verdecken. Die Messung findet mit Hilfe der Metallspulen statt, die an den Kontaktlinsen um den Rand der Linse eingebettet sind, welche Schwankungen im elektromagnetischen Feld verursachen [JK03a]. Da beide Möglichkeiten aufwändig sind und oft nicht zur Messung der POR geeignet sind, wurde eine neue Technik entwickelt, die videobasierte Hornhautreflexions-Eye-Tracker [Duc07]. In den meisten aktuellen Eye-Tracker-Systemen ist diese Technik enthalten.

Videobasierte kombinierte Pupillen/Hornhaut Reflektion

Bei dieser Eye-Tracker-Technik werden Video-Aufnahmen von den Augen gemacht, um daraus die Merkmale des Auges zur Ermittlung der POR zu extrahieren [Duc07; JK03a]. Die beobachteten Merkmale sind dabei die Hornhautreflexion und das Pupillenzentrum. Dafür werden eine Infrarotkamera und eine Bildverarbeitungssoftware zur Lokalisierung und Identifizierung der Merkmale in Echtzeit verwendet. Zunächst wird ein Infrarotlichtstrahl auf das Auge gerichtet, um starke Reflexionen in den Merkmalen des Auges zu erzeugen. Dabei gelangt der Lichtstrahl in die Netzhaut und wird zurück reflektiert, sodass die Pupille als eine helle Fläche in der Kamera erscheint und der Punkt der Hornhautreflexion identifiziert werden kann (siehe Abbildung 2.6). Anschließend wird mit Hilfe von trigonometrischen Rechenoperationen der betrachtete Punkt (POR) ermittelt. Bei dieser Technik muss zu Beginn eine Kalibrierung durchgeführt werden, um die Besonderheiten der Augenbewegungen von jeder individuellen Person anzupassen.

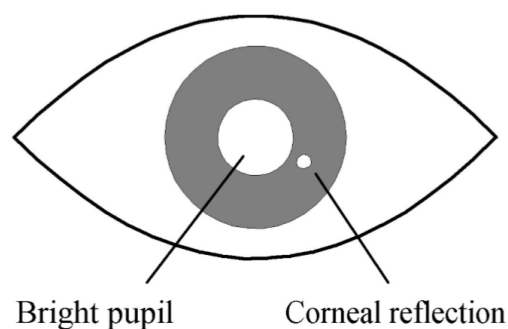


Abbildung 2.6.: Hornhautreflexion und helle Pupille im Bild der Infrarotkamera [JK03b]

Diese Technik kann auf zwei Arten verwendet werden. Zum einen besteht die Möglichkeit, das Gerät stationär zu verwenden. Für die Messung muss allerdings der Kopf fixiert werden, damit die Augenposition mit der POR übereinstimmt. Als zweite Möglichkeit steht der dynamische Einsatz zur Verfügung. Dazu wird der Eye-Tracker auf dem Kopf getragen (HMDs).

3. Verwandte Arbeiten

Die Tiefenwahrnehmung in AR ist ein bekanntes Problem und wurde in zahlreichen wissenschaftlichen Arbeiten untersucht. Für die Messung der wahrgenommenen Tiefe gibt es verschiedene Möglichkeiten. Während zahlreiche Studien auf subjektiver Tiefeneinschätzung basieren, gibt es auch Studien, die wie diese Arbeit die Tiefenwahrnehmung mittels Eye-Tracking messen.

3.1. Subjektive Tiefeneinschätzung

Der Einfluss von AR auf die Tiefenwahrnehmung wurde bereits Ende des 20. Jahrhunderts untersucht. So haben Rolland et al. [RGA95] 1995 eine der ersten Studien durchgeführt, die dieses Problem aufzeigt. Die Studie beinhaltet den Vergleich der Distanz von zwei Objekten. Für die Studie wird ein optical see-through HMD verwendet. Die Besonderheit der Objekte besteht einerseits darin, dass sie unterschiedliche Formen besitzen (Würfel und Zylinder), andererseits werden sie für den Vergleich in drei verschiedenen Darstellungsarten präsentiert. Bei den ersten beiden Arten sind beide Objekte real bzw. virtuell, während bei der dritten ein virtuelles Objekt mit einem realen Objekt verglichen wird. Bei dem Experiment vergleicht der Proband den Abstand der beiden Objekte zueinander und entscheidet, ob das Objekt auf der rechten Seite näher oder weiter entfernt von dem Objekt auf der linken Seite liegt. Der reale Abstand der Objekte zum Betrachter beträgt zwischen 0.8 und 1.2m.

Das Ergebnis dieser Studie hat gezeigt, dass die virtuellen Objekte im Vergleich zu den realen weiter entfernt wahrgenommen werden, wenn beide Arten in der gleichen Tiefe positioniert sind. Außerdem hat sich bei der Analyse die Interpretation ergeben, dass die Stärke der Tiefenwahrnehmung für virtuelle Objekte instabil ist. Als eine mögliche Ursache für dieses Ergebnis wird der **Vergenz-Akkommodations-Konflikt** aufgeführt. Auch wenn das Ergebnis den Unterschied der Wahrnehmung von virtuellen Objekten aufgezeigt hat, wurde der Einfluss von manchen Aspekten, wie der Versuchsaufbau und ihr Berechnungsmodell, auf die Ergebnisse nicht ausgeschlossen. Deshalb haben Rolland et al. in ihren späteren Experimenten mit einem besseren AR-Display gearbeitet und haben eine bessere Tiefengenauigkeit erreicht [RMAR02].

Eine aktuellere Arbeit stammt von Swan et al. [SSE15], die in ihren Experimenten eine andere Technik zur Messung der wahrgenommenen Tiefe angewendet haben. Sie haben Experimente mit zwei verschiedenen Methoden für den Vergleich der Tiefenbeurteilung von virtuellen (AR) und realen Zielen durchgeführt. Die erste Methode ist das Perceptual Matching, bei dem der Proband die Entfernung eines Ziels mit Hilfe eines Zeigeobjektes zeigt. Die Hand kann dabei als Zeigeobjekt verwendet werden. Bei der zweiten Methode wird Blind Reaching eingesetzt, die sich im Vergleich zur ersten Methode nur bezüglich der Sichtbarkeit des Zeigeobjektes unterscheidet. Hier kann der Proband das Zeigeobjekt nicht sehen (siehe Abbildung 3.1).

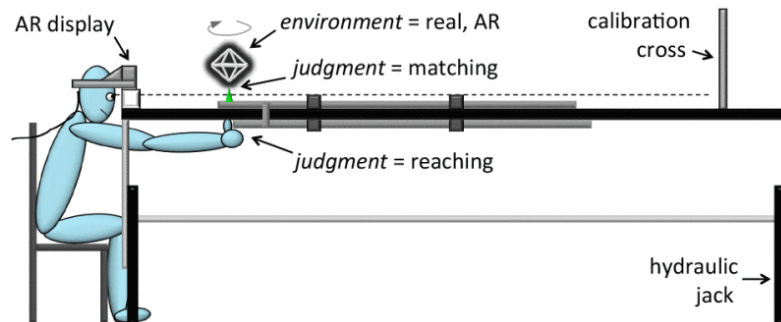


Abbildung 3.1.: Versuchsaufbau [SSE15]

In den Experimenten beurteilen die Probanden mit Hilfe dieser beiden Techniken die Tiefe von realen und virtuellen Objekten. Bei dem Objekt handelt es sich um ein langsam rotierendes Drahtmodell-Oktaeder und die Entfernung der Objekte vom Betrachter liegt zwischen 34 und 50cm. Außerdem ist sichergestellt, dass beide Szenen gleich aussehen. Für diese Experimente wird ein durchsichtiges optisches HMD verwendet, bei welchem die Fokusebenen in unendlicher Weite liegen.

Das Ergebnis der Experimente hat gezeigt, dass die Probanden mittels der Matching Methode die Distanz der realen Objekte sehr genau ermittelt haben. Bei den virtuellen Zielobjekten hingegen ist eine Überschätzung der Distanz beobachtbar. Als mögliche Ursache für solch ein Ergebnis wird der Vergenz—Akkommodations-Konflikt aufgeführt. Dieser Konflikt verursacht das Verschieben des Vergenz-Winkels nach außen, was zur Überschätzung der wahrgenommenen Distanz des AR-Ziels führt.

Die Arbeiten von Rolland et al. [RGA95] und Swan et al. [SSE15] wenden zwar unterschiedliche Techniken zur Messung der Tiefenwahrnehmung an, sie stellen aber einen realen und virtuellen Vergleich für die Untersuchung auf. In dieser Arbeit wird auch ein solcher Vergleich für die Analyse angewendet. Beide Arbeiten haben in ihren Studien nähere Distanzen untersucht. Während in [SSE15] die Abstände in Greifnähe sind, haben Rolland et al. [SSE15] bis zu 1.2m untersucht. Im Gegensatz zu [RGA95] und [SSE15], werden in dieser Arbeit Objekte, deren Entfernung zwischen 0.9m und 3.6m variiert, betrachtet. Es existieren auch Arbeiten, die die Tiefenwahrnehmung auf mittlere und weite Distanzen untersuchen. Swan et al. haben in ihrer Arbeit ein Experiment durchgeführt, bei dem der Abstand zwischen 5 und 45m beträgt [SJK+07]. Bei der Analyse hat sich folgendes Resultat ergeben: Während die Entfernung des nächsten Objektes (ca. 5m) relativ genau geschätzt wird, wird bei den Objekten, die weiter entfernt sind, ein Wechsel von Unterschätzung zu Überschätzung der Entfernung beobachtet. Der Wechsel erfolgt bei einem Abstand von 23m. In der Arbeit von Foley [Fol89] wird ein ähnliches Ergebnis für nähere Abstände von 1 bis 4m erreicht, wenn die binokulare Disparität isoliert von anderen Tiefenhinweisen betrachtet wird. Deshalb wird in der Arbeit von Swan et al. [SJK+07] vermutet, dass die Stereo-Disparität ein wichtiger Tiefenhinweis ist und möglicherweise der Grund für das Resultat ist. Verglichen mit dieser Arbeit, existiert in [SJK+07] kein expliziter Vergleich mit realen Objekten.

3.2. Tiefeneinschätzung mittels Eye-Tracking

In diesem Abschnitt werden Arbeiten vorgestellt, in denen Eye-Tracker verwendet werden um die Gaze zu messen. Für diesen Ansatz gibt es noch keine Studien mit AR. Stattdessen wurden stereoskopische Displays untersucht, bei denen auch der **Vergenz-Akkommodations-Konflikt** beobachtet wird. A. Duchowski et al. [DPHW11] haben in ihrer Arbeit ein Wheatstone Stereoskop (siehe Abbildung 3.2) und einen Eye-Tracker verwendet um den Zusammenhang zwischen der Blicktiefe, die durch die Reaktion der Augenvergrößerung bestimmt wurde, und der abgebildeten Tiefe von Objekten in einem stereoskopischen Display zu untersuchen. Die Untersuchung hat gezeigt, dass eine Tendenz zur Unterschätzung der wahrgenommenen Entfernung besteht, wenn die Objekte vor oder hinter dem Bildschirm positioniert sind.

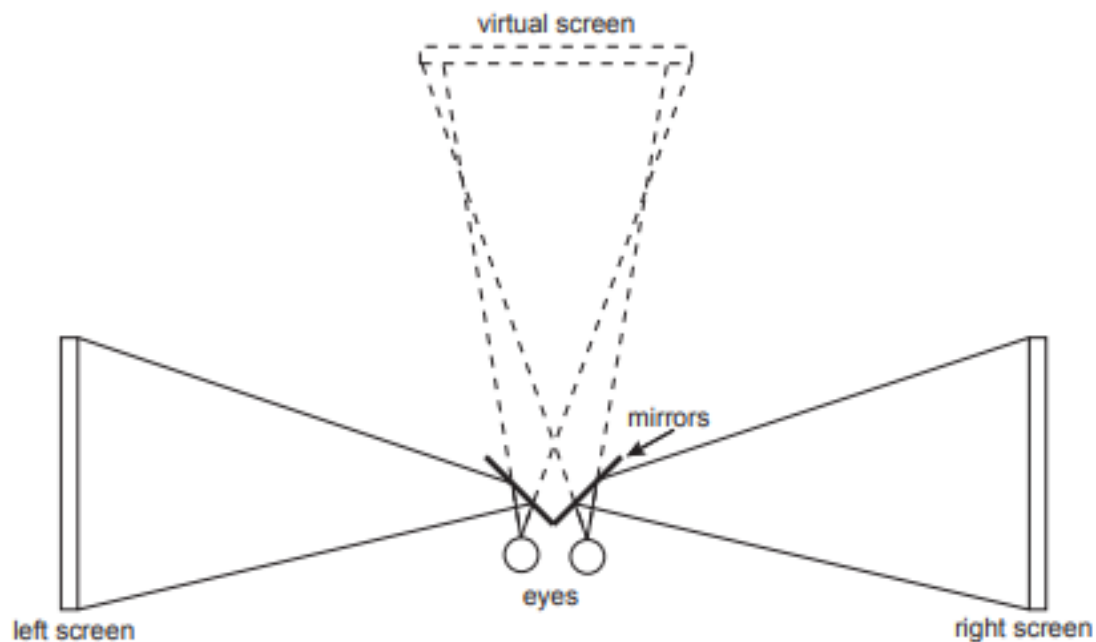


Abbildung 3.2.: Darstellung der Optik eines Wheatstone Stereoskops [DPHW11]

4. Aufgabenstellung

4.1. Aufgabenstellung

Im Alltag sind wir auf das dreidimensionale Sehen angewiesen. Um mit der Umgebung korrekt interagieren zu können, müssen wir in der Lage sein, zu erkennen, in welcher Distanz die Objekte liegen. Das gleiche gilt auch in AR. Wenn die reale Umgebung um virtuelle Objekte angereichert wird, müssen die Tiefen und Größen der virtuellen und realen Objekte übereinstimmen, um den beabsichtigten realen Standort von virtuellen Objekten zu erkennen. Nur so wird die ideale Interaktion in AR garantiert. Die Interaktion ist insbesondere für AR HMD-Technologien von Bedeutung. Ein grundlegendes Problem ist hier jedoch die Wahrnehmung von Tiefe. Aufgrund von Einflüssen, die in Kapitel 2 erläutert wurden, wird die Tiefe von AR-Objekten unterschätzt. Dazu gibt es viele Forschungen, in denen die wahrgenommene Tiefe gemessen wird. Dabei wird die subjektive Einschätzung der Tiefe durch die Probanden beurteilt, indem beispielsweise ein virtuelles Objekt zu einem realen Ziel in die Tiefe bewegt werden soll. Es wurde bis jetzt noch keine Studien durchgeführt, in denen die Tiefe aus den 3D Gaze-Positionen extrahiert und damit die Tiefenwahrnehmung untersucht wird.

Das Ziel dieser Arbeit ist zu untersuchen, ob AR Einfluss auf die Blicktiefe hat. Dafür soll unter der Verwendung einer HoloLens mit integriertem Pupil Labs Eye-Tracker die 3D-Gaze-Positionen gemessen werden, um daraus die wahrgenommene Tiefe zu ermitteln. Zur Untersuchung ist die Realisierung einer Studie notwendig, die einen Vergleich zwischen Real und Virtuell aufstellen soll. Zunächst soll eine Testszene entworfen werden, die aus anschaulichen Objekten besteht. Dabei müssen die Objekte in unterschiedlichen Tiefen positioniert werden. Anschließend soll eine echte Version der Testszene aufgebaut und eine virtuelle Version davon implementiert werden. Die beiden Szenarien müssen dabei möglichst identisch aussehen. In einer Studie sollen letztendlich in beiden Szenarien die Gaze-Daten erfasst werden, wenn die Personen auf die Objekte, die in unterschiedlichen Distanzen liegen, schauen. Bei der Analyse der Daten soll zunächst untersucht werden, wie genau die Tiefenschätzung in der realen Szene erfolgt ist. Anschließend soll die Haupthypothese, ob ein Unterschied der Blicktiefe zwischen AR und der Realität existiert, untersucht werden indem die beiden Szenen miteinander verglichen werden.

5. Konzept

In diesem Kapitel wird ein grundlegendes Konzept zur Realisierung der Testszene vorgestellt.

5.1. Anforderungen an die Testszene

Zur Realisierung der Testszene müssen mehrere Aspekte berücksichtigt werden. Zunächst muss die Szene so aufgebaut sein, dass die Realisierung in der Realität und in AR möglich ist. Deshalb muss das Objekt in der Szene eine realisierbare Form und Größe besitzen. Da das Ziel der Vergleich der realen und virtuellen Szene ist, müssen die Form, Größe und Farbe des Objektes so gewählt werden, dass sie keinen Einfluss auf die Tiefenwahrnehmung ausüben. Das bedeutet, dass die Objekte in unterschiedlichen Tiefen möglichst einheitlich aussehen sollten. Außerdem müssen die Objekte so gestaltet sein, dass der Blick der Personen auf dem Objekt fokussiert bleibt damit die Gaze-Daten richtig zugeordnet werden können. Deshalb müssen die Objekte passende Aufgaben beinhalten. Eine weitere Anforderung wird auf die Realisierung der virtuellen und realen Szene gestellt. Hier ist es wichtig, dass beide Szenen möglichst identisch aussehen. Dafür müssen neben den genannten Eigenschaften der Objekte (Form, Größe, Farbe) auch ihre Platzierung im Raum berücksichtigt werden.

5.2. Gestaltung der Testszene

Die Testszene ist die Basis der Studie da ihr Aufbau und ihre Gestaltung Einfluss auf das Ergebnis der Studie hat. Für die Gestaltung der Szene müssen verschiedene Komponenten definiert werden.

5.2.1. Auswahl des Objektes

Ein wichtiger Bestandteil der Testszene ist das Objekt. Bei der Auswahl des Objektes wurden zunächst die Anforderungen an die Testszene erfasst, daraus hat sich gefolgert, dass ein Buch eine sinnvolle Entscheidung ist. Gemäß der Aufgabenbeschreibung handelt es sich um ein unkompliziertes Objekt, das einfach zu realisieren ist weil es in der Realität existent ist und in AR einfach mit Hilfe von 3D-Grafiksoftware erstellt werden kann. Vorteilhaft für die reale Szene ist dabei, dass die Platzierung von Büchern sehr einfach ist, indem sie auf ein Tisch gestellt werden. Da die Bücher mit Hardcover ausgestattet sind, bleiben sie stabil und gerade auf dem Tisch. Die Größe der Bücher variiert weil sie abhängig von der Entfernung des Buches vom Betrachter ist. Bücher, die bis zu 2m entfernt sind, sind in Standardformat (17.4 x 1.9 x 24.1m). Bücher, die weiter entfernt liegen, sind (21.6 x 3 x 28.7 cm) groß.

5.2.3. Tiefe

In der Studie werden später die Gaze-Daten in unterschiedlichen Tiefen gemessen. Deshalb gibt es mehrere Bücher in der Szene, die in verschiedenen Abständen zum Betrachter stehen. Bei der Bestimmung der Tiefe muss wieder Rücksicht auf AR genommen werden, da die HoloLens nur in einem bestimmten Bereich eine optimale Darstellung der virtuellen Objekte ermöglicht. Aufgrund der konstanten fokalen Distanz müssen Objekte in der Nähe von 2m platziert werden, um den Vergenz-Akkommodations-Konflikt minimal zu halten, dieser Bereich wird als Komfortzone bezeichnet. In Kapitel 6 wird darauf eingegangen. In Tabelle 5.1 sind die Abstände der einzelnen Bücher aufgelistet. Um den Effekt des Konflikts zu beobachten, liegt der Abstand des ersten Buches bei 0.9m. Das zweite Buch ist genau in dem Bereich der fokalen Ebene. Die restlichen Bücher befinden sich ebenfalls noch in der Komfortzone. Die Positionen der Bücher sind so gewählt, dass beim Betrachten der Bücher keine Überdeckung erfolgen kann. Eine schematische Darstellung über die Position der Bücher ist in Abbildung 5.2 erkennbar.

Buch	Abstand
Buch 1	0.9m
Buch 2	2m
Buch 3	1.5m
Buch 4	3.5m
Buch 5	2.9m

Tabelle 5.1.: Abstand der einzelnen Bücher zum Betrachter

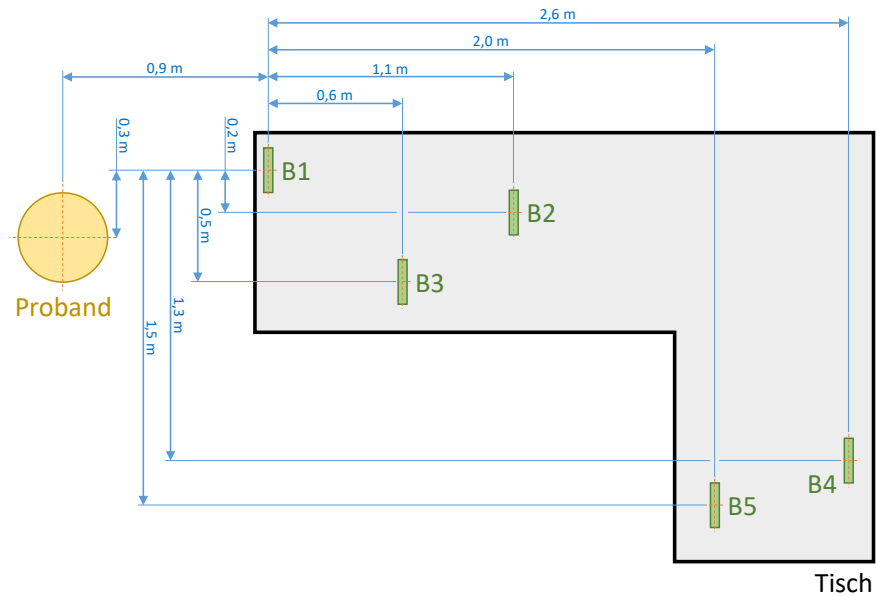


Abbildung 5.2.: Schematische Darstellung über die genaue Position der Bücher.

5.2.4. Umgebung

Neben den Objekten und ihren Abstand ist auch die Gestaltung der Umgebung wichtig. Dazu gehört unter anderem die Nutzung eines Tisches, damit die Bücher darauf gestellt werden können. Der Tisch sollte schwarz bedeckt sein, damit die Schatten, die von den Büchern geworfen werden, weniger sichtbar sind. Zudem der Hintergrund dann nicht durch die schwarzen Buchstaben.

6. Verwendete Hard- und Software

Für die Realisierung der Testszene musste eine AR-Technologie und ein Eye-Tracker verwendet werden. Dafür wurde ein optical see-through HMD System, die Microsoft HoloLens, eingesetzt, in welche der Pupil Labs Eye-Tracker integriert wurde. Für die Implementierung der Anwendung wurde Unity und die Pupil Capture Software von Pupil Labs verwendet. Im Folgenden werden die genannten Hardware und Software erläutert.

6.1. Microsoft HoloLens

Bei der HoloLens¹ handelt es sich um das erste ungebundene kommerziell erhältliche AR HMD System [Corb; EMP+17]. Sie ist das erste holographische Computer mit einem eigenen Betriebssystem. Sie ist vollständig unabhängig, da die HMD keine Anbindung zu einem Rechner erfordert. Für die Anzeige von binokularen Abbildungen besitzt die HoloLens ein durchsichtiges RGB-Display. Dabei kommen zwei durchsichtige Linsen zum Einsatz, die eine holographische Auflösung von 2.3 Millionen Lichtpunkten haben [Cora]. In den Linsen sind feine, unsichtbare Rillen enthalten, die den Eindruck hinterlassen, dass sich die virtuellen Objekte in den gewünschten Position der realen Umgebung befinden [Tay16].

Für die Abbildung von virtuellen Objekten an gewünschten Orten der realen Umgebung und für die Interaktion mit ihnen bietet die HoloLens verschiedene Sensoren an. Sie beinhaltet vier Kameras, jeweils zwei auf der linken und rechten Seite damit die HoloLens ihre eigene Position in der realen Welt feststellen kann [Tay16]. Zusätzlich verfügt die HoloLens über eine inertielle Messeinheit, die aus einem Beschleunigungssensor, Gyroskop und Magnetometer zusammengesetzt ist, um die Position und Bewegungen des Kopfes zu verfolgen.

In der Mitte der HoloLens befindet sich eine 2 MP Foto / HD Video Kamera, mit welche Bilder und Videoaufnahmen der realen Umgebung mit virtuellen Inhalten gemacht werden können [Tay16].

Um den Abstand zu realen Objekten zu ermitteln, nutzt die HoloLens eine Infrarotkamera mit einem Infrarot-Laserprojektor, die geradeaus gerichtet sind [Tay16]. Dabei sendet der Laser Infrarotlichtstrahlen zum Objekt aus. Wenn die Strahlen auf das Objekt treffen, werden sie wieder zur Infrarotkamera zurück reflektiert. Die dafür benötigte Zeit bestimmt die Entfernung zum realen Objekt.

Die HoloLens besitzt zusätzlich ein Umgebungslichtsensor, der die Beleuchtung der Umgebung erkennt. Für eine optimale Darstellung der Hologramme werden sie an die Umgebungsbeleuchtung angepasst.

¹<https://docs.microsoft.com/de-de/windows/mixed-reality/hololens-hardware-details> Zuletzt überprüft am 15.09.19

Die Abbildung von virtuellen Objekten, die sich bewegen oder an die Position des Betrachters in Echtzeit anpassen müssen, benötigen viel Rechenleistung [Tay16]. Deshalb besitzt die HoloLens drei verschiedene Prozessoren. Sie besteht aus einer Zentraleinheit (CPU) mit einem Intel Atom Prozessor (Atom x5-Z8100), einer Grafikverarbeitungseinheit (GPU) und einer kundenspezifischen holographischen Verarbeitungseinheit (HPU).

Die Steuerung von Anwendungen auf der HoloLens erfolgt über Gesten und Sprachbefehlen. Dazu braucht die HoloLens eine Tiefenkamera bzw. ein Mikrofon. Die HoloLens erkennt zwei verschiedene Handgesten - die Air-tap und die Bloom Geste [Cora]. Um zum Hauptmenü zu gehen muss die Bloom-Geste verwendet werden. Die Air-tap Geste ist ähnlich wie ein Maus Klick. Dabei muss mit der Headgaze das Zielobjekt anvisiert werden um anschließend mit der Air-tap Geste auszuwählen. Alternativ kann ein Clicker verwendet werden, der über Bluetooth mit der HoloLens kommuniziert [Tay16].

6.1.1. Optische Einschränkungen

In Kapitel Grundlagen wurden bereits die optischen Einschränkungen von HMD-Systemen erläutert. Für die HoloLens gelten die gleichen Einschränkungen. Der optische Abstand des HoloLens-Displays beträgt 2.0 m. Damit kein Vergenz-Akkommodations-Konflikt entsteht, wird von Microsoft empfohlen, Objekte möglichst in der Nähe von 2 m zu platzieren. Die optimale Zone zum platzieren von Objekten liegt zwischen 1.25m und 5m. [Cora]

6.2. Pupil Labs Binocular Add-On

Der Binocular Add-On Eye-Tracker von Pupil Labs² wurde speziell für die Microsoft HoloLens entwickelt. Der Eye-Tracker enthält zwei 200Hz Eye-Tracking Kameras und eine Weltkamera wobei er über ein USB-C Anschluss verfügt. In dieser Bachelorarbeit wurde der Eye-Tracker für das rechte Auge durch den Pupil Pro Eye-Tracker von Pupil Labs³ ausgetauscht. Im Vergleich zum Binocular Add-On besitzt dieser eine 120Hz Eye-Tracking Kamera. Die Auflösung der beiden Eye-Tracker sind auch unterschiedlich. Während der Pupil Pro Eye-Tracker eine maximale Auflösung von 1920x1080 Pixel hat, sind es beim Binocular Add-On 400x400 Pixel.

6.3. Unity

Unity⁴ ist eine Spiele-Engine, die für die Entwicklung von Spielen genutzt wird. In Unity wird zunächst ein Projekt angelegt und eine Szene erstellt. Die Szene wird mit *GameObjects* gefüllt, bei denen es sich beispielsweise um die Kamera oder den Cursor handelt [Tay16]. Auch virtuelle Objekte werden als *GameObjects* in die Umgebung platziert. In die *GameObjects* müssen Komponenten angehängt werden, um ihnen individuelle Eigenschaften zu verleihen. Eine mögliche Komponente,

²<https://pupil-labs.com/products/vr-ar/> Zuletzt überprüft am 15.09.19

³<https://pupil-labs.com/products/core/> Zuletzt überprüft am 15.09.19

⁴<https://unity.com/de> Zuletzt überprüft am 15.09.19

die für die Realisierung der Testszene benötigt wurde, ist der MeshRenderer, um ein bestimmtes Material für ein Object zu definieren. Ohne ihn ist das *GameObject* in der Szene nicht sichtbar. Weitere Komponenten sind beispielsweise Collider. Sie werden genutzt, wenn erkannt werden soll, dass zwei *GameObjects* aufeinander treffen. Collider werden auch für bestimmte Methoden, wie die Physics.Raycast-Methode, benötigt. Hier wird ein Strahl vom Ursprung in eine bestimmte Richtung geworfen, um gegen *GameObjects*, die eine Collider besitzen, aufeinanderzutreffen und beispielsweise Informationen über das *GameObject* zu erhalten. Unity besitzt zwar eine Bibliothek von Komponenten, es besteht aber auch die Möglichkeit, Skripte zur Erstellung von Komponenten in JavaScript oder C# - eine Programmiersprache von Microsoft - zu schreiben. Speziell für die HoloLens wird C# verwendet. Für die Entwicklung von HoloLens wird Visual Studio als Entwicklungsumgebung (IDE) verwendet, in der die Skripte erstellt werden, und anschließend als Komponenten zu den jeweiligen *GameObjects* hinzugefügt werden.

Alle Komponenten, *GameObjects* und Skripte werden in *Assets* gespeichert. Um Komponenten und Skripte von Frameworks/Toolkits nutzen zu können, müssen diese in *Assets* importiert werden. Für die Nutzung der HoloLens in Unity muss das *HoloToolkit* importiert werden. Darin sind Komponenten enthalten, die speziell für die Entwicklung der HoloLens benötigt werden. Die Basis-Komponenten, die für die Entwicklung der HoloLens benötigt werden, sind der InputManager, der DefaultCursor sowie die Mixed-RealityCamera. Durch die Integration von weiteren Komponenten und *GameObjects* lassen sich Anwendungen für die HoloLens realisieren.

Das *ARToolkit*⁵, welches für die Realisierung der Testszene benötigt wurde, kann ebenfalls in Unity importiert werden. Das *ARToolkit* ist eine Softwarebibliothek, die bei der Entwicklung von AR Anwendungen ermöglicht, dass virtuelle Objekte relativ zur realen Umgebung dargestellt werden. In üblichen AR Anwendungen besteht das Problem, dass die Objekte keine absolute Position in der realen Welt besitzen, sondern relativ zum Betrachter liegen. Eine mögliche Lösung dafür bietet das *ARToolkit*, indem es Computer Vision Algorithmen verwendet. Dabei berechnen die Video-Tracking-Bibliotheken die tatsächliche Kameraposition und -ausrichtung des AR-Geräts in Bezug auf physikalische Marker.

6.4. Pupil Software

Pupil besitzt mehrere Software, die für Pupil Headsets genutzt werden können. Es gibt Pupil Capture, Pupil Player und Pupil Service, wobei die erstere genannte für diese Arbeit eingesetzt wurde und im folgenden genauer erläutert wird. Die Pupil Capture erfasst und verarbeitet die Videostreams in Echtzeit, die er von der Weltkamera und Augenkamera erhält. Dies ermöglicht die Erkennung der Pupillen in Echtzeit, Aufnahme sowie die Verfolgung der Gaze. Beim Öffnen der Pupil Capture erscheint das Hauptfenster (siehe Abbildung 6.1), das Live-Video-Feeds der Weltkamera anzeigt. In Abbildung 6.1 stellt der rot markierte Bereich (1) die Leistungsdiagramme dar, davon werden die CPU, FPS und die Konfidenz der Pupillen angezeigt. Für eine optimale Performance der Eye-Tracker müssen die Konfidenzen für beide Augen in der Nähe von 1.0 betragen. Dafür müssen manchmal Anpassungen an den einzelnen Eye-Trackern vorgenommen werden. Dazu werden die Augen-Fenster geöffnet (siehe Abbildung 6.2) und die Helligkeit der Eye-Tracker eingestellt. Zudem müssen unter Umständen auch die Pupil min und max-Werte verändert werden. Die weiteren

⁵<https://github.com/qian256/HoloLensARToolKit>

6. Verwendete Hard- und Software

Software-Komponenten von Pupil werden kurz erläutert. Pupil Player gibt die Video- und Blickdaten, die mit Pupil Capture aufgenommen wurden, wieder und bietet dafür eine Visualisierung [KPB14]. Pupil Service hat die gleiche Funktion wie die Pupil Capture mit der Ausnahme, dass es keine GUI besitzt, da es nur über Netzwerkbefehle gesteuert werden soll. Der Source Code für die Pupil Software wurde zunehmend in Python geschrieben, während die einzelnen Module in C geschrieben wurden.

Für die Erkennung der Pupillen wird die "dark PupilMethode verwendet [KPB14]. Hierbei nimmt die Augenkamera Videos innerhalb eines bestimmten Bereichs des IR-Spektrums auf. Dabei verwendet die Augenkamera eine RI-Bandpassfilter mit einer oberflächenmontierten IR-LED, um das Auge des Benutzers zu beleuchten. Dadurch kann der Pupillen-Erkennungs-Algorithmus die dunkle Pupille lokalisieren. Der Algorithmus ist im Gegensatz zu üblichen Eye-Trackern unabhängig von der Hornhaut-Reflexion. Zudem wird die Genauigkeit des Algorithmus nicht durch die Brille oder Kontaktlinsen beeinflusst. Besonders an dem Erkennungs-Algorithmus ist außerdem die Verbesserung der Erkennung basierend auf dem Feedback, der durch Augen-Kamera-Videos des Benutzers gesammelt wird. Die Erkennung der Pupille basiert zunächst durch Umwandlung des Bild in Graustufen. Weitere Informationen können aus der Arbeit von Kassner et al. [KPB14] entnommen werden.

Um die Gaze-Position zu erhalten, werden die ermittelten Pupillen-Positionen von den Augen auf den Szenen-Raum anhand einer Übertragungsfunktion abgebildet, die aus zwei bivarianten Polynomen mit einstellbarem Grad besteht. Dabei werden die Parameter der Polynome ermittelt, indem eine Kalibrierung des Eye-Trackers durchgeführt wird.[KPB14]

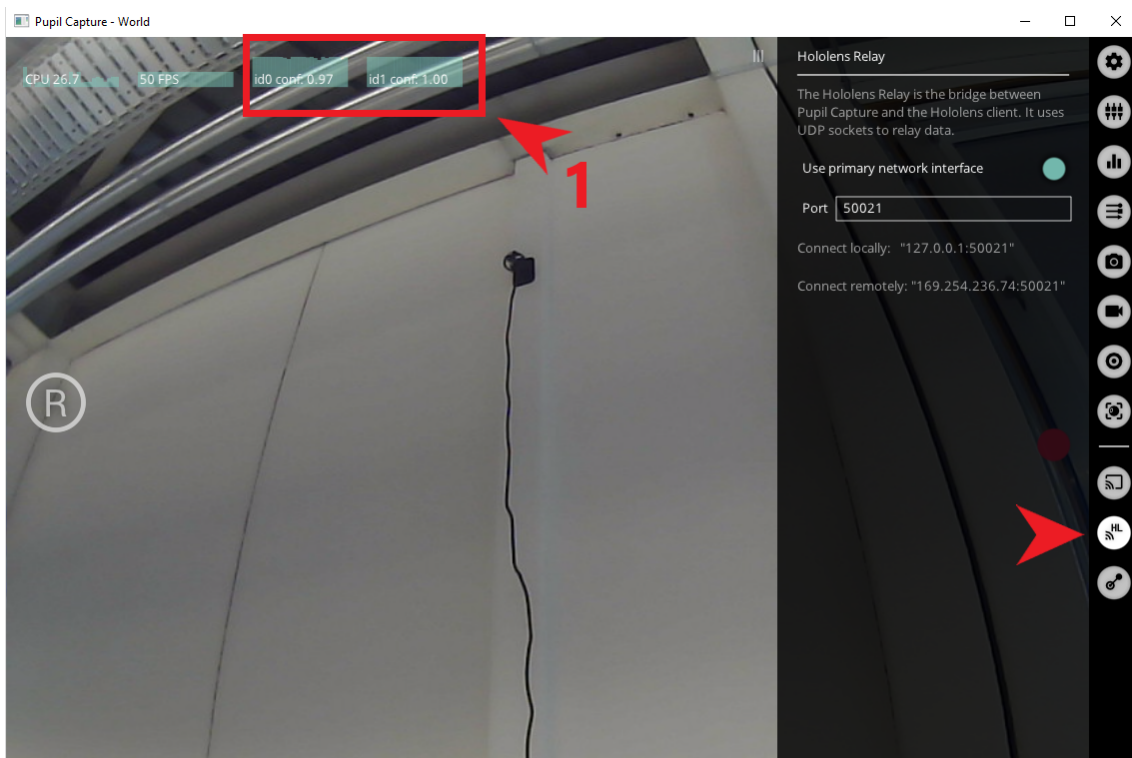


Abbildung 6.1.: Pupil Capture Haupt-Fenster

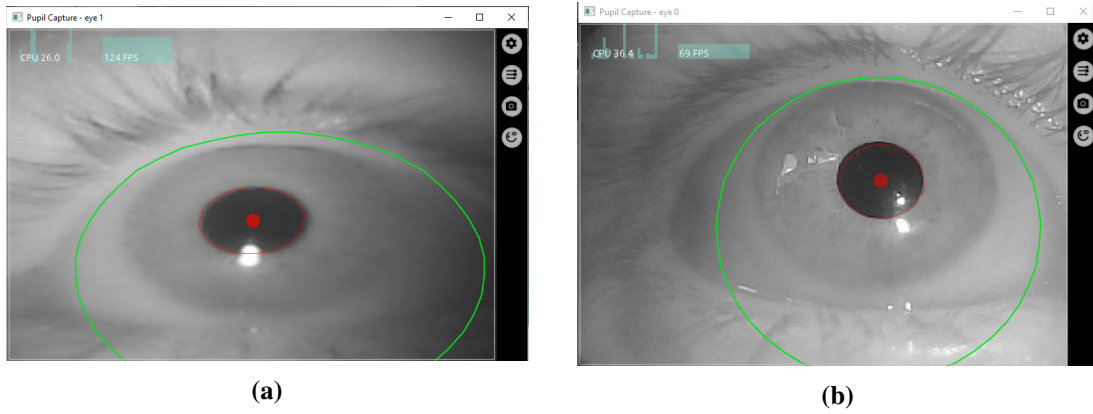


Abbildung 6.2.: Die Pupil Capture Fenster der einzelnen Eye-Tracker.

7. Realisierung der Testszene

Die Testszene wird durch die reale und virtuelle Szene realisiert. Dafür sind die Erstellung der Aufgaben und die Implementierung einer Anwendung in Unity notwendig.

7.1. Erstellung der Aufgaben

In Kapitel 5 wurde bereits erwähnt, welche Arten von Aufgaben verwendet werden. Die Gestaltung der Wort- und Zahlenrätsel lehnt sich an die aus dem Alltag bekannten Rätsel an. Die Komplexität der Rätsel wurde darauf ausgelegt, sodass sie durchschnittlich in einer Minute gelöst werden können. Die Aufgaben wurden in Inkscape erstellt. Für die reale Szene wurden sie ausgedruckt und die Bücher damit eingebunden, für die virtuelle Szene wurden sie als Textur für die virtuellen Bücher verwendet. Dabei wurden die Bücher der virtuellen Szene aus dem Assetstore kostenlos heruntergeladen. Die Größe wurde nachträglich an die Größe der realen Bücher angepasst.

7.2. Implementierung - Vorgehen

Für die Studie muss eine Anwendung implementiert werden, die auf der HoloLens funktioniert und für beide Szenen verwendet werden kann. Dabei müssen verschiedene Aspekte der Implementierung berücksichtigt werden. Zunächst muss aufgrund der Tatsache, dass die HoloLens und der Eye-Tracker getrennte Hardware-Geräte sind, die nicht miteinander interagieren, eine Kommunikation zwischen den beiden Geräten aufgebaut werden (siehe Abschnitt 7.3). Erst nach einer erfolgreichen Verbindung ist die HoloLens in der Lage, die Gaze-Daten, die vom Eye-Tracker gemessen werden, zu verarbeiten. Dazu wird ein Paket von Pupil Labs angeboten, die durch einen Import in das Unity-Projekt ein pupilplugin-Ordner erstellt. Darin befindet sich eine Szene, welche die Kalibrierung der Eye-Tracker durchgeführt wird (siehe Abschnitt 7.4). Die Szene lädt daraufhin die Testszene. Hier muss folgendes beachtet werden: Da die virtuellen Bücher zu Beginn nicht auf dem realen Tisch positioniert sind, muss zunächst eine Transformation mittels AR-Marker durchgeführt werden (siehe Abschnitt 7.5). Anschließend findet die eigentliche Studie statt. Dazu müssen neben den Gaze-Daten noch weitere Daten aufgezeichnet werden, die in einer .CSV-Datei gespeichert werden (siehe Abschnitt 7.6). Die Unterscheidung der virtuellen und realen Szene findet dabei über die Art der Visualisierung von den Büchern statt.

7.3. HoloLens – Eye-Tracker Kommunikation

Da die HoloLens und der Eye-Tracker zwei getrennte Hardware-Geräte sind, muss zusätzlich eine Verbindungsschnittstelle zwischen den beiden Geräten existieren. Dafür stellt Pupil Labs einen Plugin bereit, das HoloLens Relay [Laba], das sich in der Software Pupil Capture befindet. Das Plugin ermöglicht, dass die beiden Geräte über UDP kommunizieren. Nachdem Pupil Capture auf Rechner, mit welchem der Eye-Tracker verbunden ist, gestartet wurde, wird ein UDP-Socket erstellt, wobei die HoloLens den Client repräsentiert, während das Plugin den Server darstellt. Über den Socket werden Anfragen und Antworten an den Client gesendet. Für eine solche Verbindung in Unity wird ein HMD-Eyes Paket¹ von Pupil Labs bereitgestellt, die die notwendigen Skripte sowie eine Kalibrierungs-Szene enthält. Da für die Kalibrierung ebenfalls eine Verbindung mit Pupil Capture notwendig ist, werden die Skripte in dieser Szene verwendet. Für eine erfolgreiche Verbindung müssen zunächst Anpassungen in der Szene vorgenommen werden. Dafür gibt es im Pupil Manager Objekt die PupilSettings Komponente, worin die IP-Adresse des Computers, mit dem der Eye-Tracker verbunden ist, und ein freier Port eingegeben werden [Laba]. Außerdem ist zu beachten, dass die HoloLens und der Computer sich im gleichen Netzwerk befinden müssen, wobei das Netzwerk die Erlaubnis für den Transport von UDP geben muss. Nachdem die Verbindung hergestellt wurde, findet die Kalibrierung statt. Nach einer erfolgreichen Kalibrierung wird der Zugriff auf die Gaze gewährt. Um in der Testszene die Gaze verwenden zu können, muss auf die Thema „Gaze“, die über eine Abonnement-Methode aufgerufen wird, zugegriffen werden. In der Theorie sind keine weiteren Anpassungen notwendig, um auf die Gaze-Daten zuzugreifen. Die 3D Gaze-Positionen werden ursprünglich durch 16 Byte große Pakete über UDP an Unity gesendet. Jedoch bestand hier das Problem, dass Unity auch manchmal 17 Byte große Pakete bekommen hat, die nicht verarbeitet werden konnten. Das Problem konnte behoben werden, indem im UDPCommunication Skript eine Anpassung vorgenommen wurde, bei dem das erste Byte des 17 Byte großen Pakets ignoriert wurde.

7.4. Kalibrierung

In der HMD-Eyes Szene ist ein Kalibrierungsprozess enthalten. Darin kann zunächst der Kalibrierungsmodus eingestellt werden, ob eine 2D oder 3D Kalibrierung stattfinden soll. In Pupil Settings können die Tiefen, die bei der Kalibrierung gemessen werden, angepasst werden. Für diese Arbeit wurden sie auf die Bücher angepasst mit: 0,9, 1,5, 2,0, 3,5 m. Außerdem wurden 90 Datenpunkte pro Tiefe gemessen. Damit nach der Kalibrierung die Testszene geladen werden kann, wird diese in Pupil Manager angegeben. In Abbildung 7.1 wird die Kalibrierungsszene dargestellt.

7.5. Büchertransformation

In Unity werden die Objekte immer relativ zum Ursprung positioniert. Der Ursprung wird über die HoloLens definiert, indem die Stelle an der die Anwendung auf der HoloLens startet als Ursprungspunkt gesetzt wird. Deshalb werden Marker verwendet, die eine feste Positionierung

¹<https://github.com/pupil-labs/hmd-eyes/releases/tag/v1.0-beta>

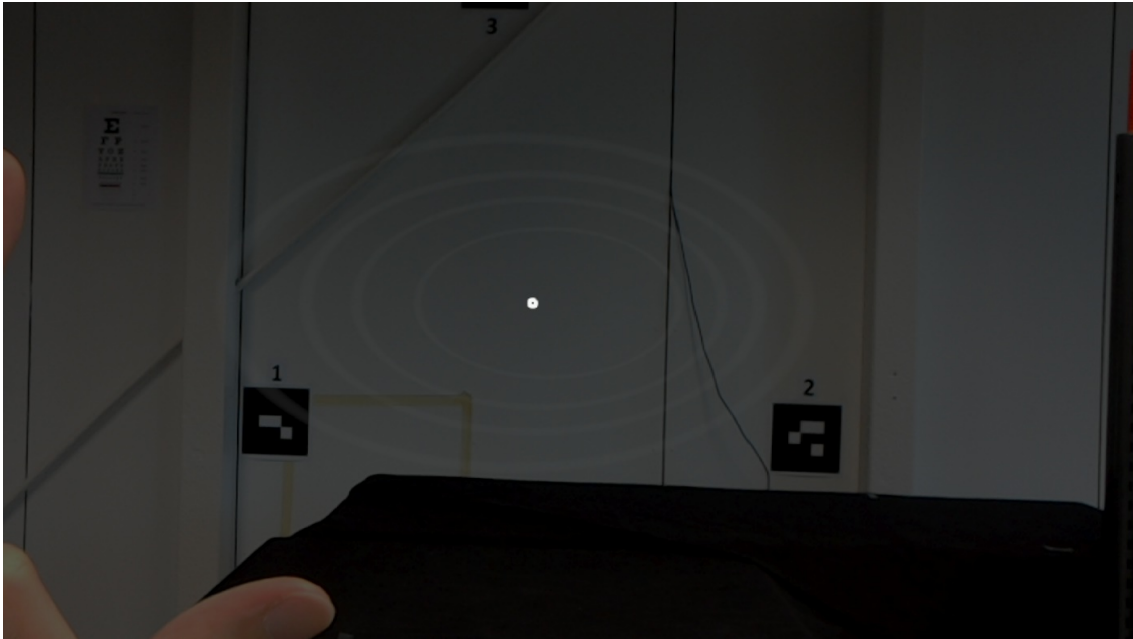


Abbildung 7.1.: Kalibrierungsszene der Eye-Trackers.

von virtuellen Objekten in der realen Umgebung ermöglichen. Zur Realisierung der Testszene werden fünf Marker verwendet, die von *ARToolkit*² bereitgestellt werden. Dazu müssen sie in der realen Umgebung platziert werden um anschließend von der HoloLens erkannt zu werden und die Buchpositionen relativ zu den erkannten Marker-Positionen auszurichten. Für diesen Vorgang muss zunächst das *ARToolkitUWP* Paket³ in das Unity Projekt importiert werden, welches die benötigten Skripte zur Erkennung der Marker beinhaltet. In einem ARUWP Controller Objekt werden die Skripte verwendet, wobei das ARUWP Marker Skript für jeden einzelnen Marker angelegt wird. Hier werden die Größe der Marker, der jeweilige Marker-Typ (mit ID) und die Visualisierung eines Objektes, nachdem der Marker erkannt wurde, bestimmt. Innerhalb dieses Objektes findet also die Erkennung des jeweiligen Markers mit der entsprechenden ID statt, deren Position bestimmt wird. Die Positionen der Marker werden in den entsprechenden MarkerAnchor Objekten gespeichert, wobei solche Objekte Kindknoten des MarkerAnchorManager Objektes sind. Nachdem in MarkerAnchor die Position des Markers 20-mal erfasst wurde, wird die durchschnittliche Position als Marker-Position verwendet und an den Manager geschickt. Außerdem ändert das virtuelle Objekt, welches sich unterhalb des Markers befindet, seine Farbe von rot auf grün wenn es 20-mal erkannt wurde (siehe Abbildung 7.2).

Wenn der Manager alle fünf Marker-Positionen erhalten hat, findet die Transformation der Bücher statt. Dazu wird ein Root Objekt angelegt, welches die Bücher als Kindknoten enthält. Auf das Root Objekt wird eine Transformation angewendet, indem die Single Value Decomposition (SVD) [AHB87] verwendet wird. Mit Hilfe von SVD wird berechnet, welche Translation und Rotation benötigt wird, um die Position der Bücher relativ zu den Markern zu verschieben.

Die Transformation ist anfangs nicht reibungslos erfolgt. Dies wurde durch die HoloLens verursacht,

²<https://github.com/artoolkit/artoolkit5/tree/master/doc/patterns>

³<https://github.com/qian256/HoloLensARToolKit>

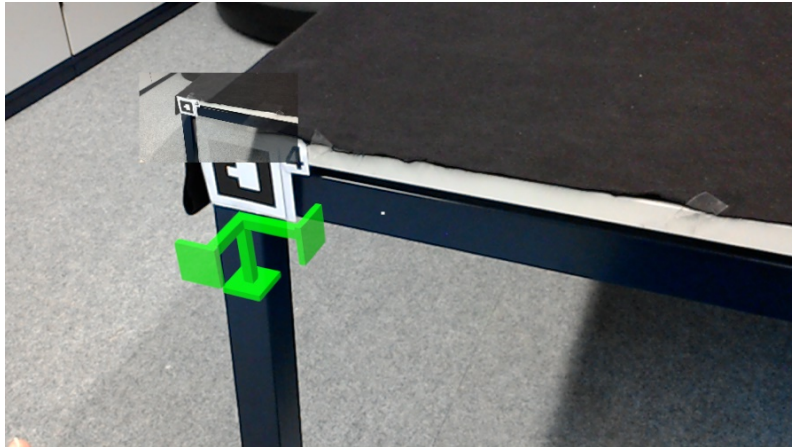


Abbildung 7.2.: AR Marker, das durch die HOloLens erkannt wurde.

da sie die Position der Marker bzw. die Abstände zwischen den Markern nicht richtig erkennen konnte. Dadurch wurden die Bücher komplett falsch positioniert. Deshalb wurden die von der Hololens erkannten ungefähren Abstände als die tatsächlichen Abstände zwischen den Markern in Unity eingetragen. Die Anpassung hat eine Verbesserung der Transformation bewirkt. Für eine optimale Transformation wurde zusätzlich ein World-Up Vektor definiert, um die Bücher alle auf einer Ebene darzustellen. In Situationen, bei denen die Hololens die Marker vollkommen fehlerhaft erkennt, werden die Bücher auch falsch transformiert. Deshalb wurde die Funktionalität hinzugefügt, bei dem über Tastendruck die Transformation rückgängig gemacht werden kann.

7.6. Datenerfassung

Für eine Analyse der Studie sind das Sammeln von Daten essentiell. Deshalb werden die Daten in einer CSV-Datei gespeichert, die sich im lokalen Cache der Hololens befindet. Beim Start der Anwendung wird immer eine neue Datei mit einer eindeutigen ID angelegt. Eine eindeutige Zuordnung der Gaze-Positionen und Bücher wird sichergestellt, indem zuerst mit dem Headgaze Cursor auf das entsprechende Buch geklickt werden muss um die Daten zu speichern. Durch erneutem Klick hört die Aufzeichnung auf.

8. Studie

Das Ziel der Studie ist es zu untersuchen, ob AR Einfluss auf die Tiefenwahrnehmung hat. Dabei soll gezeigt werden, ob die Hypothese H1: „Es existiert ein Unterschied der wahrgenommenen Blicktiefe in AR im Gegensatz zur Realität.“ statistisch belegt werden kann.

8.1. Studiendesign

Für die Studie wurde das „within-subject“ Studiendesign gewählt. Bei dieser Studiendesignart nehmen alle Studienteilnehmer an allen Versuchsbedingungen teil. Die Studie beinhaltet eine reale und virtuelle Szene (siehe Abbildung 8.1), in denen Bücher in verschiedenen Abständen auf einem Tisch positioniert sind. Die Studienteilnehmer mussten in beiden Szenen die Wörter- und Zahlenrätsel, die jeweils auf den Büchern abgebildet sind, lösen. Während der Studie trugen sie die Microsoft HoloLens (1st Gen) mit eingebautem Pupil Labs Eye-Tracker, sodass die Tiefe beim Lösen der Aufgaben gemessen wurde. Damit die Reihenfolge der Versuchsbedingungen keine systematischen Auswirkungen auf das Verhalten der Studienteilnehmer und somit auf das Ergebnis hat, wurde die Randomisierung berücksichtigt. Dazu wurde zu Beginn definiert, dass Studienteilnehmer mit der ID1 bis ID14 mit der realen Szene beginnen, während die restlichen (ID15 bis ID28) mit der virtuellen Szene beginnen. Da die Studienteilnehmer ihre IDs selber aus einem Behälter ziehen mussten, wurde die Randomisierung gewährleistet. Für eine Studie müssen verschiedene Variablen definiert werden. Dazu gehören die unabhängige, abhängige und Störvariablen. Die ersteren beiden genannten Variablen sind abhängig voneinander, da die unabhängige Variable eine Variable ist, die man verändert während die abhängige diejenige ist, die die Veränderung – hervorgerufen durch die unabhängige Variable – misst. In dieser Studie werden AR und die Realität miteinander verglichen und stellen deshalb die unabhängige Variable dar. Die Veränderung der Umgebung wird durch die wahrgenommene Tiefe mit Hilfe eines Eye-Trackers gemessen. Sie stellt die abhängige Variable dar. Eine weitere Variable ist die Störvariable, die ebenfalls einen Einfluss auf die abhängige Variable hat aber nicht gemessen werden kann. In dieser Studie wird zunächst die Eye-Tracking Qualität aufgeführt, da diese sehr ungenaue Ergebnisse bezüglich der Tiefe liefert. Dabei ist das Maß der Ungenauigkeit nicht bekannt. Eine weitere Störvariable stellt die Erkennung von Markern dar. Hier besteht das Problem, dass die Bücher bei der realen und virtuellen Szene nicht identisch platziert werden, da durch die ungenaue Erkennung der Marker die virtuellen Bücher ungenau transformiert werden. Dabei beträgt die Genauigkeit mindestens 80%. Um eine mögliche Störvariable zu verhindern, wurde der Studienzweck vor den Teilnehmern geheim gehalten. Weitere Störvariablen wurden weitestgehend eliminiert, indem die Studie in einem Labor unter kontrollierten Bedingungen stattfand und die Randomisierung berücksichtigt wurde.

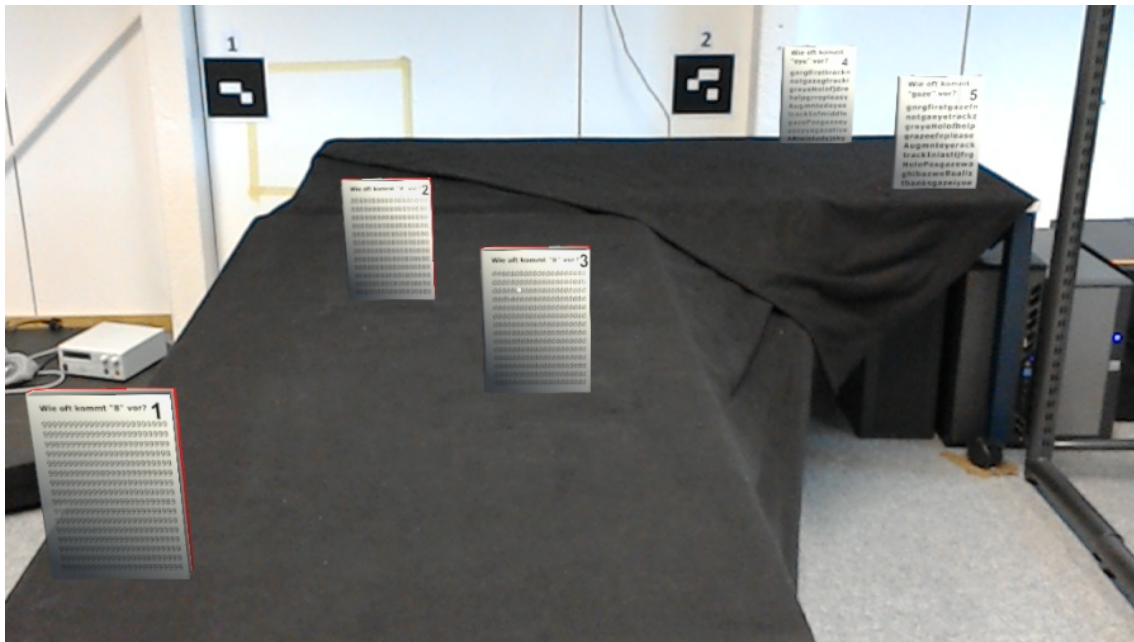


Abbildung 8.1.: virtuelle Testszene

8.2. Studienablauf

Die Studie hat in einem Laborraum am Visualisierungsinstitut der Universität Stuttgart stattgefunden.

Nachdem die Studienteilnehmer begrüßt und darauf hingewiesen wurden, ihr Handy auszuschalten, haben sie eine Aufwandsentschädigung von 10€ erhalten und mussten sich dazu in einer Liste eintragen. Danach mussten sie die Einverständniserklärung unterschreiben. Zusätzlich zu der Einverständniserklärung wurden sie nochmals darüber informiert, dass die Daten anonym verarbeitet werden. Deshalb durften sie aus einem Behälter ihre ID ziehen. Zudem wurde der Zweck der Studie nochmal mündlich beschrieben. Dabei wurde ihnen der echte Zweck der Studie nicht verraten, sondern nur allgemein erklärt, dass es sich um einen Vergleich der realen und virtuellen Umgebung handelt, bei dem der Einfluss von AR untersucht wird. Als nächstes haben sie den ersten Teil des Fragebogens ausgefüllt, bei dem sie ihre demografischen Daten angegeben haben. Um ihr Sehvermögen zu überprüfen, wurde ein Sehtest durchgeführt. Dazu wurde der Snellentest verwendet, bei dem eine Reihe von Zahlen in unterschiedlichen Größen aus einer Entfernung von 3m betrachtet wurden. Nur bei erfolgreichem Bestehen des Testes wurde die Studie fortgesetzt.

Als nächstes wurden diejenigen, die nicht wussten, was eine HoloLens ist, diese erläutert. Dazu wurden zunächst allgemeine Informationen über die AR Technologie und die Visualisierung von 3D-Objekten gegeben. Anschließend wurde die Bedienung der HoloLens erläutert und unter anderem die Air-tab und Blütengeste beigebracht. Schließlich durften die Teilnehmer die HoloLens vorsichtig aufsetzen. Die HoloLens wurde so eingerichtet, sodass sie auf dem Kopf rutschfest befestigt wurde und das ganze Display im Sichtfeld lag. Anschließend wurde Pupil Capture geöffnet um den Eye-Tracker einzustellen. Dazu wurden zunächst die Eye-Tracker-Komponenten so verschoben,

dass beide Augen durch die Kamera gesehen werden. Anschließend mussten Anpassungen im Capture vorgenommen werden. Dabei wurden zunehmend die Exposure Time und die minimalen und maximalen Pupillengrößen angepasst.

Nachdem alle Einstellungen abgeschlossen wurden, konnte die Kalibrierung der HoloLens stattfinden. Die Kalibrierung der HoloLens (siehe Abbildung 8.2) war erforderlich, damit die Hologramm Anzeige entsprechend dem Augenabstand des Benutzers eingestellt werden konnte.

Ab diesem Zeitpunkt mussten die Teilnehmer zu dem Bereich, wo die eigentliche Studie stattfindet. Dafür mussten sie sich auf den Stuhl setzen, welches sich direkt vor dem Tisch, auf dem die Bücher platziert waren, befand.

Nach erfolgreicher Kalibrierung der HoloLens konnte die für diese Arbeit implementierte Anwendung gestartet werden. Während die Anwendung gestartet wurde, wurden die Teilnehmer darauf hingewiesen, dass sie sich während der Studie nicht unnötig bewegen sollen und die HoloLens nicht verschieben dürfen. Anschließend wurde das Vorgehen für die Eye-Tracking Kalibrierung erläutert. Dabei wurde erwähnt, dass fünf Ellipsen in verschiedenen Größen und Tiefen erscheinen werden und auf diesen ein weißer Punkt auftauchen wird. Sie sollten auf diesen Punkt schauen bis dieser verschwindet und auf einer anderen Stelle wiedererscheint. Wenn die Kalibrierung erfolgreich war, konnte die jeweilige Szene geladen werden.

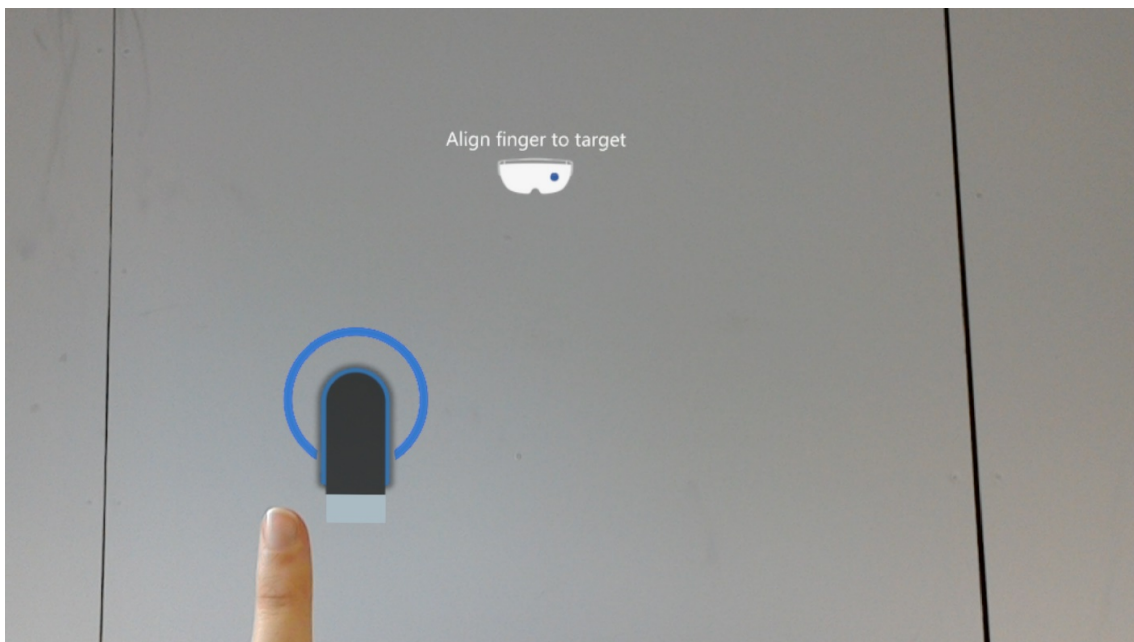


Abbildung 8.2.: Kalibrierungsszene der HoloLens

Abhängig von der ID des jeweiligen Teilnehmers fand entweder die Studie zuerst mit der realen oder virtuellen Szene statt. Bei beiden Möglichkeiten musste die gleiche Prozedur durchlaufen werden.

Die Vorgehensweise sah folgenderweise aus:

Nachdem die Kalibrierung beendet und eine neue Szene geladen wurde, wurde den Teilnehmern erklärt was gerade auf dem Display erscheinen sollte: Sie konnten links oberhalb des Displays eine kleine Anzeige sehen, auf der die Kamera der HoloLens erkannt wurde. Anschließend wurde kurz erklärt, dass die Bilder mit schwarz-weißen Mustern, die an der Wand aufgehängt sind,

als Marker bezeichnet werden. Sie sollten mit der HoloLens-Kamera die Marker jeweils einzeln anschauen, bis das rote virtuelle Objekt direkt unterhalb vom jeweiligen Marker, sich grün färbt. Da die einzelnen Marker nummeriert waren, wurde ihnen gesagt, dass sie alle Marker nacheinander anschauen sollen. Die Transformation fand statt, wenn alle Marker erkannt wurden. Wenn es Probleme bei der Erkennung der Marker gab und die Transformation falsch durchgeführt wurde, musste der Markererkenntnis-Vorgang wiederholt werden. Dazu mussten die Tasten T und Z auf der Bluetooth-Tastatur betätigt werden. Bei einer erfolgreichen Transformation wurden die virtuellen Bücher auf den Tisch transformiert. Anschließend wurde die Taste U gedrückt, damit sie wirklich auf einer Ebene sind.

Nachdem die Transformation endgültig beendet war, wurden die Anzeige und die virtuellen Objekte um die Marker herum entfernt.

Nach dieser Prozedur wurde die entsprechende Szene gestartet. Für den Start der virtuellen Szene wurden die virtuellen Bücher durch Betätigen der Taste I sichtbar gemacht. Für den Start der realen Szene musste nichts gemacht werden. Nun durften die Teilnehmer mit der jeweiligen Szene beginnen. Dazu wurde ihnen folgendes erklärt:

- „Ihre Aufgabe ist es, die fünf Bücher, die jeweils von eins bis fünf nummeriert sind, nacheinander anzuschauen. Dabei müssen Sie die Aufgaben auf den Büchern lösen.“
- „Bevor Sie die Aufgabe lösen, müssen Sie folgendermaßen vorgehen: Wenn Sie das erste Buch betrachten, müssen sie den Headgaze der HoloLens auf das Buch richten und auf den Clicker drücken. Daraufhin verschwindet der Headgaze, solange es auf dem jeweiligen Buch ist. Dann dürfen Sie beginnen, die Aufgabe zu lösen. Sie müssen die Antwort laut sagen. Wenn die Antwort in Ordnung ist, dürfen Sie wieder auf den Clicker drücken und der Headgaze erscheint wieder. Das gleiche wiederholen Sie für alle Bücher.“

Damit die Gaze-Positionen der Teilnehmer gespeichert wurden, musste zum Starten der Aufzeichnung die Taste R betätigt werden. Nachdem man am Ende erneut auf R gedrückt hat, wurde die Aufnahme gespeichert.

Nachdem die Teilnehmer die Aufgaben der jeweiligen Szene bearbeitet hatten, musste sie einen Moment warten, bis die nächste Szene vorbereitet wurde. Wenn es sich bei der ersten Szene um die virtuelle gehandelt hat, mussten die virtuellen Bücher wieder unsichtbar gemacht werden und die realen Bücher auf den Tisch platziert werden. Andernfalls musste genau andersherum vorgegangen werden.

Wenn beide Szenen bearbeitet wurden, durften die Teilnehmer die HoloLens wieder absetzen. Zum Schluss durften sie den zweiten Teil des Fragebogens ausfüllen.

8.3. Probanden

Um den Stichprobenumfang zu bestimmen wurde eine a-priori Analyse durchgeführt, da dieser später bei der Analyse benötigt wird, um die Größe des Effekts von inferentieller Statistik zu erkennen. Dazu muss man die Effektgröße, den alpha-Wert und die Power bestimmen. Dabei beschreibt der alpha-Wert die Wahrscheinlichkeit, ein signifikantes Ergebnis zu erhalten obwohl keines existiert. Die Power ist die Wahrscheinlichkeit, ein wahres signifikantes Ergebnis zu erhalten.

Die Standardwerte für $\alpha=0.05$ und $\text{power}=0.8$ wurden auch in dieser Studie gewählt. Außerdem wurde ein großer Effekt angestrebt, sodass die Effektgröße auf $d=0.8$ gesetzt wurde. Die Analyse hat für einen zweiseitigen t-Test ergeben, dass der benötigte Stichprobenumfang $n=25.52$ beträgt.

An der Studie haben 26 Personen teilgenommen. Davon waren 12 weiblich und 14 männlich. Dabei wurden die Daten von einem männlichen Teilnehmer aufgrund von Hardware-Ausfall während der Studie von der Analyse ausgeschlossen. Von den in die Analyse einbezogenen Personen betrug das durchschnittliche Alter 23.28. Die Altersspanne lag zwischen 18 und 33. Unter den Teilnehmern berichteten 13 von korrigierten Sehstärke, davon waren 11 Brillenträger und 2 trugen Kontaktlinsen.

Vor der eigentlichen Studie musste ein Sehtest (Snellentest) durchgeführt werden. Der Test ergab, dass 14 Personen eine normale Sehkraft hatten. Bei den restlichen Personen wurde eine geringe Sehschwäche beobachtet, aber aufgrund der Berücksichtigung, dass die Schrift auf den Büchern verglichen mit dem Test viel größer war, wurde die Studie fortgesetzt.

Unter den Teilnehmern gab es 14 Bachelorstudenten, 6 Masterstudenten, 2 Studenten, die ein Staatsexamen anstreben, 1 Person mit abgeschlossener Ausbildung und 1 Person mit einer Hochschulreife. Davon studierten 16 Personen in einem Studiengang im IT-Bereich (Informatik, Softwaretechnik, Medieninformatik). Erste Erfahrungen in AR oder VR hatten 20 Personen unter den Teilnehmern. Dabei haben die meisten als AR/VR-Hardware die Oculus Rift (7), die HTC VIVE (6), Windows VR-Brillen (5) und die HoloLens(3) zuvor verwendet. Nur drei Personen gaben an, dass sie regelmäßig ein AR/VR-Gerät verwenden.

8.4. Auswertung

Nachdem die Studie wie in den letzten Abschnitten beschrieben, stattgefunden hatte, wurden daraus Daten gesammelt. Diese mussten anschließend analysiert und ausgewertet werden. Für die Analyse und die entsprechenden Graphen wurde RStudio, eine Entwicklungsumgebung für die Programmiersprache R, die für statistische Berechnungen und Grafiken benötigt wird, verwendet.

8.4.1. Beobachtungen

Als die Teilnehmer die Aufgaben auf den virtuellen und realen Büchern gelöst haben, wurden die ihre Lösungen notiert. In Abbildung 8.3 werden die Lösungen miteinander verglichen. Bei den Büchern 1 bis 3 wurden in der virtuellen Szene die Aufgaben durchschnittlich mehr richtig gelöst. Buch 2 in der virtuellen Szene wurde von allen Teilnehmenden richtig gelöst. Die Bücher 4 und 5 wurden in der realen Szene durchschnittlich mehr richtig gelöst.

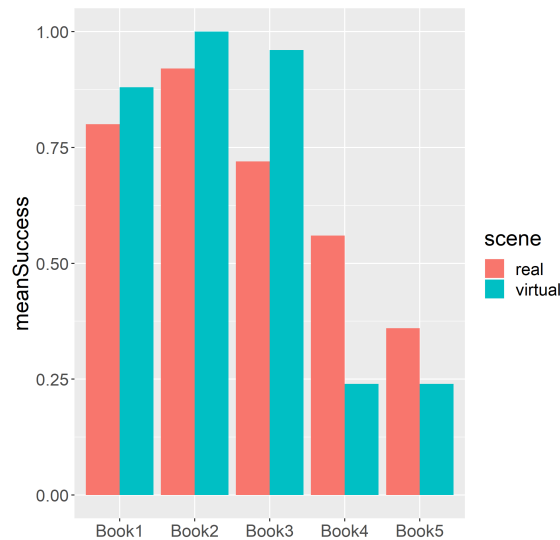


Abbildung 8.3.: Durchschnittlicher Erfolg beim Lösen der Aufgaben auf den Büchern

8.4.2. Auswertung des Fragebogens

Im zweiten Teil des Fragebogens wurde gefragt, ob die Teilnehmer die realen und virtuellen Bücher gut erkennen konnten. Unter den 25 Studienteilnehmern gab es eine Person, die das reale Buch 2 nicht erkennen konnte. In der virtuellen Szene hat eine Person Buch 1 nicht gut erkannt, zwei Personen konnten Buch 2 und Buch 3 nicht gut erkennen. Den Teilnehmern wurde gefragt, ob sie Probleme beim Fokussieren der Bücher hatten. Fünf Teilnehmer hatte Probleme, wobei ein Teilnehmer angemerkt hatte, dass das Problem bei den virtuellen Büchern bestand. Auf die Frage, ob sie Unterschiede bei den echten und virtuellen Büchern erkannt haben, haben bis auf zwei alle mit Ja beantwortet. Als Ursache wurden meistens die Aussage gemacht, dass die virtuellen Bücher durchschaubarer sind und eine gewisse Transparenz haben. Zudem wurden die realen Bücher schärfer empfunden. Bei den virtuellen Büchern wurde von zwei Personen behauptet, dass die Bücher leuchten. Weitere Aussagen, die von einzelnen Personen gemacht wurden sind, dass

- sie ein leichtes Flackern der virtuellen Bücher wahrgenommen haben,
- sie eine Verfärbung bei den virtuellen Büchern erkennen konnten und
- die virtuellen Bücher an den Seiten leicht "nebelig" waren.

Zusätzlich mussten die Teilnehmer anhand einer Likert-Skala bewerten, wie ähnlich sie die virtuellen und realen Bücher fanden. In Abbildung 8.4 wird beobachtet, dass zum Großteil die Bücher sehr ähnlich empfunden wurden. Die durchschnittliche Bewertung beträgt 4.04.

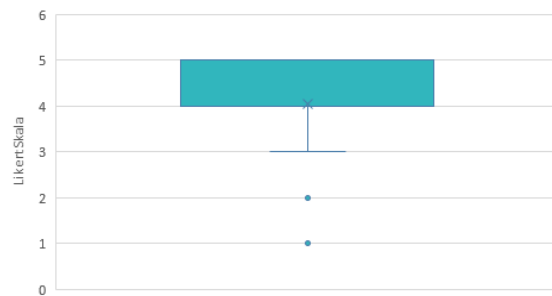


Abbildung 8.4.: Der Boxplot zeigt die Bewertung der Ähnlichkeit zwischen der virtuellen und realen Bücher. Dabei bedeutet 1, dass es überhaupt nicht zutrifft, 5 bedeutet, dass es vollkommen zutrifft.

Bei der Auswertung einer Studie können verschiedene Faktoren Einfluss auf das Ergebnis haben. Dazu gehört auch Situationen, in denen sich die Teilnehmer unwohl fühlen. Deshalb wurde im Fragebogen die Frage gestellt, ob sie sich während der Studie unwohl gefühlt haben. Während vier Personen dazu keine Aussage gemacht haben, kam von den meisten die Aussage, dass sie Augenbrennen und Kopfschmerzen hatten. Zudem wurde angemerkt, dass die HoloLens unangenehm warm wurde. Drei Teilnehmer litten außerdem an Übelkeit. Den Druck am Kopf, der durch HoloLens verursacht wurde, empfanden die Teilnehmer als unangenehm. Zusätzlich wurde von einer Person angemerkt, dass die Augen durch nahes fokussieren einer hohen Belastung ausgesetzt waren.

8.4.3. Ausreißer

Die während der Studie erhobenen Daten werden zunächst in einem Boxplot dargestellt (siehe Abbildung 8.5). In Abbildung 8.5a wird beobachtet, dass die Daten für beide Szenen Ausreißer enthalten, die sich bis zu -150000 m erstrecken. Da in der Studie die betrachtete Distanz maximal 3.6 m beträgt, werden Datenpunkte außerhalb von $[-10,10]$ als Messfehler betrachtet. Diese werden als Ausreißer definiert und aus dem Datensatz entfernt. In Abbildung 8.5b ist erkennbar, dass die Daten aus Boxplot-Sicht immer noch sehr viele Ausreißer enthalten, sodass sie beim zeichnen eine Linie bilden.

In den folgenden Abschnitten wird die Analyse mit den gefilterten Daten stattfinden.

8.4.4. Deskriptive Statistik

In diesem Abschnitt werden die aufgezeichneten Daten anhand von verschiedenen Messungen wie Median, Mittelwert (M) und Standardabweichung (SD) zusammengefasst und beschrieben. Daraus werden erste Zusammenhänge betrachtet, die später in der inferentiellen Statistik analysiert werden.

In Abbildung 8.5b kann die Streuung der Daten abgelesen werden. Dabei beschreibt die horizontale Linie innerhalb der farbigen Fläche den Median. Direkt unterhalb und oberhalb des Medians liegen 50% der Datenpunkte, die durch den farbig markierten Bereich beschrieben wird. Die Abbildung zeigt für beide Szenen ähnliche Werte. Die reale Szene hat einen Median von -1.06 , während der Wert bei dem virtuellen -1.19 ist. In der realen Szene ($M=-1.20$, $SD=1.87$) wird jedoch die Tiefe

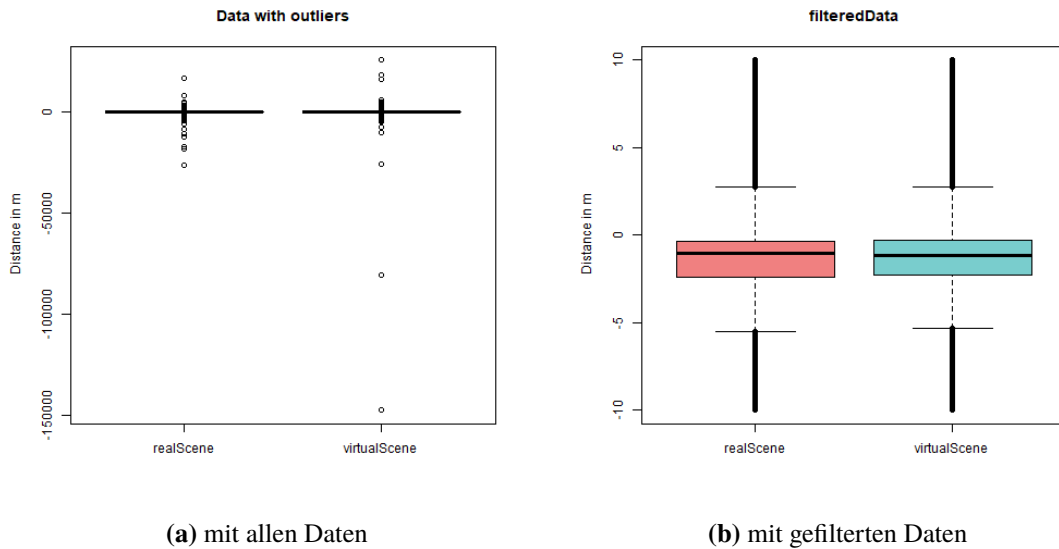


Abbildung 8.5.: Die Daten der realen und virtuellen Szene, die als Boxplots dargestellt werden.

durchschnittlich besser wahrgenommen als in der virtuellen Szene ($M=-1.26$, $SD=2.05$). Die Größe der Standardabweichung zeigt, dass die Verteilung der Datenpunkte in beiden Szenen sehr groß ist. Dabei ist die Verteilung im Negativen größer als im Positiven. Das bedeutet, dass die Teilnehmer zunehmend weiter nach vorne geblickt haben, als es das Buch ist.

Der Vergleich der einzelnen Bücher in Abbildung 8.6 zeigt, dass in Buch 1 eine geringe Streuung der Datenpunkte außerhalb des Medians vorliegt, während sie bei Buch 2 und 3 etwas stärker ist und bei Buch 4 und 5 sehr stark. In Abbildung 8.7 erkennt man, dass die Tiefe von Buch 1 in der realen Szene ($M=-0.27$, $SD=0.98$) durchschnittlich etwas besser wahrgenommen wird als in der virtuellen Szene ($M=-0.29$, $SD=1.39$). Das gleiche wird auch in Buch 2 beobachtet (reale Szene: $M=-0.98$, $SD=1.45$; virtuelle Szene $M=-1.04$, $SD=2.11$). Bei Buch 3 wird die Tiefe in der realen Szene ($M=-0.66$, $SD=1.37$) durchschnittlich minimal schlechter wahrgenommen als die virtuelle ($M=-0.65$, $SD=1.84$). Die Tiefe in der realen Szene in Buch 4 ($M=-1.97$, $SD=2.06$) wird durchschnittlich besser wahrgenommen als in der virtuellen ($M=-2.23$, $SD=2.02$). Das gleiche wird auch für Buch 5 beobachtet (reale Szene: $M=-1.47$, $SD=2.16$ virtuelle Szene: $M=-1.96$, $SD=2.11$).

Verteilung der Daten

Die Verteilung der Daten kann zum einen graphisch in Form von Histogrammen erfolgen, oder mit Hilfe von statistischen Signifikanztests. Dabei wird untersucht, ob die Daten ungefähr normal verteilt sind. Eine Normalverteilung liegt vor, wenn die Dichtefunktion eine Glockenform annimmt. In diesem Fall sind der Median und der Mittelwert identisch.

Für diese Studie wurden zwei verschiedene statistische Testverfahren eingesetzt, um zu überprüfen, ob die Daten normal verteilt sind. Bei dem ersten Verfahren handelt es sich um den Shapiro Wilk Test. Der Test wurde sowohl für die gesamten Daten der realen und virtuellen Szene angewendet, als auch für die einzelnen Bücher. Für beide Aspekte wurde keine Normalverteilung entdeckt. Bei dem zweiten Testverfahren handelt es sich um den Kolmogorov Smirnov Test. Auch hier wurde

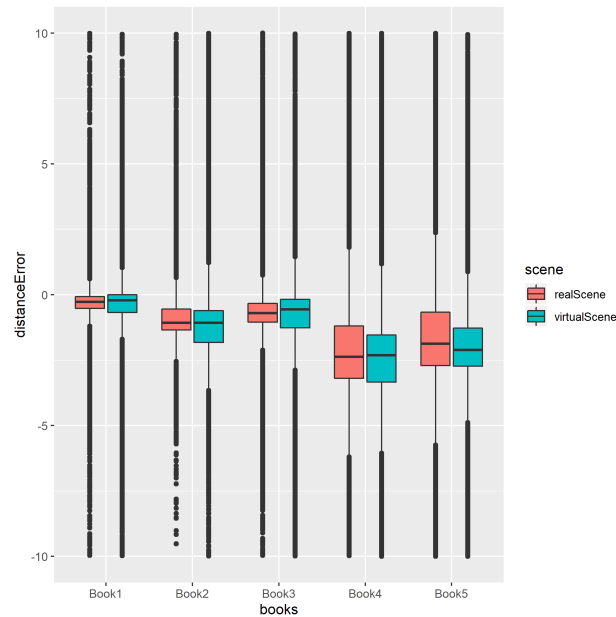


Abbildung 8.6.: Boxplot der einzelnen Bücher jeweils in den realen und virtuellen Szenen

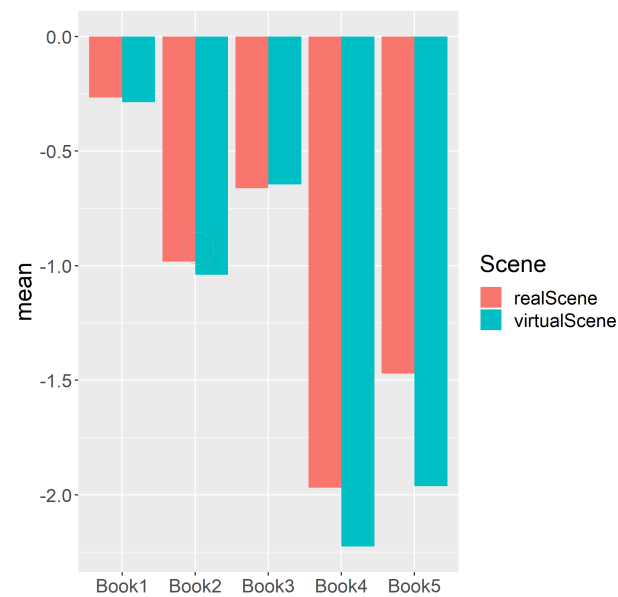


Abbildung 8.7.: Balkendiagramm von den Durchschnittswerten der einzelnen Bücher in der realen und virtuellen Szene

keine Normalverteilung erkannt. Bei statistischen Testverfahren besteht das Problem, dass sie bei zu großer Stichprobengröße keine Normalverteilung feststellen können. Deshalb ist es sinnvoll, zusätzlich eine visuelle Analyse mit Hilfe von Histogrammen zu machen.

In Abbildung 8.8 erkennt man, dass das Histogramm von den gesamten Daten keine perfekte Normalverteilung darstellt. Jedoch ist eine Kurve ersichtlich, die sich an eine Normalverteilung nähert. Für die Daten der realen und virtuellen Szene (siehe Abbildung 8.9) sowie die der einzelnen Bücher (siehe Abbildung 8.10) wird das gleiche beobachtet. Für die Auswertung der Studie werden die Daten auf zwei Arten betrachtet. Es wird eine Analyse durchgeführt, bei der die Daten als normalverteilt betrachtet werden. Hier wird ein parametrischer Test angewendet und auf einen signifikanten Unterschied untersucht. Wenn angenommen wird, dass die Daten nicht normalverteilt sind, dann muss ein nicht-parametrischer Test durchgeführt werden. In den nächsten Abschnitten werden beide Analyseverfahren eingesetzt.

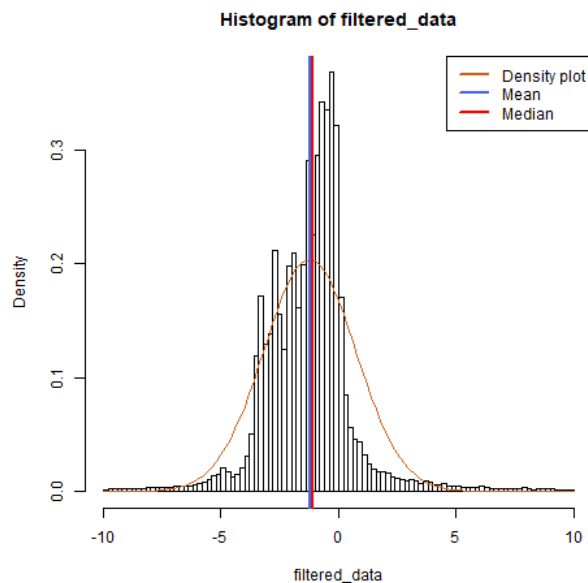


Abbildung 8.8.: Histogramm von den gesamten Daten

8.4.5. Inferentielle Statistik

Bei der deskriptiven Statistik wurde beobachtet, dass die Mittelwerte der beiden Szenen Unterschiede aufweisen. Zudem gab es zwischen den einzelnen Büchern ebenfalls einen Unterschied. Bei der inferentiellen Statistik soll nun untersucht werden, ob ein statistisch signifikanter Unterschied feststellbar ist. Sie wird zeigen, ob die Hypothese H_1 akzeptiert oder abgelehnt wird. Dazu muss zunächst eine Gegenhypothese, die Null-Hypothese, formuliert werden. Sie geht davon aus, dass Gleichheit vorliegt und wird folgendermaßen definiert:

H_0 : „Es existiert kein Unterschied zwischen AR und der Realität bei der wahrgenommenen Blicktiefe.“

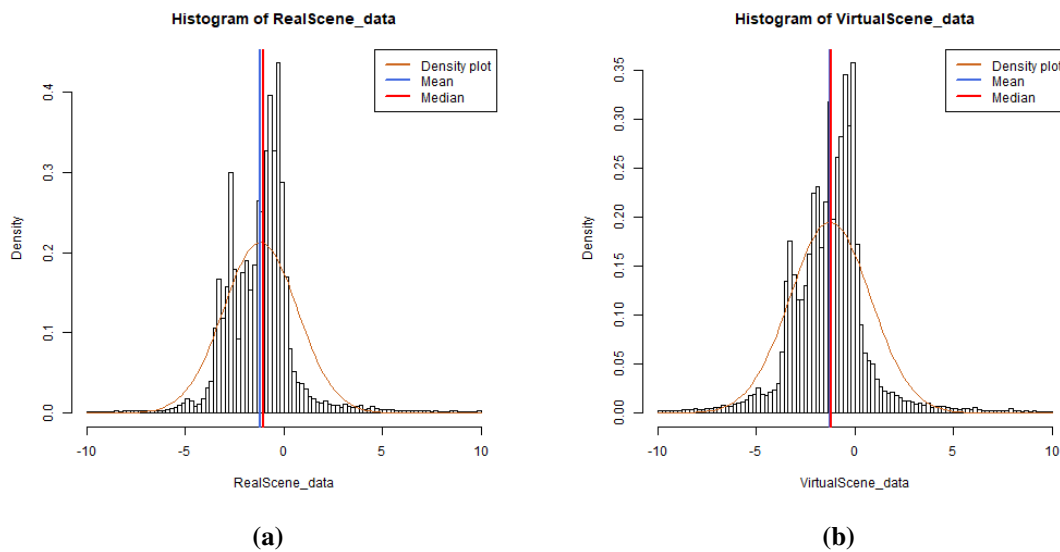


Abbildung 8.9.: Histogramm der realen (8.9a) und virtuellen Szene (8.9b)

Parametrischer Test

Es gibt verschiedene Testverfahren, die zum parametrischen Test gehören. Für die Analyse dieser Studie wird der unabhängige zwei-Stichproben t-Test verwendet. Er wird eingesetzt, um den Mittelwert von zwei Gruppen zu vergleichen, wobei die Teilnehmer nur in einer der Gruppen sein können. Obwohl in dieser Studie die Teilnehmer an beiden Versuchsbedingungen teilgenommen haben (also zu beiden Gruppen gehören), wurde ausnahmsweise ein unabhängiger t-Test durchgeführt, da die erhobenen Daten in der virtuellen und realen Szene nicht gleich lang sind.

Der unabhängige t-Test setzt neben der Normalverteilung der Daten auch die Homogenität der Varianz (Gleichheit der Populationsvarianzen) voraus. Jedoch wurde mit dem Levene's Test gezeigt, dass keine Varianzhomogenität besteht. Es muss aber nicht auf den t-Test verzichtet werden, da alternativ der zweiseitige Welch's t-Test verwendet werden kann. Bei dem Test wird versucht, die Null-Hypothese abzulehnen, sodass die Alternativ-Hypothese H_1 angenommen wird. Dazu muss der p-Wert kleiner als das Signifikanzniveau sein. Für die Analyse wurden zunächst die beiden Szenen miteinander verglichen. Es gab einen statistisch signifikanten Unterschied zwischen den Fehlerdistanzen in der realen und virtuellen Szene, wobei die Fehlerdistanz der virtuellen Szene durchschnittlich ($m_v - m_r$) -0.060m größer war ($95\%-\text{CI}[0.048, 0.07], t(368991)=9.48, p < 0.05$). Die Bücher wurden auch auf Unterschiede geprüft. In Tabelle 8.1 stehen die Ergebnisse für die einzelnen Bücher. Außer bei Buch 3 wurde bei allen anderen ein signifikanter Unterschied festgestellt.

Für die einzelnen Teilnehmer wurden auch jeweils ein t-Test durchgeführt. Das Ergebnis (Siehe Tabelle 8.2) hat gezeigt, dass alle bis auf zwei Teilnehmer einen signifikanten Unterschied aufweisen.

8. Studie

Buch	$M_v - M_r$	Konfidenzintervall CI	t-Verteilung	p-Wert
Buch 1	-0.0194	[0.0027, 0.0360]	t(76917) = 2.2871	<0.05
Buch 2	-0.0573	[0.0242, 0.0905]	t(41116) = 3.3884	<0.05
Buch 3	0.0162	[-0.0402, 0.0077]	t(64532) = -1.3261	0.1848
Buch 4	-0.256	[0.2318, 0.2816]	t(102237) = 20.197	<0.05
Buch 5	-0.490	[0.4611, 0.5208]	t(68194) = 32.217	<0.05

Tabelle 8.1.: Ergebnis des zweiseitigen Welch t-Tests von den einzelnen Büchern

ID	$M_v - M_r$	Konfidenzintervall CI	t-Verteilung	p-Wert
1	0.109	[0.0768, 0.1415]	t(7623.3) = 6.6256	<0.05
2	-0.006	[-0.0527, 0.0404]	t(7551.1) = -0.26099	0.7941
3	-0.093	[-0.2185, 0.03278]	t(6743.4) = -1.4488	0.1474
4	-0.091	[-0.1338, -0.0482]	t(14135) = -4.1638	<0.05
5	-0.535	[-0.5875, -0.4821]	t(15688) = -19.884	<0.05
7	-0.830	[-0.9228, -0.7372]	t(14554) = -17.534	<0.05
8	0.438	[0.3986, 0.4782]	t(14712) = 21.595	<0.05
9	-0.830	[-0.8826, -0.7780]	t(14993) = -31.107	<0.05
10	-0.177	[-0.2140, -0.1397]	t(15961) = -9.3219	<0.05
11	-1.660	[-1.7171, -1.6030]	t(6223.6) = -57.018	<0.05
12	-0.209	[-0.2496, -0.1692]	t(11109) = -10.214	<0.05
13	0.471	[0.3732, 0.5690]	t(9882.6) = 9.4297	<0.05
14	-2.733	[-2.9065, -2.5594]	t(7669.4) = -30.866	<0.05
15	1.937	[1.9038, 1.9703]	t(16801) = 114.17	<0.05
16	-0.520	[-0.5442, -0.4961]	t(16898) = -42.392	<0.05
18	0.901	[0.8630, 0.9396]	t(18591) = 46.09	<0.05
19	-2.073	[-2.2076, -1.9374]	t(2790.4) = -30.081	<0.05
20	1.0284	[1.0036, 1.0532]	t(21530) = 81.319	<0.05
21	-0.388	[-0.4253, -0.3498]	t(11563) = -20.138	<0.05
22	0.688	[0.6331, 0.7423]	t(15788) = 24.68	<0.05
23	0.176	[0.1284, 0.2230]	t(10272) = 7.281	<0.05
24	0.0518	[0.0152, 0.0886]	t(13014) = 2.7718	<0.05
26	-0.859	[-0.9197, -0.7988]	t(11744) = -27.872	<0.05
27	-0.888	[-0.9170, -0.8598]	t(12902) = -60.978	<0.05
28	1.669	[1.5871, 1.7509]	t(5564.1) = 39.95	<0.05

Tabelle 8.2.: Ergebnis des zweiseitigen Welch t-Tests von den einzelnen Studienteilnehmern

Nicht-parametrischer Test

Der Kruskal-Wallis-Test ist ein nicht-parametrischer Test, der als eine mögliche Alternative für den unabhängigen zwei-Stichproben t-Test verwendet wird, wenn die Anforderungen an einen parametrischen Test nicht erfüllt sind. Er wurde für die Analyse verwendet. Der Test hat gezeigt, dass es einen statistisch signifikanten Unterschied zwischen den Fehlerdistanzen in der realen und

virtuellen Szene gibt ($\chi^2 = 36.791$, $df = 1$, $p\text{-value} < 0.05$). Der Test für die einzelnen Bücher (siehe Tabelle 8.3) hat das gleiche Ergebnis geliefert. Bei allen Büchern wurde ein signifikanter Unterschied festgestellt. Der Test bei den einzelnen Teilnehmern hat unterschiedliche Ergebnisse geliefert. Während es bei zwei Personen keinen signifikanten Unterschied gab, wurde bei den restlichen ein Unterschied festgestellt (siehe Tabelle 8.4).

Buch	χ^2 -Verteilung	p-Wert
Buch 1	29.614	<0.05
Buch 2	149.19	<0.05
Buch 3	179.07	<0.05
Buch 4	445.61	<0.05
Buch 5	851.15	<0.05

Tabelle 8.3.: Ergebnis des Kruskal-Wallis Test von den einzelnen Büchern

ID	χ^2 -Verteilung	p-Wert
1	41.567	<0.05
2	254.48	<0.05
3	0.83837	0.3599
4	6.3435	<0.05
5	514.91	<0.05
7	415.07	<0.05
8	683.22	<0.05
9	1110.7	<0.05
10	71.061	<0.05
11	3675.6	<0.05
12	287.47	<0.05
13	0.95033	0.3296
14	694.61	<0.05
15	12197	<0.05
16	1588.6	<0.05
18	3399.6	<0.05
19	3361.3	<0.05
20	8401.3	<0.05
21	729.37	<0.05
22	1523	<0.05
23	36.588	<0.05
24	221.53	<0.05
26	1363.1	<0.05
27	3764.7	<0.05
28	3162.1	<0.05

Tabelle 8.4.: Ergebnis des Kruskal-Wallis Test von den einzelnen Studienteilnehmern

8.4.6. Diskussion

Der t-Test für die reale und virtuelle Szene hat gezeigt, dass die Blicktiefe bei den virtuellen Büchern im Vergleich zu den realen durchschnittlich 6cm weiter vorne liegt. Der nicht-parametrische Test hat auch einen signifikanten Unterschied festgestellt. Bei den einzelnen Büchern wurde der gleiche Effekt beobachtet. Außer bei Buch 3 wiesen alle Bücher einen signifikanten Unterschied auf (siehe Tabelle 8.1). Der nicht-parametrische Test konnte für alle Bücher einen Unterschied feststellen (siehe Tabelle 8.3). Hier zeigen sich Unterschiede zwischen den beiden Testverfahren. Die einzelnen Bücher haben Unterschiede bei der Differenz der durchschnittlichen Fehlerdistanz gezeigt. Während die Differenz für Buch 1 und Buch 2 die Werte bei 1.94cm und 5.73cm liegt, werden in Buch 4 und Buch 5 Werte wie 25.6cm und 49cm erreicht. Da die ersten beiden Bücher zwischen 0.9m und 2m und die letzten beiden Bücher zwischen 2.9m und 3.5m liegen, deutet dies darauf hin, dass die Länge des Abstandes Einfluss auf die Wahrnehmung der Blicktiefe haben könnte. Im parametrischen Test wurde für Buch 3 kein signifikantes Ergebnis erreicht. Der nicht-parametrische Test hat jedoch gezeigt, dass ein signifikanter Unterschied besteht. Bei der Analyse der einzelnen Teilnehmer wurden im parametrischen Test bei allen außer bei zwei Personen ein signifikanter Unterschied festgestellt (siehe Tabelle 8.2). Dabei beträgt die durchschnittliche Differenz 18cm. Die Teilnehmer haben in der virtuellen Szene also 18cm weiter nach vorne geschaut als in der realen Szene. Der nicht-parametrische Test hat ähnliche Ergebnisse geliefert (siehe Tabelle 8.4). Auch hier wurde bei allen Teilnehmern außer bei zwei ein Unterschied festgestellt. Dabei sind die nicht-signifikanten Ergebnisse der zwei Teilnehmer nicht die gleichen wie im parametrischen Test. Das bedeutet, dass die beiden Tests Unterschiede aufweisen.

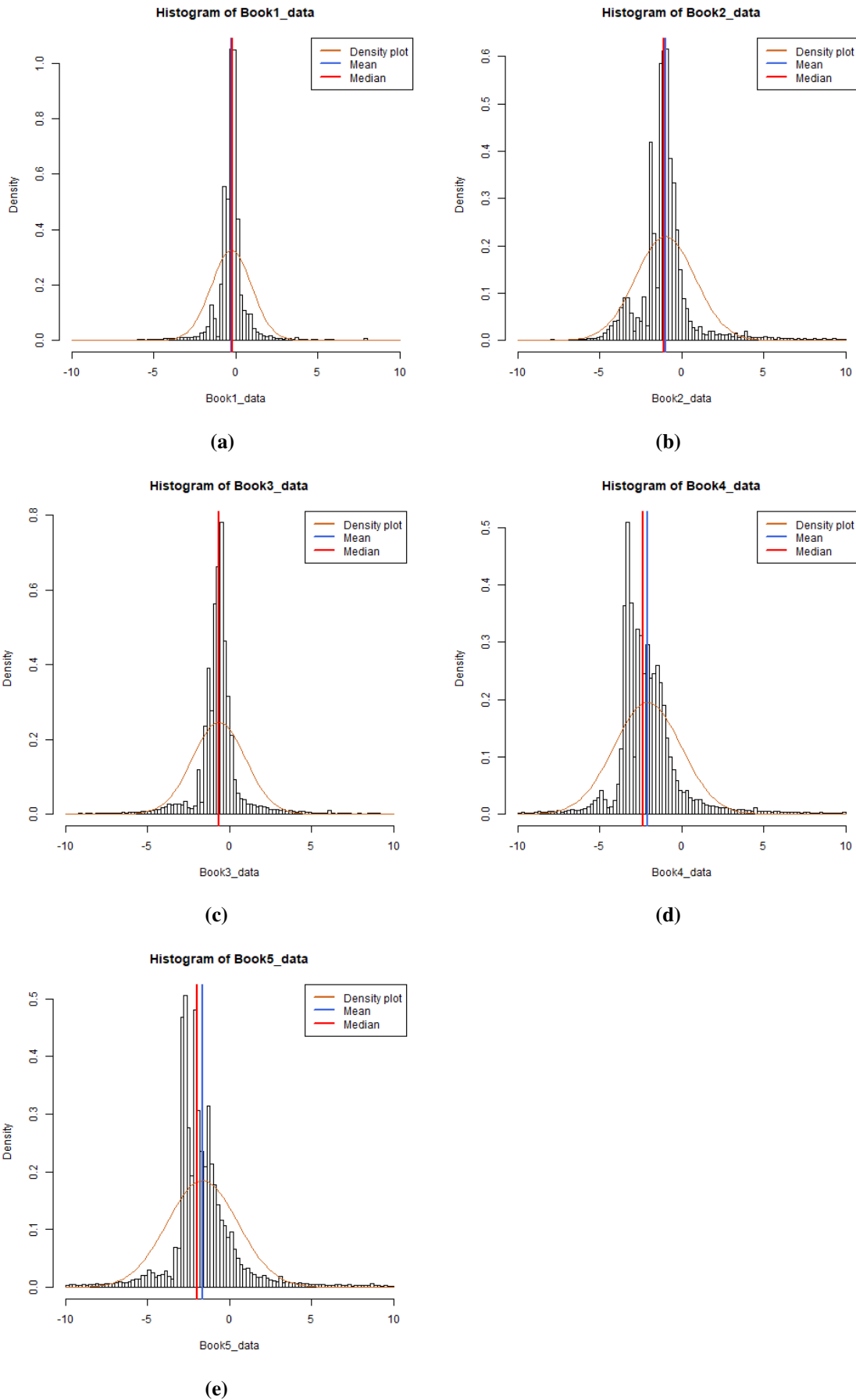


Abbildung 8.10.: Histogramme von den einzelnen Büchern

9. Zusammenfassung und Ausblick

9.1. Zusammenfassung

Im Rahmen dieser Arbeit wurde der Einfluss von AR auf die Wahrnehmung der Blicktiefe untersucht. Dazu wurde eine HoloLens mit einem eingebauten Eye-Tracker, der von Pupil Labs für die HoloLens entwickelt wurde, verwendet. Dabei wurde die Hypothese formuliert, dass ein Unterschied der wahrgenommenen Blicktiefe in AR im Gegensatz zur Realität existiert. Für die Untersuchung der Blicktiefenschätzung sollte in einer Studie ein Vergleich von realer und virtueller Szene erfolgen. Dazu musste zunächst eine Testszene konzipiert werden, um anschließend eine reale und virtuelle Version dieser Szene zu entwickeln.

Die Testszene sollte Zielobjekte beinhalten, die in unterschiedlichen Abständen zum Betrachter liegen. Dabei wurden Bücher als Objekte verwendet. Es wurden fünf Bücher auf einen schwarz bedeckten Tisch positioniert, deren Abstand zwischen 0.9 m und 3.6 m zum Betrachter betrug. Damit der Blick auf dem entsprechenden Buch liegt, mussten Aufgaben auf die Bücher abgebildet werden. Dafür wurden Wörter- und Zahlenrätsel erstellt.

Für die Realisierung der realen Szene mussten lediglich die Aufgaben auf Papier gedruckt und die Bücher damit eingebunden werden. Für die virtuelle Szene mussten die Positionen der virtuellen Bücher so transformiert werden, dass sie auf einem realen Tisch platziert wurden. Dazu wurden AR-Marker in die reale Umgebung untergebracht, damit die HoloLens die Position der Marker ermitteln und die Bücher relativ zu den Markern positionieren konnte.

Zum Schluss wurde in einer Studie ein realer und virtueller Vergleich aufgestellt, indem die Probanden jeweils in beiden Szenen die Zielobjekte betrachten mussten, während der Eye-Tracker die Blicktiefe gemessen hat.

Das Ergebnis der Studie hat gezeigt, dass die Wahrnehmung der Blicktiefe in AR einen Unterschied zur Realität zeigt. Das Resultat war jedoch mit gewissen Einschränkungen verbunden. Da der Eye-Tracker noch Mangel bei der Erfassung der 3D-Gaze-Positionen aufweist, gab es in den Daten viele Ausreißer, die einen möglichen Einfluss auf das Ergebnis haben könnten.

9.2. Ausblick

In Kapitel Studie wurde bereits angemerkt, dass der Eye-Tracker bezüglich der Blicktiefe sehr ungenaue Daten liefert. Die erfassten Daten könnten genauer analysiert werden, um Ausreißer optimal zu entfernen. Dazu können die Daten beispielsweise darauf untersucht werden, ob sie gewisse Muster aufweisen.

Um zu untersuchen, ob das Resultat der Studie durch AR hervorgerufen wurde, könnte in zukünftigen Arbeiten die Studie mit einem qualitativ besseren Eye-Tracker wiederholt werden. Da beobachtet wurde, dass es möglicherweise eine Beziehung zwischen der Wahrnehmung der Blicktiefe in AR und den Abständen gibt, könnten Anpassungen in der Testszene vorgenommen werden, indem auch sehr nahe (< 0.9 m) oder sehr weit entfernte Zielobjekte (> 3.6 m) in die reale Welt positioniert werden. Um zu untersuchen, ob die Objektwahl das Ergebnis beeinflusst hat, könnten nach anderen alternativen Objekten gesucht und verwendet werden.

Da die HoloLens 2 einen integrierten Eye-Tracker besitzt und sehr bald erworben werden kann, kann die Studie damit wiederholt werden. Die Ergebnisse können anschließend mit den Ergebnissen dieser Arbeit verglichen werden.

Literaturverzeichnis

- [ABB+01] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, B. MacIntyre. „Recent advances in augmented reality“. In: Hrsg. von IEEE. 2001 (zitiert auf S. 13).
- [AHB87] K. S. Arun, T. S. Huang, S. D. Blostein. „Least-squares fitting of two 3-D point sets“. In: *IEEE Transactions on pattern analysis and machine intelligence* 5 (1987), S. 698–700 (zitiert auf S. 43).
- [Azu97] R. T. Azuma. „A survey of augmented reality“. In: *Presence: Teleoperators & Virtual Environments* 6.4 (1997), S. 355–385 (zitiert auf S. 13).
- [Cora] M. Corporation. *HoloLens (1st Gen) hardware details*. URL: <https://docs.microsoft.com/de-de/windows/mixed-reality/hololens-hardware-details> (zitiert auf S. 35, 36).
- [Corb] M. Corporation. *Microsoft HoloLens*. URL: <https://www.bechtle.com/shop/medias/5a1430f79ce96955066d0b10.pdf?context=bWFzdGVyFHJvb3R8NjA0ODN8YXBwbGljYXRpb24vcGRmfGhLNC9oM2UvOTQ3MDY4ODYyNDY3MC5wZGZ8MmUwZjQ0YWMYmUxMTk3MGE0MmJkNWQxNjI4NmU1MzFiN2Q2MDU4MDYxYjM0MmJiYjkzNzQ5NmY0Njc0YjJhZA> (zitiert auf S. 35).
- [Dah06] M. Dahm. *Grundlagen der Mensch-computer-interaktion*. Pearson Studium München, 2006 (zitiert auf S. 14, 16–18).
- [DM96] D. Drascic, P. Milgram. „Perceptual issues in augmented reality“. In: *Stereoscopic displays and virtual reality systems III*. Bd. 2653. International Society for Optics und Photonics. 1996, S. 123–134 (zitiert auf S. 19, 20, 22).
- [DPHW11] A. T. Duchowski, B. Pelfrey, D. H. House, R. Wang. „Measuring gaze depth with an eye tracker during stereoscopic display“. In: *Proceedings of the ACM SIGGRAPH symposium on applied perception in graphics and visualization*. ACM. 2011, S. 15–22 (zitiert auf S. 27).
- [Duc07] A. T. Duchowski. „Eye tracking methodology“. In: *Theory and practice* 328.614 (2007), S. 2–3 (zitiert auf S. 22, 23).
- [DWSS17] C. Diaz, M. Walker, D. A. Szafrir, D. Szafrir. „Designing for depth perceptions in augmented reality“. In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2017, S. 111–122 (zitiert auf S. 19).
- [EMP+17] G. Evans, J. Miller, M. I. Pena, A. MacAllister, E. Winer. „Evaluating the Microsoft HoloLens through an augmented reality assembly application“. In: *Degraded Environments: Sensing, Processing, and Display 2017*. Bd. 10197. International Society for Optics und Photonics. 2017, S. 101970V (zitiert auf S. 35).
- [Fol89] J. M. Foley. „Stereoscopic distance perception“. In: (1989) (zitiert auf S. 26).
- [GB09] Goldstein, E. Bruce. *Sensation and perception*. Cengage Learning, 2009 (zitiert auf S. 15–18).

- [JK03a] R. J. Jacob, K. S. Karn. „Eye tracking in human-computer interaction and usability research: Ready to deliver the promises“. In: *The mind's eye*. Elsevier, 2003, S. 573–605 (zitiert auf S. 22, 23).
- [JK03b] R. J. Jacob, K. S. Karn. „Eye tracking in human-computer interaction and usability research: Ready to deliver the promises“. In: *The mind's eye*. Elsevier, 2003, S. 573–605 (zitiert auf S. 23).
- [KPB14] M. Kassner, W. Patera, A. Bulling. „Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction“. In: *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. ACM. 2014, S. 1151–1160 (zitiert auf S. 38).
- [Kra16] G. Kramida. „Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays“. In: *IEEE Transactions on Visualization and Computer Graphics* 22.7 (Juli 2016), S. 1912–1931. DOI: [10.1109/TVCG.2015.2473855](https://doi.org/10.1109/TVCG.2015.2473855) (zitiert auf S. 20).
- [KSF10] E. Kruijff, J. E. Swan, S. Feiner. „Perceptual issues in augmented reality revisited“. In: *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE. 2010, S. 3–12 (zitiert auf S. 20–22).
- [Laba] P. Labs. *Developer Documentation*. URL: <https://github.com/pupil-labs/hmd-eyes/blob/034a3c613cee7cd725c49189944229d7bb7120fe/docs/Developer.md> (zitiert auf S. 42).
- [Labb] P. Labs. *hololensRelay*. URL: https://github.com/pupil-labs/pupil/blob/master/pupil_src/shared_modules/hololens_relay.py (zitiert auf S. 32).
- [LMF+09] Lambooi, Marc, Fortuin, Marten, Heynderickx, Ingrid, IJsselsteijn, Wijnand. „Visual discomfort and visual fatigue of stereoscopic displays: A review“. In: *Journal of Imaging Science and Technology* 53.3 (2009), S. 30201–1 (zitiert auf S. 16, 18, 19).
- [MPWL13] M. Mehrabi, E. M. Peek, B. C. Wuensche, C. Lutteroth. „Making 3D work: a classification of visual depth cues, 3D display technologies and their applications“. In: *Proceedings of the Fourteenth Australasian User Interface Conference-Volume 139*. Australian Computer Society, Inc. 2013, S. 91–100 (zitiert auf S. 20).
- [MTK95] P. Milgram, H. Takemura, A. U. and Fumio Kishino. *Augmented reality: a class of displays on the reality-virtuality continuum*. 1995. DOI: [10.1117/12.197321](https://doi.org/10.1117/12.197321). URL: <https://doi.org/10.1117/12.197321> (zitiert auf S. 13, 14).
- [Qia97] N. Qian. „Binocular disparity and the perception of depth“. In: *Neuron* 18.3 (1997), S. 359–368 (zitiert auf S. 18).
- [RGA95] J. P. Rolland, W. Gibson, D. Ariely. „Towards quantifying depth and size perception in virtual environments“. In: *Presence: Teleoperators & Virtual Environments* 4.1 (1995), S. 24–49 (zitiert auf S. 25, 26).
- [RHFL10] S. Reichelt, R. Häußler, G. Fütterer, N. Leister. *Depth cues in human visual perception and their realization in 3D displays*. 2010. DOI: [10.1117/12.850094](https://doi.org/10.1117/12.850094). URL: <https://doi.org/10.1117/12.850094> (zitiert auf S. 16).
- [RMAR02] J. P. Rolland, C. Meyer, K. Arthur, E. Rinalducci. „Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays“. In: *Presence: Teleoperators & Virtual Environments* 11.6 (2002), S. 610–625 (zitiert auf S. 25).

- [SJK+07] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, H. S. Smallman. „Egocentric depth judgments in optical, see-through augmented reality“. In: *IEEE transactions on visualization and computer graphics* 13.3 (2007), S. 429–442 (zitiert auf S. 26).
- [SSE15] J. E. Swan, G. Singh, S. R. Ellis. „Matching and reaching depth judgments with real and augmented reality targets“. In: *IEEE transactions on visualization and computer graphics* 21.11 (2015), S. 1289–1298 (zitiert auf S. 25, 26).
- [SSTT12] R. Snowden, R. J. Snowden, P. Thompson, T. Troscianko. *Basic vision: an introduction to visual perception*. Oxford University Press, 2012 (zitiert auf S. 14, 15, 17, 18).
- [Sut68] I. E. Sutherland. „A Head-mounted Three Dimensional Display“. In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I. AFIPS '68 (Fall, part I)*. San Francisco, California: ACM, 1968, S. 757–764. DOI: [10.1145/1476589.1476686](https://doi.org/10.1145/1476589.1476686). URL: <http://doi.acm.org/10.1145/1476589.1476686> (zitiert auf S. 13).
- [Tay16] A. G. Taylor. *Develop Microsoft hololens apps now*. Springer, 2016 (zitiert auf S. 35, 36).
- [TJ96] Tovée, M. J. *An introduction to the visual system*. Cambridge University Press, 1996 (zitiert auf S. 16).
- [Wik19] t. f. e. Wikipedia. *Snellen chart*. 2019. URL: https://en.wikipedia.org/wiki/Snellen_chart (zitiert auf S. 76).
- [WJX+16] Z. Wang, X. Jin, F. Xue, R. Li, H. Zha, K. Ikeuchi. „Perceptual enhancement for stereoscopic videos based on horopter consistency“. In: *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM. 2016, S. 15–18 (zitiert auf S. 19).

Alle URLs wurden zuletzt am 17.09.2019 geprüft.

A. Anhang

A.1. Testszene

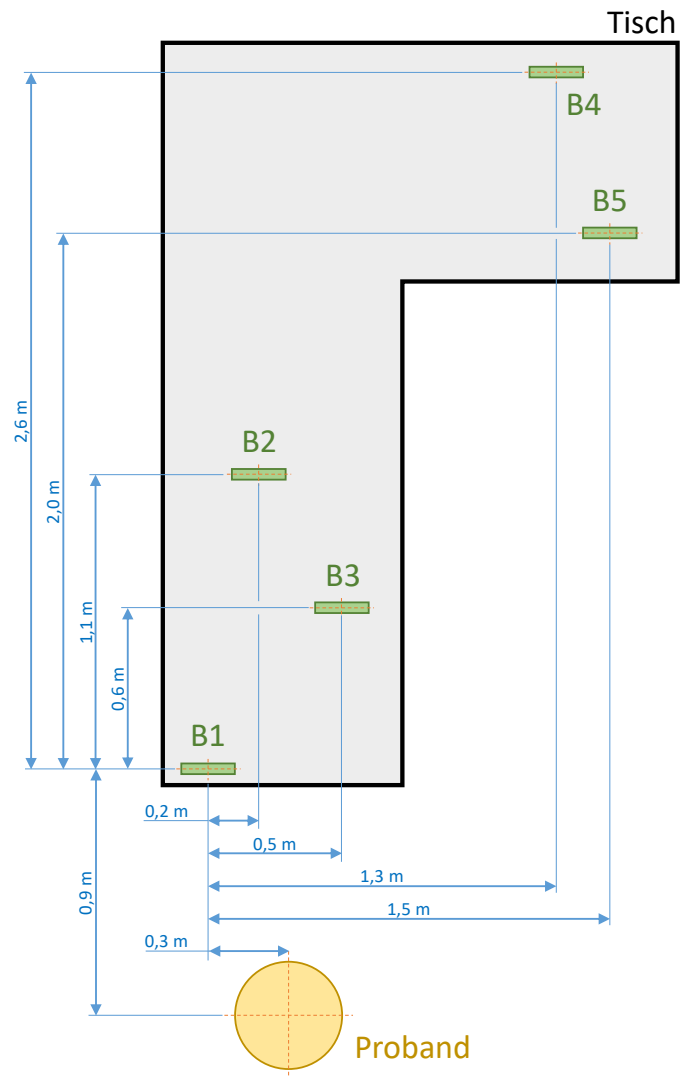


Abbildung A.1.: Skizze der Testszene in Hochformat

A.2. Einverständniserklärung



Universität Stuttgart

Universität Stuttgart • VISUS • Allmandring 19 • 70569 Stuttgart

Information und Einverständniserklärung zur Teilnahme an der Forschungsstudie „Vergleich der virtuellen und realen Umgebung mittels Eye-Tracking“

Liebe(r) Studieninteressent/in,

wir laden Sie ein, an folgender Studie im Rahmen unserer Forschung des Sonderforschungsbereichs/ Transregio 161 teilzunehmen. Wir untersuchen den Einfluss von Augmented Reality zum Lösen von einfachen Rätselaufgaben. In unserer aktuellen Studie, „**Vergleich der virtuellen und realen Umgebung mittels Eye-Tracking**“, werden Sie Aufgaben in einer realen und virtuellen Szene bearbeiten. Die erhobenen Daten dienen dazu herauszufinden, welchen Einfluss Augmented Reality hat.

Studienablauf

Der Ablauf der Studie sieht wie folgt aus:

1. Sie führen eine Kalibrierung für die Hololens und den Eyetracker durch
2. Die Szene wird geladen und Sie werden dazu aufgefordert die Marker zu betrachten.
3. Nachdem die Marker erfolgreich erkannt wurden, werden Sie in der virtuellen und realen Szene jeweils die Fragen auf den Büchern beantworten.
4. Nachdem Sie alle Aufgaben erfolgreich gelöst haben, werden Sie einen Fragebogen ausfüllen.

Visualisierungsinstitut (VISUS)

Direktoren

Prof. Dr. Thomas Ertl
Prof. Dr. Daniel Weiskopf

Ansprechpartner

Nils Rodrigues
Michael Becher
Jun.-Prof. Michael Sedlmair

Kontakt

Allmandring 19
70569 Stuttgart
T 0711 685-88634
nils.rodrigues@visus.uni-stuttgart.de

www.visus.uni-stuttgart.de
www.twitter.com/vis_visus

15.09.2019



Bank
Baden-Württembergische Bank
Stuttgart – BW-Bank

IBAN
DE51 6005 0101 7871 5216 87

SWIFT/BIC
SOLADEST600

Umsatzsteuer-IdNr.
DE147794196

Seite 1



Weiterverarbeitung der aufgenommenen Daten

Visualisierungsinstitut

Die Daten, die während der Studie aufgezeichnet werden, werden nur in anonymer Form gespeichert und können nicht auf eine bestimmte Person zurückgeführt werden. Die Daten werden zusammen mit den Ergebnissen anderer Versuchspersonen verarbeitet um gemittelte Werte zu erhalten, die dann statistisch ausgewertet und in wissenschaftlichen Zeitschriften oder Konferenzbeiträgen veröffentlicht werden. Sie stimmen der Nutzung und der Veröffentlichung Ihrer Daten für wissenschaftliche Zwecke zu. Diese Zustimmung kann jederzeit durch ein einfaches Schreiben widerrufen werden. Dabei müssen keine Gründe, die zu dieser Entscheidung beigetragen haben, genannt werden.

Datenschutz

Alle Angaben, die Rückschlüsse auf Ihre Person zulassen, werden von den MitarbeiterInnen des Projektes vollständig vertraulich behandelt, d.h. nicht an Dritte außerhalb des Forschungsprojektes weitergegeben oder veröffentlicht. Sollten Sie die Informationsweitergabe an Dritte wünschen, so geschieht dies ausschließlich auf Ihre schriftliche Veranlassung hin. Nur autorisierte MitarbeiterInnen haben Zugang zu dem erhobenen Datenmaterial. Die Computer auf denen solche personenbezogenen Daten gespeichert sind, sind mit dem Internet verbunden und speziell gegen unautorisierten Zugriff gesichert. Ein legaler Zugriff auf die Daten von außerhalb ist nicht möglich.

Kosten und Aufwandsentschädigung

Für die Teilnahme an der Studie entstehen Ihnen keine Kosten. Für Ihre Teilnahme erhalten sie eine pauschale Aufwandsentschädigung von 10 € (EUR) aus Mitteln des Sonderforschungsbereich/ Transregio 161.

Risiken und Rechte

Die Teilnahme an der Studie ist mit gewissen Risiken verbunden. Eine erhöhte Infrarotbestrahlung (Wärmestrahlung), die durch den Pupil Labs Eye-Tracker hervorgerufen wird, kann schädlich für die Augen sein. Es ist möglich, dass Sie während der Studie trockene Augen kriegen und deshalb häufiger Blinzeln müssen. Ihre Teilnahme an der Studie ist freiwillig. Durch Ihre Einwilligung gehen Sie keinerlei Verpflichtungen ein. Sie können Ihre Einwilligung jederzeit widerrufen, ohne dass Ihnen ein rechtlicher Nachteil entsteht. Sie können die Studie jederzeit und ohne Angaben von Gründen abbrechen, ohne dass Ihnen ein rechtlicher Nachteil entsteht.

Einverständniserklärung

- Ich habe die vorausgehende Erklärung gelesen bzw. vorgelesen und übersetzt bekommen und verstanden.
- Ich wurde von dem Versuchsleiter über die Studie aufgeklärt und meine Fragen wurden für mich befriedigend und umfassend beantwortet.
- Ich weiß, dass Daten über meine Person nur anonym verarbeitet werden, und dass alle autorisierten ProjektmitarbeiterInnen, die Zugang zu Angaben und Daten zu meiner Person haben, unter Schweigepflicht stehen.
- Mir ist bekannt, dass die erhobenen Informationen auf Computern gespeichert und verarbeitet werden, die mit dem Internet verbunden sind.
- Ich nehme freiwillig an der Studie teil und weiß, dass ich die Teilnahme jederzeit abbrechen kann.

Nachname, Vorname

Ort, Datum und Unterschrift des Teilnehmers

Datum und Unterschrift des Versuchsleiters

Vielen Dank für Ihre Teilnahme!



Abbildung A.3.: Zweiter Teil der Einverständniserklärung

A.3. Fragebogen

**ID «ID»
Fragebogen 1. Teil
Demographische Angaben**

1. **Geschlecht:** Männlich
 Weiblich
 Andere

2. **Alter:** _____

3. **Welches ist Ihr höchster Bildungsabschluss?**
 Hochschulreife
 Bachelor
 Master
 Diplom
 Sonstige _____

Welches Studienfach habe Sie studiert/ studieren Sie?

Welchen Studienabschluss streben Sie an?
 Bachelor
 Master
 Staatsexamen
 Sonstige _____

4. **Tragen Sie eine Sehhilfe?**
 Nein
 Brille
 Kontaktlinsen

5. **Welche Augenfarbe haben Sie?** _____ hell dunkel

6. **Haben Sie bereits Erfahrung in Augmented Reality (AR)/Mixed Reality (MR)/Virtual Reality (VR)?**
 Ja
 Nein

7. Mit welchen AR/MR/VR – Geräten haben Sie bereits Erfahrung?

- Google Glass
- Hololens
- Meta 2
- Magic Leap
- Windows VR-Brille (Acer, Asus, Dell Visor, HP, Lenovo Explorer ...)
- HTC Vive
- Oculus Rift
- Sonstiges: _____

8. Benutzen Sie ein AR/MR/VR Gerät regelmäßig? Wenn ja, wie viele Stunden in der Woche?

- Ja, _____ Stunden in der Woche
- Nein

Sehtest (wird vom Versuchsleiter ausgefüllt)

Damit wir in der Studie zuverlässige Ergebnisse bekommen können, müssen wir einen Sehtest durchführen. Der Test ist nur spezifisch auf diese Studie ausgelegt und ist **nicht** für eine medizinisch relevante Einschätzung geeignet.

Snellen-Test

Zeilen, die nicht gelesen werden konnten:

- 1 20/200
- 2 20/100
- 3 20/70
- 4 20/50
- 5 20/40
- 6 20/30
- 7 20/25
- 8 20/20
- 9
- 10
- 11

Abbildung A.5.: Zweiter Teil der Einverständniserklärung

**ID «ID»
Fragebogen 2. Teil**

9. Kreuzen Sie an, welche Aussagen für Sie zutreffen.

- Ich konnte alle virtuellen Bücher gut erkennen und lesen
- Ich konnte **nicht** alle virtuellen Bücher gut erkennen und lesen
- Buch 1
- Buch 2
- Buch 3
- Buch 4
- Buch 5

10. Bitte kreuzen Sie an, welche Aussagen für Sie zutreffen.

- Ich konnte alle realen Bücher gut erkennen und lesen
- Ich konnte **nicht** alle realen Bücher gut erkennen und lesen
- Buch 1
- Buch 2
- Buch 3
- Buch 4
- Buch 5

11. Hatten Sie Schwierigkeiten, den Blick auf die Bücher zu fokussieren?

- Ja
- Nein

12. Konnten Sie einen Unterschied zwischen den echten und virtuellen Büchern feststellen?

- Nein
- Ja

Wenn ja, welchen? _____

13. Bitte geben Sie an, inwieweit Sie mit der folgenden Aussage zustimmen.

	(1) Stimme gar nicht zu	(2) Stimme eher nicht zu	(3) Weder noch	(4) Stimme eher zu	(5) Stimme voll zu
Die virtuellen und realen Bücher sehen ähnlich aus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. Haben Sie sich während der Studie unwohl gefühlt? (Augenbrennen, Kopfschmerzen, Übelkeit, ...)

- Ja, _____
- Nein

15. Weitere Anmerkungen:

Abbildung A.6.: Zweiter Teil der Einverständniserklärung

A.4. Sehtest

E	1	20/200
F P	2	20/100
T O Z	3	20/70
L P E D	4	20/50
P E C F D	5	20/40
E D F C Z P	6	20/30
F E L O P Z D	7	20/25
D E F P O T E C	8	20/20
L E F O D P C T	9	
F D P L T C E O	10	
P E Z O L C F T D	11	

Abbildung A.7.: Snellentest [Wik19]

A.5. Histogramme der Daten

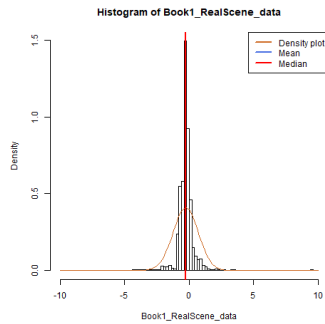


Abbildung A.8.: Histogramm von Buch1 der realen Szene.

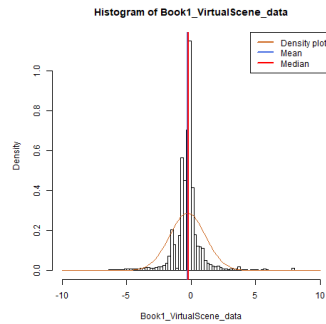


Abbildung A.9.: Histogramm von Buch1 der virtuellen Szene.

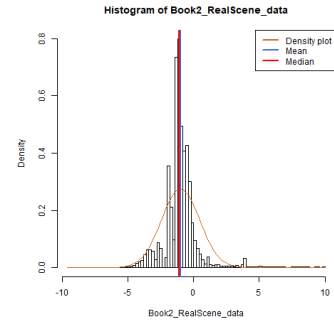


Abbildung A.10.: Histogramm von Buch2 der realen Szene.

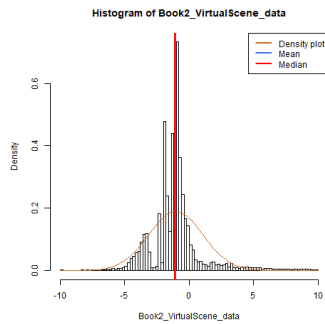


Abbildung A.11.: Histogramm von Buch2 der virtuellen Szene.

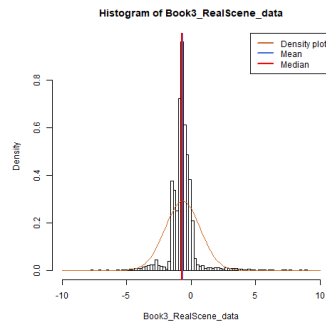


Abbildung A.12.: Histogramm von Buch3 der realen Szene.

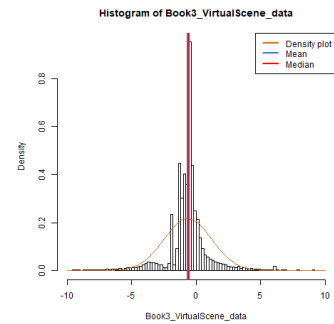


Abbildung A.13.: Histogramm von Buch3 der virtuellen Szene.

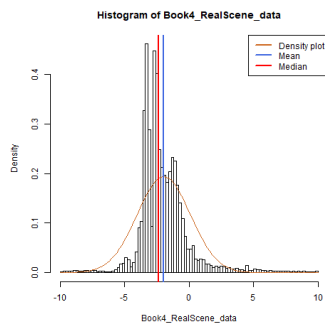


Abbildung A.14.: Histogramm von Buch4 der realen Szene.

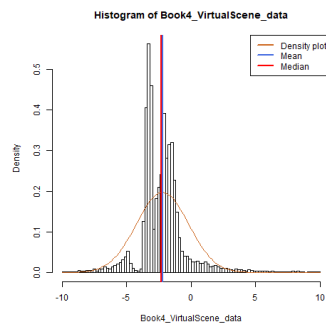


Abbildung A.15.: Histogramm von Buch4 der virtuellen Szene.

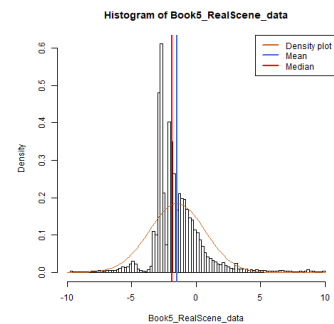
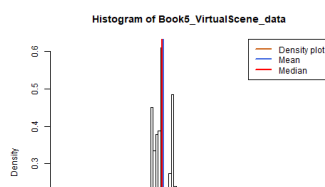


Abbildung A.16.: Histogramm von Buch5 der realen Szene.



Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Ort, Datum, Unterschrift