

Entwicklung eines Natural User Interface mit Hilfe moderner AR und AI Technologien

MASTERARBEIT

Angewandte Informatik

an der

Fachhochschule Erfurt

von

Sebastian Rieger

Abgabedatum 15.11.2017

Bearbeitungszeitraum	24 Wochen
Matrikelnummer	10286908
Betreuer der Masterarbeit	Prof. Rolf Kruse
Zweitbetreuer der Masterarbeit	Prof. Steffen Avemarg

Erklärung

Ich, Sebastian Rieger, versichere hiermit, dass ich die vorliegende Masterarbeit mit dem Thema „Entwicklung eines Natural User Interface mit Hilfe moderner AR und AI Technologien“ selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Ort Datum

Unterschrift

Inhaltsverzeichnis

1	Abstract	4
2	Einleitung	5
3	Augmented Reality	6
3.1	Typen von visuelle Ausgabegeräte	6
3.1.1	Handheld-Geräte	7
3.1.2	Video-See-Through-Displays	8
3.2	Auswertung der Möglichkeiten für ein Natural User Interface	10
4	Artificial Intelligence	12
5	Beschreibung eines Natural User Interface	12
6	Versuch der Entwicklung eines Natural User Interface anhand von ... und ...	12
7	Fazit	12
8	zusammenfassung	12
9	Abkürzungsverzeichnis	13

1 Abstract

2 Einleitung

In den letzten zwanzig Jahren, hat sich die Bedienung von Computern grundlegend geändert. Vor nicht all zu vielen Jahren, gab es nur die Möglichkeit mit Hilfe von Maus und Tastatur mit einem Computer zu interagieren.

Mit dem Aufkommen von Touch-Screens jedoch, änderte sich auch die Benutzung von Computern. Es war nun möglich direkt mit dem Bildschirm zu interagieren ohne den Umweg über die Maus.

Als dann wenig später die ersten Sprachsteuerungen auf den Markt kamen, änderte sich die Interaktion mit dem Computer erneut. So ist es nun möglich Computern mittels Sprache Befehle zu erteilen oder Texte zu sprechen, welche automatisch transkribiert werden.

Heute ist es mit manchen Smartphones schon möglich Nachrichten wie SMS zu schreiben und zu versenden, ohne das Telefon überhaupt in die Hand zu nehmen. Dies ist nur durch neuste Entwicklungen in der *künstliche Intelligenz* (KI) möglich.

Im selben Zeitraum hat sich parallel auch die *Virtuelle Realität* (VR) entwickelt. Sie versucht Menschen mit Hilfe von verschiedenen Brillen in eine virtuelle Welt zu versetzen. Die Entwicklungen solcher VR Systeme wurde vor allem von der Spieleindustrie getrieben, da sie immer neue Versuche unternimmt Spieler besser in die Welt des Spiels zu versetzen.

Eine Abstufung der VR ist die *Argumented Reality* (AR), welche versucht die analoge, reale Welt um digitale Inhalte zu erweitern. Hierbei sind die Möglichkeiten für den Einsatz von AR fast unbegrenzt. Es ist also quasi möglich, jede menschliche Tätigkeit durch die AR zu unterstützen.

Genau hier setzt der Schwerpunkt der Arbeit an. Im Verlauf soll versucht werden ein *Natural User Interface* (NUI) unter Zuhilfenahme moderner AR und KI Technologien zu erstellen.

Hierfür werden aktuelle AR und KI Systeme daraufhin untersucht, wie sie im Zusammenspiel ein NUI bilden können, welches allein durch Sprache und Gesten mit einem Computer interagiert.

Es soll ein übliche Tätigkeit mit Hilfe einer AR Technologie in der realen Welt abgebildet werden, welche dann von einem Menschen nur unter Zuhilfenahme von Sprache und Gesten gesteuert und bearbeitet werden kann.

3 Augmented Reality

In der heutigen Zeit ist fast jedem der Begriff Augmented Reality geläufig. Jedoch gibt es im wissenschaftlichen Umfeld keine einheitliche Definition. Georg Klein definiert die AR als „Anreicherung der realen Welt um computergenerierte Zusatzobjekte“. [?]]

Viele Abhandlungen zu diesem Thema beziehen sich auf das „Reality-Virtuality Continuum“, welches von Milgram, Takemura, Utsumi und Kishino welches in Abbildung 1 zu sehen ist. Dieses stellt bildlich dar, wie sich eine AR-Anwendung einordnen lässt. Links ist die reale Umgebung (Real Environment) und rechts die Virtuelle Umgebung (Virtual Environment) zu sehen.

Heutige VR-Anwendungen müssen also ganz rechts eingeordnet werden. Zwischen der realen und der virtuellen Umgebung gibt es jedoch noch weitere Abstufungen. So gibt es die „Augmented Reality“, welche die reale Welt um virtuelle Elemente ergänzt. Sie bezieht sich eher auch die reale Umgebung, weshalb sie weitere links angeordnet ist. Die „Augmented Virtuality“ ist das genaue Gegenteil der AR. Hier wird die virtuelle Welt des Computers um reale Gegenstände ergänzt.

Im wissenschaftlichen Umfeld wird als Oberbegriff für „Augmented Reality“ und „Augmented Virtuality“ oft „Mixed Reality“ oder „Enhanced Reality“ verwendet. [?]]

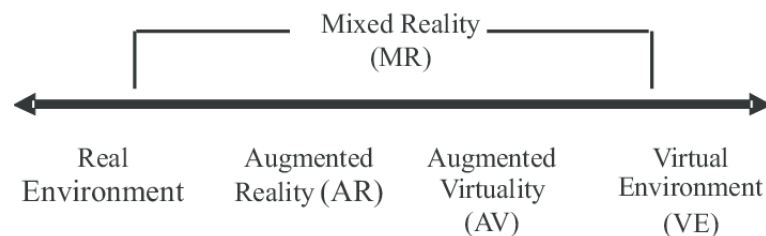


Abbildung 1: Reality-Virtuality Continuum [?]]

Der Begriff der *Augmented Virtuality* (AV) ist heute weniger gebräuchlich, da es keine wirklichen Anwendungsfelder gibt, in denen eine virtuelle Welt um reale Gegenstände ergänzt werden muss. Im Umkehrschluss wird heute jedoch die AR immer bedeutender, da bei fast jeder menschlichen Tätigkeit eine Unterstützung durch Computergenerierte Objekte oder Informationen möglich ist.

Im medizinischen Umfeld, können Ärzten zum Beispiel Positionen von empfindlichen Gewebe oder zu entfernenden Fremdkörpern bei einer Operation angezeigt werden. Eine andere Einsatzmöglichkeit wäre das unterstützende Anzeigen von Gefahren oder Fluchtwegen für Feuerwehrleute, die sich einem stark verrauchten Brandobjekt befinden. Computer können durch Infrarot und Wärmebild das Sichtvermögen eines Feuerwehrmannes in einer Gefahrensituation so entscheidend erhöhen um sich selbst und andere zu retten.

Ein Gerät, welches später für die Erstellung eines NUI dienen kann muss folgende Kriterien erfüllen:

- Das Gerät muss Freihändig bedienbar sein
- Das Gerät muss die AR-Inhalte exakt zur realen Welt wiedergeben
- Das Gerät muss Handgesten und Sprache des Nutzers erkennen können

3.1 Typen von visuelle Ausgabegeräte

Im folgenden Abschnitt werden mögliche visuelle Ausgabegerätetypen genauer beschrieben und auf ihre Einsatzmöglichkeiten untersucht.

3.1.1 Handheld-Geräte

Als Handheld werden in Geräte bezeichnet, die wie der Name schon sagt von einer Person in der Hand gehalten werden. Im AR-Umfeld sind hier Smartphones oder Tablets gemeint, welche heutzutage durch ihre Dual- oder Quad-Core-Prozessoren und teilweise mehrere Gigabyte RAM in der Lage sind, auch komplexere AR-Anwendungen zur Ausführung zu bringen.

In Smartphones und Tablets nimmt die Kamera das Bild der realen Welt auf und stellt es in Echtzeit auf dem Display dar. Ein Programm (App) ist dann in der Lage, die aufgenommene reale Welt um virtuelle Elemente zu erweitern und diese innerhalb des Kamerabildes auf dem Bildschirm zu platzieren.

Hierbei ergibt sich jedoch ein Problem, da sich der Blickwinkel des Betrachters vom Blickwinkel der Kamera unterscheidet, was zu einer unsauberen Platzierung der Objekte im Bild führt. In Abbildung 2 ist zu sehen, wie sich der Blickwinkel der Kamera und der Blickwinkel des Betrachters unterscheiden.

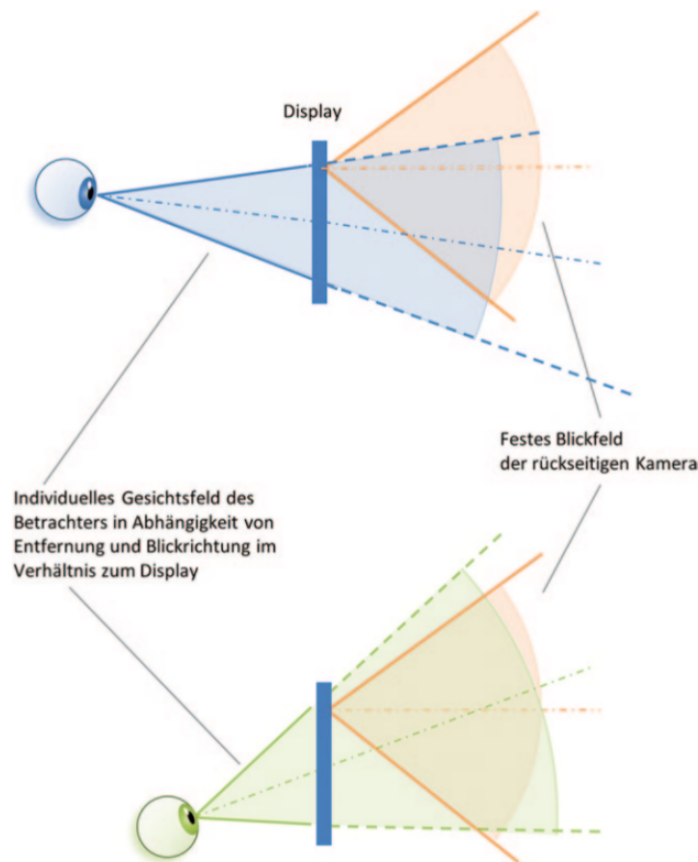


Abbildung 2: Unterschiedliche Blickfelder von Betrachter und Kamera bei Handheld-AR [?]

Um den unterschied, zwischen den Blickwinkeln auszugleichen, muss das entsprechende Gerät die beiden Blickwinkel aufeinander abstimmen. Dies kann jedoch kein zur Zeit auf dem Markt befindliches Gerät leisten.

Einen ersten Ansatz hierzu hat die University of California gemacht, indem sie ein Tablet mit einem Kinect Sensor verbunden haben. Mit Hilfe des zusätzlichen Sensors und dem „KinectFusion Algorithmus“ konnten sie eine „Magic Lens“ entwickeln, die beide Blickwinkel in etwa aufeinander abstimmt. [?]

In Abbildung 3 ist einmal der Unterschied zwischen einer Aufnahme mit Magic Lens und ohne schematisch dargestellt.

Im linken Bildabschnitt ist gut zu sehen, dass Bäume und andere Gegenstände im Bildschirm exakt an der Position in der realen Welt liegen. Im Gegensatz dazu sind Bäume und andere Gegenstände im rechten Bildabschnitt merklich verzerrt.



Abbildung 3: Schematische Darstellung des Magic Lens Effekts [?]

Ein weiteres Problem ist, dass AR-Anwendungen für die optimale Skalierung und Platzierung von virtuellen Objekten im Raum zusätzlich Tiefeninformationen des aktuellen Bildes benötigen, denn nur dann ist auch gewährleistet, dass virtuelle Objekte in der richtigen Größe im realen Raum positioniert werden.

Dieses Problem versucht gerade das Projekt Tango von Google zu lösen. Mit Hilfe eines Lasers und einer Infrarotkamera können Tiefeninformationen zum aktuellen Bild aufgenommen und verarbeitet werden. [?]

Nur mit der Tango Technologie ist es bisher möglich, AR-Objekte richtig skaliert und im richtigen Blickwinkel präzise im Raum zu positionieren. Weitergehende Informationen sind auf den Developer-Seiten¹ von Google zu finden. Bisher unterstützen jedoch nur das „Lenovo Phab2“ und das „Asus ZenFon AR“ diese Technologie, da hier eine zusätzliche Sensoreinheit im Smartphone verbaut sein muss.

Abschließen kann über AR-Anwendungen ausgesagt werden, dass sie Grundsätzlich lauffähig auf Handheld Geräten wie Tablets oder Smartphones sind. Die Bedingungen sind jedoch nicht zumindest heutzutage noch optimal, wie eben dargestellt wurde.

3.1.2 Video-See-Through-Displays

Video-See-Through Geräte existieren im Vergleich zu AR-Anwendungen schon ziemlich lange. Schon Anfang der 1940er Jahre wurde erste Head-Up-Displays in Kampfflugzeugen eingesetzt, um die Piloten mit zusätzlichen Informationen zu versorgen, welche immer im Blick behalten werden müssen. [?]

Von Anfang an, wurden die zusätzlichen Informationen auf eine Glasfläche projiziert, um das Sichtfeld nicht unnötig einzuschränken. Es wurde entweder direkt auf das Cockpit Fenster projiziert oder die Piloten hatten kleine durchsichtige Displays in den Helm eingearbeitet, auf die mit Hilfe von Spiegeln projiziert werden konnte.

Mit der Veröffentlichung des iPhones im Jahre 2007 wurde das Thema AR immer interessanter, da mit ihm die technische Grundlage gelegt wurde. 2008 erschien der AR-Browser Wikitude wodurch das Feld immer mehr Fuß fasste und sich nun rasant weiter entwickelt. [?]

Die durch die immer größerer Verbreitung von AR-Anwendungen wurde auch die Hardware entsprechend weiterentwickelt und angepasst, was zu den ersten AR-Brillen führte.

Durch die Kombination von Handheld-Geräten und Head-Up-Displays entstanden bis 2014 die ersten Video-See-Through-Brillen oder auch AR-Brillen genannt. Die Entwicklung dieser Brillen

¹<https://get.google.com/tango/>

steckt noch in den Kinderschuhen, aber dennoch gibt es schon zwei auf dem Markt verfügbare Geräte, die für die spätere Entwicklung eines NUI in Betracht gezogen werden können. Zum einen ist das die Epson Moverio BT-200 und zum anderen die Microsoft HoloLens, welche zur Zeit nur für Entwickler verfügbar ist.

Die HoloLens und die Moverio sind beide Prismen basiert. In Abbildung 4 ist die Funktionsweise von Prismen basierten See-Through-Displays schematisch dargestellt. Ein normales Display ist an einer Seite eines Prismas angebracht. Das Prisma leitet das Licht des Displays weiter, so dass es in das Auge des Betrachters weitergeleitet wird. Weil das Prisma nach vorn durchsichtig ist, wird auch das Umgebungslicht direkt in das Auge des Betrachters geleitet. Innerhalb des Prismas vermischen sich die beiden Bilder, wobei das Displaylicht das Umgebungslicht überlagert. Für den Betrachter sieht es somit aus, als würden die Objekte in der realen Welt zu finden sein.

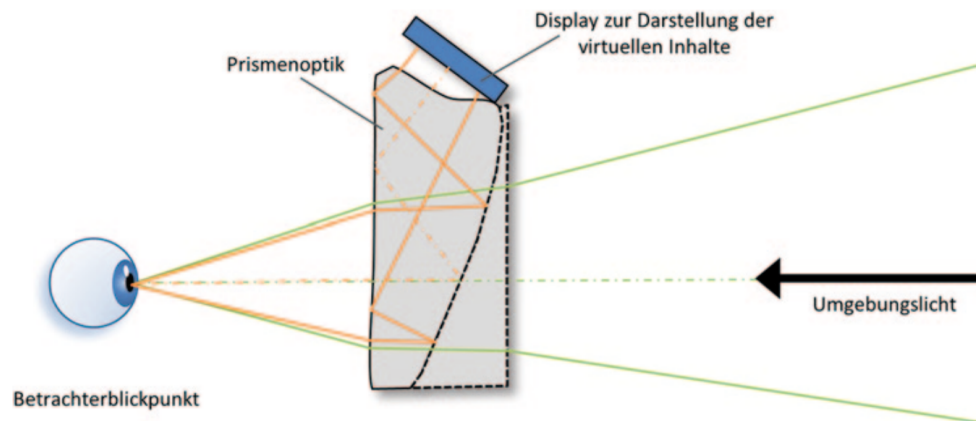


Abbildung 4: Schematische Darstellung des Aufbaus eines See-Through-Displays [?]

Die Microsoft HoloLens wurde 2015 von Microsoft vorgestellt und ist momentan die wahrscheinlich beste AR-Brille auf dem Markt. Sie verfügt über mehrere Kameras und Tiefensensoren, womit die HoloLens in der Lage ist sich selbstständig im Raum zu orientieren.

Ausgestattet mit einem Intel Atom x5-Z810 Prozessor und 2 Gigabyte RAM, ist die Brille sehr gut aufgestellt und kann Programme und Berechnungen ohne zusätzliche Hardware ausführen. Als Betriebssystem kommt Windows 10 zum Einsatz. Die Akkulaufzeit wird mit zwei Stunden angegeben, wobei dies wie immer Abhängig von der geforderten Leistung ist. [?]

Für Entwickler hat Microsoft ein *Software Development Kit* (SDK) bereitgestellt mit dessen Hilfe es möglich ist Anwendungen für die HoloLens zu entwickeln. [?]

Über verschiedene Sensoren kann die HoloLens Bewegungen und Gesten des Nutzers wahrnehmen, erkennen und verarbeiten.

Die HoloLens ist standardmäßig in der Lage, grundlegende Gesten zu erkennen, mit denen zum Beispiel ein Tippen, Halten, Rotieren und Zoomen möglich ist. [?] Auf genauere technische Details zur Programmierung der HoloLens wird im Kapitel ????? genauer eingegangen.

Eine Besonderheit der HoloLens ist natürlich das mit Microsoft ein großer Softwarekonzern hinter der Entwicklung steht und somit ein schnelles voranschreiten der Technik quasi sicher ist.



Abbildung 5: Aufgesetzte HoloLens [?]

Die Epson Moverio BT-200 eine weitere AR-Brille ist die sogenannte Epson Moverio BT-200, welche auf Android-Basis arbeitet. Im Gegensatz zur HoloLens, kann die Moverio wie eine normale Brille getragen werden und sieht auch so aus.

Intern sorgt ein nicht genauer benannter 1,2 GHz Dual-Core-Prozessor für die nötige Rechenleistung. Die Brille verfügt über 8 Gigabyte Speicher, der sich mittels SD-Karten erweitern lässt.

Ähnlich wie die HoloLens verfügt auch die Moverio über eine Tiefenbildkamera, welche es ermöglicht AR-Inhalte exakt zu platzieren. Ein Beschleunigungssensor registriert zusätzlich die Bewegungen des Kopfes des Trägers.

Gesteuert wird die Brille über ein Controller-Kästchen, welche im Lieferumfang enthalten ist. Dieses verfügt zur Steuerung verschiedene Button wie Home oder Zurück, welche auch von Android bekannt sind. Zusätzlich kann das System über einen eingebauten Touchscreen im Controller Gesten des Nutzers erkennen und interpretieren.

Der Akku soll 6 Stunden durchhalten, was ein längeres Vergnügen verspricht, als es bei der HoloLens gegeben ist.

Der „Google Play“-Store ist auf der Brille nicht verfügbar, jedoch portiert Epson einige Anwendungen damit diese auf der Brille Lauffähig sind.

Bilder/moverio.jpg

Abbildung 6: Epson Moverio mit Controller [?]

3.2 Auswertung der Möglichkeiten für ein Natural User Interface

Für die Erstellung eines NUI gibt es Frontendseitig zwei Möglichkeiten. Zum einen gibt es die Handheld-Geräte, zum anderen gibt es noch die AR-Brillen.

Als Handheld-Gerät kann heute fast schon jedes Smartphones oder Tablet verwendet werden. Ohne eine eingebaute Magic-Lens 3.1.1 oder die Google Tango Technologie, ist das präzise arbeiten mit solchen Geräten nicht möglich. Hinzu kommt noch, dass der Benutzer des Interfaces immer mit einer Hand das entsprechende Gerät halten muss, wodurch diese dann nicht mehr zum Arbeiten zur Verfügung steht.

Das ständige halten eines Smartphones oder Tablets ist auch anstrengend für den Nutzer und seine Armmuskulatur.

Handheld-Geräte sind eine gute und auch einfache Lösung um Filme zu schauen oder Spiele zu spielen. Eine Langzeitnutzung wie in einem 8 stündigen Arbeitstag steht jedoch außer Frage.

Hier kommen die AR-Brillen ins Spiel. Sie ermöglichen eine Anreicherung der realen Welt ohne dass ein Nutzer permanent ein Gerät halten muss, da die Brillen durch die natürliche Kopfform halten.

Die Epson Moverio Brille ist ein erster Ansatz um AR-Brillen einer breiten Masse zur Verfügung zu stellen. Jedoch kann sie durch viele Nachteile nicht mit der HoloLens mithalten. Im Gegensatz zur Moverio kann sich die HoloLens selbstständig im Raum orientieren.

Ein weiterer K.O.-Punkt ist, dass die HoloLens Handgesten des Nutzers direkt erkennt ohne wie bei der Moverio auf ein zusätzlichen Controller zurück greifen zu müssen.

Im Vergleich zu Handheld-Geräten ist die Moverio nicht besser, da auch ständig ein Gerät zur Steuerung in der Hand des Nutzers verbleiben muss und somit nicht frei für andere Aufgaben ist.

Die Microsoft HoloLens ist bisher die einzige Möglichkeit um ein NUI zu entwickeln, wie es die Arbeit vorsieht. Sie kann völlig freihändig gesteuert werden ohne dass der Nutzer zusätzliche Hardware zum steuern oder Anzeigen in der Hand halten muss.

Im Verlauf der Arbeit wird durch die klare Überlegenheit der HoloLens diese als einzige weiter betrachtet und für die Entwicklung verwendet. Alle anderen alternativen konnten nicht überzeugen.

Anforderung sehr hoch	Priorisierung ✓	Handheld Geräte ✓	Microsoft HoloLens ✓	Epson Moverio BT-200 ✓
exakte AR-Inhalte	sehr hoch	✓	✓	✓X
Handgestenerkennung	gering	✓	✓	✓
Spracherkennung	hoch	✓	✓	X

Tabelle 1: Tabellarischer Vergleich der betrachteten Systeme

- 4 Artificial Intelligence
- 5 Beschreibung eines Natural User Interface
- 6 Versuch der Entwicklung eines Natural User Interface anhand von ... und ...
- 7 Fazit
- 8 zusammenfassung

9 Abkürzungsverzeichnis

KI *künstliche Intelligenz*

VR *Virtuelle Realität*

AR *Argumented Reality*

NUI *Natural User Interface*

AV *Augmented Virtuality*

SDK *Software Development Kit*

Abbildungsverzeichnis

1	Reality-Virtuality Continuum [?]	6
2	Unterschiedliche Blickfelder von Betrachter und Kamera bei Handheld-AR [?] . .	7
3	Schematische Darstellung des Magic Lens Effekts [?]	8
4	Schematische Darstellung des Aufbaus eines See-Through-Displays [?]	9
5	Aufgesetzte HoloLens [?]	9
6	Epson Moverio mit Controller [?]	10

Tabellenverzeichnis

1	Tabellarischer Vergleich der betrachteten Systeme	11
---	---	----

Literatur