

Gesten im dreidimensionalen Raum

Christian Blank

HAW Hamburg, Technik und Informatik,
Berliner Tor 7, Hamburg, Germany
christian.blank@hawhamburg.de
<http://www.haw-hamburg.de>

Zusammenfassung. Die Darstellung von digitalen Informationen und die Benutzung dieser Informationen durch den User hat sich seit der Einführung von Touch- und insbesondere von Multitouch-Devices sehr verändert. Die Informationen werden zumeist auf einem zweidimensionalen Bildschirm ausgegeben. Jedoch gibt es immer mehr Geräte, die auch die dreidimensionale Wiedergabe erlauben. Mit der Verbesserung und Etablierung dieser Technik muss auch über neue Bedienmöglichkeiten nachgedacht werden.

Grundlegend sind Maus, Tastatur und zweidimensionale Gesten nicht geeignet, um intuitiv mit Informationen im dreidimensionalen Raum zu arbeiten. Ein Ausweg ist die Nutzung von natürlichen, dreidimensionalen Gesten. Für die Erkennung von Gesten muss der Computer Bewegungen des Nutzers zunächst erkennen und dann analysieren.

Diese Arbeit gibt einen Überblick über den geschichtlichen Verlauf und den derzeitigen Stand der Entwicklungen bei der Aufnahme von Interaktionen zwischen Mensch und Computer, insbesondere die Interaktion mittels Gesten.

Schlüsselwörter: 3D;Gesten;HCI;Interaktion

1 Einleitung

Neben der Tastatur zählt die Maus auch aktuell zu den am weitesten verbreiteten Eingabegeräten. Jedoch werden spätestens seit der Einführung des iPhones im Jahr 2007 Multitouch-Devices immer beliebter ([App07]). Die steigende Beliebtheit und Akzeptanz führt dazu, dass alte Bedienkonzepte verdrängt und immer neue Möglichkeiten gesucht werden, um die Interaktion mit dem Computer zu vereinfachen und angenehmer zu gestalten. Waren es vor ein paar Jahren noch einige Ausnahmen, so wird heute fast jedes Notebook mit einem Multitouchscreen oder zumindest einem -pad ausgestattet und alle modernen Betriebssysteme unterstützen mittlerweile die Eingabe von Multitouchgesten.

Auch im Bereich der Wiedergabe von Inhalten gibt es seit einiger Zeit den Trend zur dreidimensionalen Darstellung. Zu Anfang waren es nur Kinohäuser, die eine Wiedergabe des Bildmaterials in 3D unterstützten. Mittlerweile besitzen zunehmend mehr Haushalte ein 3D-fähiges Fernsehgerät. Die Darstellung wird dabei immer realistischer und kommt bei manchen Technologien auch ohne zusätzliche Brillen aus, die den Betrachter behindern könnten. Die Oculus Rift hat einen weiteren Impuls gegeben und gezeigt, wie 'echt' sich die virtuelle Welt inzwischen anfühlt.

Die bisherigen Interaktionsmöglichkeiten mittels Tastatur, Maus und zweidimensionalen Gesten sind zwar sehr gut für die Arbeit mit zweidimensionalen Inhalten geeignet, aber nicht für Inhalte, die auch in der dritten Dimension dargestellt werden. Neben der Arbeit mit virtuellen Inhalten können dreidimensionale Gesten auch für die Kommunikation oder die Arbeit mit realen Objekten verwendet werden (vgl. Mixed Reality und CSCW).

1.1 Ziele

Ein Ziel dieser Arbeit ist es, einen Überblick über die derzeitigen Konzepte zu geben und die wichtigsten Arbeiten, Konferenzen und Forschungsgruppen zu präsentieren. Daneben sollen auch geschichtliche Aspekte beleuchtet und Anwendungsmöglichkeiten für die Steuerung durch dreidimensionale Gesten erarbeitet werden. Zudem wird eine eigene Einschätzung über das zukünftige Vorhaben und die Risiken, die bei der Bearbeitung dieses Themas auftreten können, abgegeben.

1.2 Aufbau

Die vorliegende Arbeit ist in fünf Abschnitte unterteilt. Die Einleitung befindet sich in Abschnitt 1. Danach folgt eine Einführung in die Mensch-Computer-Interaktion in Abschnitt 2 mit einem Überblick über die geschichtliche Entwicklung der verschiedenen Computerschnittstellen. In Abschnitt 3 wird der Bereich der Informatik genauer beschrieben, der sich mit der Gestenerkennung beschäftigt. Dieser Teil bildet den informatischen Teil meiner Arbeit. Der nicht-informatische Teil der Arbeit wird in Abschnitt 4 beschrieben. Zum Schluss wird die vorliegende Arbeit zusammengefasst, es wird ein Ausblick auf die künftigen Aufgaben und Ziele gegeben und in Abschnitt 5 werden die Risiken analysiert, die diese Aufgabe mit sich bringt.

2 Mensch-Computer-Interaktion

Seit einigen Jahrzehnten beschäftigen sich Informatiker und auch Geisteswissenschaftler, wie etwa Psychologen und Soziologen, mit der Frage, wie in Zukunft mit Computern kommuniziert werden könnte und was nötig sein muss, um eine breite Akzeptanz durch die Benutzer zu erhalten ([Suc87], [Suc07], [Hei12]). Mensch-Computer-Interaktion¹ ist die Disziplin, die sich mit der Gestaltung, Evaluierung und der Implementierung von interaktiven Computersystemen für den menschlichen Gebrauch und mit der Analyse von zentralen Phänomenen in deren Umfeld beschäftigt ([Bou08]). Es gibt zwei wesentliche Anforderungen an eine Schnittstelle: *Come as you are* und *Veränderlicher Kontext* werden im Folgenden näher erläutert.

Come as you are. An die Interaktion mit Computern werden mehrere Anforderungen gestellt. Einer der wichtigsten Aspekte ist der Wunsch nach einer einfachen, allgegenwertigen Lösung. "*Come as you are*" beschreibt dabei nicht nur den Wunsch, dass ein User keine Geräte oder Marker am Körper tragen sollte, sondern auch, dass die benötigte Sensorik möglichst gut verfügbar ist und einfach benutzt werden kann ([DBG13]). Das von Microsoft vorgestellte Project Natal, das mittlerweile als Microsoft Kinect vertrieben wird, macht einen großen Schritt in diese Richtung ([WA12]).

Veränderlicher Kontext. Die Schnittstelle, mit der ein Nutzer arbeitet, beeinflusst auch immer seine Umgebung und die Aufgabe, die er mit der Hilfe der Schnittstelle lösen will. Die Aufgabe und die Umgebung wiederum beeinflussen die Schnittstelle. Die Schnittstelle besitzt neben den Verbindungen zu Nutzer, Aufgabe und Umgebung auch eine direkte Verbindung zur Technologie.

Interaktive Systeme müssen für einen veränderlichen Kontext gestaltet werden und sich an diesen anpassen oder angepasst werden können. [Hel08]

Der Kontext wird dabei durch die Umgebung, die Aufgabe und den Nutzer definiert.

In diesem Abschnitt werden verschiedene Arten von Schnittstellen vorgestellt und es wird auf die historische Entwicklung eingegangen.

2.1 Multimodale Interaktion

Man spricht von einer multimodalen Schnittstelle oder Interaktion, wenn mehr als nur ein Sinn an der Kommunikation beteiligt ist. Die meisten Schnittstellen, die heute zur Verfügung stehen, ermöglichen nur eine Interaktion mit maximal zwei Modalitäten. Betrachtet man jedoch moderne Smartphones und Spielekonsolen, so kommt zu der Eingabe über Touchscreen bzw. Controller und der Spracheingabe noch das Force Feedback, das als eine rudimentäre Umsetzung von haptischem Feedback angesehen werden kann. Eine Kombination aus verschiedenen Eingabetypen kann in Zukunft zu einer größeren Nutzung von Computern in Bereichen führen, in denen sie derzeit noch nicht weit verbreitet sind. [Ovi99]

Als Möglichkeiten der Interaktion eignen sich **Controller**, wie etwa ein Joystick oder auch Maus und Tastatur oder **Lagesensoren**, wie sie in fast allen mobilen Geräten und auch in vielen Spielecontrollern verbaut werden. Auch **Sprachsteuerung** und **Tangibles** lassen sich zur Interaktion nutzen, sind aber im Gegensatz zu der **Toucheingabe** wenig verbreitet. Eine weitere sehr natürliche Möglichkeit mit einem Computer zu interagieren, ist über **Mimik** und **Gestik**.

¹ engl.: Human-Computer interaction (HCI)

2.2 Geschichte der MCI

Es gibt viele verschiedene Möglichkeiten für die Interaktion mit Computern. Dabei kann man sehr gut vier verschiedene Phasen ausmachen, die sehr unterschiedliche Konzepte repräsentieren. Angefangen von den Batch-Interfaces, die zu Anfang des 20. Jahrhunderts aufkamen, über die Command-Line-Interfaces und die Graphical User Interfaces bis hin zu den Natural User Interfaces.

Batch-Interface. Die erste Interaktion mit Maschinen und Computern wurde über Batch-Interfaces realisiert. Dabei wurden Lochkarten zunächst gestanzt und dann in einen Computer eingeführt. Die Interaktion entstand also nicht in Echtzeit, sondern mit einer Verzögerung. [Hel08]

Command-Line-Interfaces. In der Mitte der 1960er Jahre wurde der Schritt zu einer neuen Phase gemacht. Die CLI waren zunächst noch papierbasiert (vgl. Fernschreiber), jedoch nach kurzer Zeit wurden auch Videodisplays verwendet. Mittels einer Tastatur konnte ein Vokabular an Befehlen direkt eingegeben werden. [Hel08]

Graphical User Interfaces. Bereits 1951 entwickelte das US-Militär einige graphische Schnittstellen ([Hel08]), jedoch wurden erst 1973 die ersten GUI für kommerzielle Systeme am Xerox PARC entwickelt ([TML⁺79]). Die Bedienung erfolgt durch Maus, Tastatur, Joystick oder ähnlichem und ist einfach erlernbar, da es ein direktes Feedback auf die Aktion eines Nutzers gibt und viele Metaphern aus der Realität aufgegriffen werden (Papierkorb, Arbeitsplatz, Ordner). Ein Problem, das gerade in der Entstehungszeit der GUI sehr schwierig zu lösen war, ist die zusätzliche Rechenleistung, die nötig ist um eine GUI anzuzeigen. Durch die ständig steigende Rechenleistung moderner Computer ist die Belegung von Rechenkapazität durch eine GUI jedoch ohne weiteres möglich.

Natural User Interfaces. Bei den NUI werden virtuelle und reale Objekte auf die gleiche Weise gesteuert. Die Kommunikation mit dem Computer ist direkt, intuitiv und leicht zugänglich. Die NUI soll dabei nicht CLI und GUI ersetzen, sondern nur eine Alternative zu ihnen bieten ([Man01]). Es fallen viele verschiedene Interfaces unter die Kategorie der NUI, etwa Mimik- und Gestensteuerung, Spracheingabe oder auch haptische Touchpads.

Neben den vier genannten Gruppen kann man noch weitere Konzepte finden, die nicht genau einer Gruppe zuzuordnen sind (vgl. Ubiquitous Computing [Wei94]). Zudem gibt es weitere Entwicklungen, die ein völliges Umdenken im Umgang mit dem Computer erfordern (vgl. Sensual Interfaces [Woe07]).

3 Gestenerkennung

Im Gegensatz zu den ersten drei Interfaces, die in Abschnitt 2 beschrieben wurden, gibt es im Bereich der NUI noch einen sehr großen Forschungs- und Entwicklungsbedarf. Obwohl bereits seit den 1990er Jahren intensiv geforscht wurde, konnten sich erst einige der Ideen im Alltag durchsetzen. Wie bereits erwähnt, ist es kein Wunsch von NUI herkömmliche Interaktionsformen zu ersetzen, sondern eine Alternative zu bieten, die bei vielen Aufgaben eine wesentlich natürlichere Bearbeitung ermöglicht, da sie sich mehr an die gestellte Aufgabe anpasst. Im Gegensatz dazu verwenden GUI immer wieder die Arbeitsplatz-Metapher und passen alle Aufgaben so an, dass sie diesem mehr oder weniger genügen. [Wei94][Hel08]

Dieser Abschnitt enthält neben einer Definition für Gesten auch eine erste Gegenüberstellung zwischen zweidimensionalen und dreidimensionalen Gesten. Es werden die allgemeinen Schritte erklärt, die zur Gestenerkennung nötig sind und es werden verschiedene aktuelle Lösungen gezeigt, mit deren Hilfe Gesten erkannt werden können. Zum Ende des Abschnittes werden noch relevante Konferenzen, Arbeitsgruppen und Veröffentlichungen vorgestellt.

3.1 Gesten

Unter vielen anderen sind Gesten ein Bestandteil von natürlichen Benutzerschnittstellen. Sie spielen in der zwischenmenschlichen Kommunikation eine wichtige Rolle und können für sich alleine stehen oder im Zusammenhang mit einem anderen Kommunikationsmittel verwendet werden.

Geste: Eine Geste ist die Bewegung des Körpers von einer Person, die einem Beobachter eine Information mitteilen soll. [WF98]

Durch die Definition werden Bewegungen, die bspw. beim Zeichnen oder einem Tastendruck durchgeführt werden, nicht als Geste anerkannt. Sie sollen nicht die eigentliche Information transportieren, sondern sie nur dazu beitragen. Es gibt noch weitere Definitionen, die teilweise statische Haltungen einbeziehen oder explizit ausschließen.

Bei der Bedienung von Tablets, Smartphones, Tabletops und PCs erleichtern Gesten, die meist über ein Touchdisplay oder -trackpad aufgenommen werden, die Ausführung von Aktionen. Diese zweidimensionalen Gesten sind einfach zu erlernen und zu verwenden. Somit sinkt die Einstiegshürde für die Interaktion mit einem Computer und die Nutzergruppe kann erweitert werden.

Auf Touchscreens oder -pads können nur zweidimensionale Gesten ausgeführt werden, da die dritte Dimension nicht von den Sensoren erfasst werden kann². Auch bei Sensoren, die Bewegungen im Raum vollständig erfassen können, wird oftmals eine virtuelle Ebene eingesetzt, auf der die Nutzer interagieren können (vgl. Abbildung 1). Die meisten zweidimensionalen Gesten sind Zeigegesten, wie etwa das Tippen oder Zoomen.



Abb. 1. Der Finger muss für eine gültige Eingabe die virtuelle Ebene durchstoßen.

Im Unterschied zu den zweidimensionalen Gesten gibt es bei den dreidimensionalen Gesten noch kein Eingabealphabet, das sich allgemein durchgesetzt hat. Das liegt zum einen an der wesentlich größeren Anzahl an möglichen Gesten und zum anderen an der schwierigen Erkennung der Gesten und des Kontextes, in dem die Geste ausgeführt wird. Somit gibt es für einzelne Anwendungen jeweils eigene Gesten, die der Nutzer erlernen und unterscheiden muss. Die gewünschte Vereinfachung wird damit nicht erreicht.

3.2 Schritte

Ein Computer benötigt für die Gestenerkennung mehrere Schritte (Abbildung 2), die bei allen Verfahren gleich sind, egal welche Art von Sensorik verwendet wird.

In der *Signalverarbeitung* werden die Daten, die durch die Sensoren aufgenommen wurden, zusammengefasst und vorverarbeitet (Rauschen entfernen etc.). Je nach Quelle müssen dabei unterschiedliche Dinge beachtet werden. Unter anderem kann das Datenvolumen sehr variieren. Kamerabasierte Verfahren liefern komplette Bilder, andere Sensoren vielleicht nur einige Winkel.

² Apple Inc. arbeitet an einer Technik, die es ermöglicht, auch auf Touchscreens dreidimensionale Gesten zu erkennen[KB13].

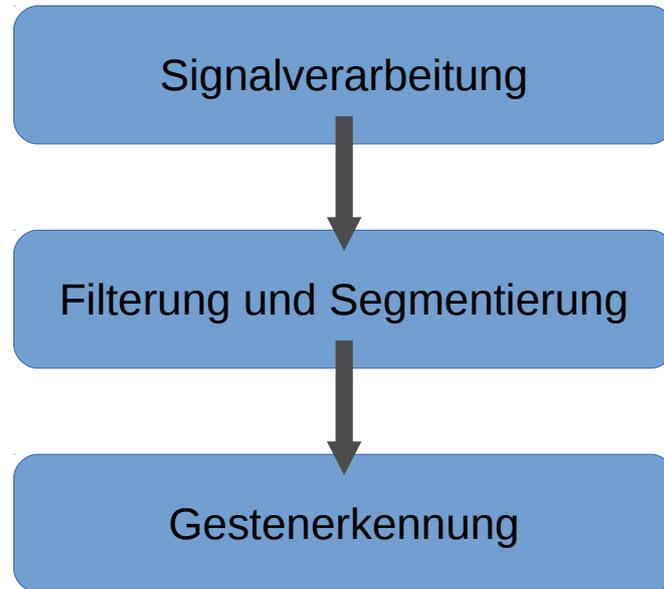


Abb. 2. Die Verarbeitungsschritte zum Erkennen einer Geste.

Im nächsten Schritt können die gesammelten Daten *gefiltert* werden, sodass nur noch die relevanten Informationen übrig bleiben. Diese werden dann je nach Ansatz unterschiedlich ausgewertet. Einige untersuchen den kompletten Datenstrom nach einer initialen Bewegung oder Haltung und dem Ende einer Geste, andere sehen die komplette Aufnahme als eine Geste an, die durch ein externes Event (bspw. ein Tastendruck) beginnt und endet.

Die eigentliche *Gestenerkennung* findet im letzten Schritt statt. Mithilfe von Techniken aus der künstlichen Intelligenz können Gesten durch den Computer ‘verstanden’ werden. Zu diesen Techniken gehören Hidden Markov Modelle, künstliche neuronale Netze und Conditional Random Fields.

3.3 Techniken

Es gibt bereits einige Möglichkeiten, um einen Computer mithilfe von Gesten zu steuern. Verschiedene Firmen und Forschungsgruppen haben sich dabei mit unterschiedlichen Ansätzen befasst. In diesem Abschnitt werden eine Auswahl an Umsetzungen genannt.

Microsoft hat mit der Kinect eine weit verbreitete 3D-Kamera entwickelt³ und nutzt diese unter anderem bei ihren See-Through-Displays ([HKI⁺12], [LOIB13]), um die Bewegungen der Hand und andere Gegenstände im Aufnahmebereich zu verfolgen. So können durch Gesten Karteikarten ausgewählt (zu sehen in Abbildung 3) oder Textpassagen kopiert werden. Eine andere Möglichkeit ist das Spielen mit einem virtuellen Ball.

Im Gegensatz zu dem Verfahren, das Microsoft nutzt, müssen die Nutzer bei einem markerbasierten Motioncapturing zusätzliche Hilfsmittel am Körper tragen, um von dem Computer erkannt zu werden. Genutzt wird dieses Verfahren meist in der Film- und Spieleindustrie, aber auch viele virtuelle Käfige statten Nutzer oder Eingabegeräte mit Markern aus, um sie zu verfolgen.

Die SpaceGlasses von Meta und Google Project Tango verwenden eine ähnliche Technologie wie die Kinect. Während bei dem Project Tango aber die Interaktion über das Display im Vordergrund steht, erkennt SpaceGlasses die Armbewegungen des Trägers und wertet diese aus ([Goo13], [Met13]). Auch die Leap Motion kann dieser Kategorie zugeordnet werden. Sie kann ein Handpaar aufnehmen und dieses in ein 3D-Modell transferieren ([Hod13]). Die Daten, die sie ausgibt, können von Applikationen verwendet werden.

³ Die Kinect wurde ursprünglich von PrimeSense entwickelt, aber neuere Versionen sind Eigenentwicklungen von Microsoft.



Abb. 3. Der See-Through-Monitor, der unter anderem von Jinha Lee (MIT Media Labs, Microsoft Research) entwickelt wurde [LOIB13].

Eine weitere Möglichkeit Gesten aufzunehmen sind Datenhandschuhe. Die ersten von ihnen wurden schon 1977 entwickelt und im Laufe der Jahre immer weiter verbessert. Anfang der 1990er Jahre wurden am MIT unter anderem ICONIC entwickelt. Dieses nutzen eine multimodale Interaktion, bestehend aus Sprache und Gesten ([KS94], [SZ94]). Eine Weiterentwicklung der Datenhandschuhe ist das Armband Myo von Thalmic Labs, dass in Abbildung 4 zu sehen ist ([Nuw13]). Es misst die elektrischen Impulse, die bei der Muskelkontraktion entstehen und wandelt diese in Bewegungsinformationen um. Dadurch kann der Computer feststellen, wie die Hände gedreht sind und welcher Stellung sich die einzelnen Finger befinden ([NKA10]).

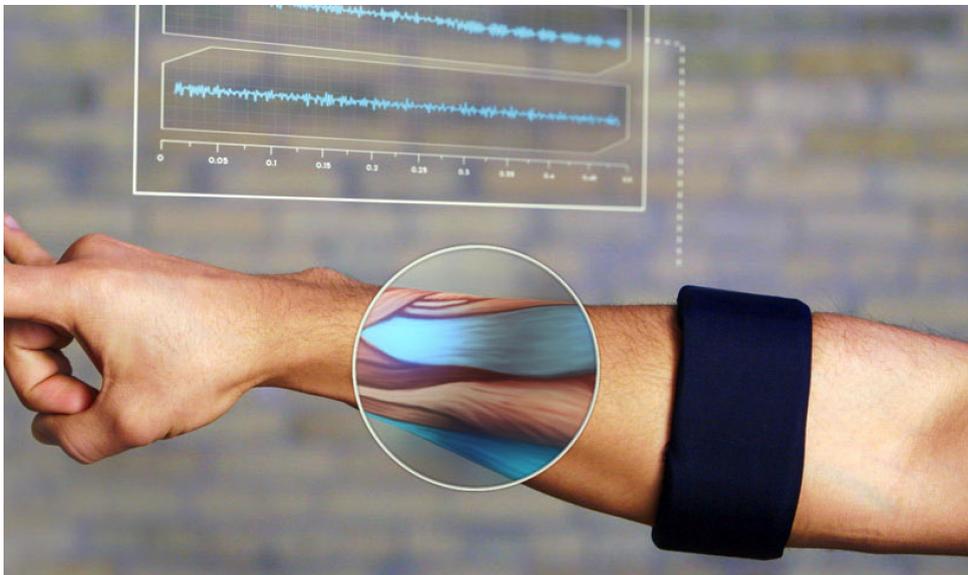


Abb. 4. Das Myo-Armband misst elektrische Impulse der Muskulatur, damit die Bewegung des Armes für den Computer sichtbar werden.

Eigentlich zählen Bewegungen des Gesichtes zur Mimik und nicht zur Gestik, aber dennoch sind Kopf- und Augen-Tracker wichtige Hilfsmittel für die Analyse von Gesten, da sie etwas über den Kontext verraten, in der eine Geste ausgeführt wird. So wird unter anderem auch bei [HKI⁺12] und [LOIB13] Augentracking verwendet, um die korrekte perspektivische Ansicht zu ermöglichen. Wird ein kontinuierlicher Eingabestrom ausgewertet, dann ist das Wissen über den Blickpunkt des Nutzers sehr wichtig für eine Auswertung.

3.4 Publikationen, Arbeitsgruppen und Konferenzen

Da an dem Thema der Mensch-Computer-Interaktion bereits seit Jahren gearbeitet wird, gibt es ein sehr breites Feld an Publikationen, Konferenzen und Arbeitsgruppen. Die Gestenerkennung und die Bedienung von Computern mittels Gesten sind immer wieder ein wichtiger Bestandteil der Forschung. Die wichtigsten werden in diesem Abschnitt aufgelistet.

Tabelle 1. Wichtige Konferenzen 2014 im Bereich der Mensch-Computer-Interaktion und der 3D-Interfaces

Konferenz	Datum	Ort	Thema
IEEE 3D User Interfaces	29.03.	Minneapolis	“Interacting in Spatial Augmented Reality”
Computer Human Interaction	26.04.	Toronto	-
HCI International	22.06.	Heraklion,GR	“HCI Design for Immersive Systems and 3D Environments”
SIGGRAPH Conference	10.08.	Vancouver	“Naturally Digital”

Viele der großen Konferenzen finden im nordamerikanischen Raum statt. Die HCI International wird 2014 in Griechenland ausgetragen. Kapstadt war 2013 Austragungsort der IFIP Conference on Human-Computer Interaction (INTERACT). Eine Weiterführung wurde bisher nicht bestätigt⁴. Die SIGGRAPH Conference ist eine sehr große Konferenz für alle möglichen Bereiche der Computergrafik. Teilbereiche beschäftigen sich auch mit Computerschnittstellen.

Tabelle 2. Publikationen im Bereich MCI und den Themen 3D-Interfaces und Bildverarbeitung

Autor	Titel	Jahr	Art
Bernin, A.	Einsatz von 3D-Kameras zur Interpretation von räumlichen Gesten im Smart Home Kontext	2011	BA
Dieck, H.	Modellbasierte Handposenerkennung	2013	BA
Dix, A.	Human-computer Interaction	2004	Buch
Jiang, X.	Advances in Depth Images Analysis and Applications	2012	Buch
Hellige, H. D.	Mensch-Computer-Interface	2008	Buch
Koch, W.	Tracking and Sensor Data Fusion	2013	Buch
LaViola Jr, J.	3D Gestural Interaction: The State of the Field	2013	Paper
Nitschke, C.	3D Reconstruction	2007	Buch
Potratz, O.	Ein System zur physikbasierten Interpretation von Gesten im 3D-Raum	2011	MA
Wang, J.	Human Action Recognition with Depth Cameras	2014	Buch

In der Tabelle 2 wird eine Auswahl an Publikationen aus den letzten zehn Jahren aufgelistet. Einige bieten einen sehr guten Überblick über die MCI ([Hel08]), die Sensor-Fusion ([Koc13]) oder die Datenverarbeitung von 3D-Informationen ([Nit07], [JBGO]), andere geben einen detaillierteren Einblick in spezifischere Themengebiete, wie etwa das Handtracking ([Die13]) oder die räumlichen Gesten ([Ber11], [Pot11]). Eine Zusammenfassung über den aktuellen Stand der Forschung im Bereich der dreidimensionalen Gesten ist in [LaV13] zu finden.

Im Gegensatz zu den SIGs⁵ SIGCHI und SIGGRAPH gehört die SIGHCI nicht zur ACM⁶ sondern zur Association for Information Systems (AIS). Das deutsche Äquivalent zu SIGCHI und SIGHCI ist der Fachbereich MCI der Gesellschaft für Informatik. Die SIGCHI, der Fachbereich MCI und auch die SIGHCI beschäftigen sich mit allen Bereichen der Mensch-Computer-Interaktion. Die SIGGRAPH beinhaltet alle Themen rund um die Computergrafik.

⁴ Stand: 04.03.2014

⁵ Special Interest Groups

⁶ Association for Computing Machinery

Aufgrund des sehr großen Spektrums an Themen und der intensiven Forschung im Bereich der MCI ist es schwierig, *Keyplayer* zu identifizieren. Vergleicht man die Häufigkeit der Zitate und die Veröffentlichungen in den letzten Jahren, dann können **Ben Shneiderman** von der University of Maryland im Bereich MCI und **Joseph J. LaViola Jr.** von der University of Central Florida im Bereich 3D-Gesten genannt werden.

4 Einsatzmöglichkeiten

Die Einsatzmöglichkeiten für dreidimensionale Gesten können sehr vielseitig sein. In diesem Abschnitt werden Anwendungen vorgestellt, die es so möglicherweise schon gibt, die aber durch den Einsatz von dreidimensionalen Gesten wesentlich verbessert werden könnten.

Virtueller Käfig. In der Automobilindustrie werden virtuelle Käfige eingesetzt, um kostengünstige Prototypen zu erstellen und zu begutachten. Sie werden aber auch im Verkauf und in der Schulung der Techniker verwendet. Bisher steuern die Nutzer die Ansicht immer noch mit einem Controller, meist eine Art Pistolengriff mit Markern. Wenn sie Gesten zur Steuerung nutzen könnten, dann wäre es ein wesentlicher Vorteil, weil dadurch ein deutlich natürlicheres Gefühl entstehen würde. Zum Beispiel könnte der Techniker durch die Betätigung eines virtuellen Hebels die Motorhaube öffnen und sie dann mit seinen Händen nach oben drücken bis sie einrastet.

Medizin. In der Chirurgie werden oft Computer zur Unterstützung genutzt. Sie dienen als Nachschlagewerk, als Patientenakte, als Sehhilfe, die die vorhandenen Aufnahmen durch Bildbearbeitung verbessern können, oder als Monitor bei minimal-invasiven Eingriffen. Die Bedienung muss meist durch eine weitere Person erfolgen, weil der operierende Arzt nicht ständig zwischen Computer und Patient wechseln kann. Besser wäre aber eine berührungslose Steuerung des Computers durch den Operateur. Mithilfe von dreidimensionalen Gesten könnte der Arzt den Computer steuern und sich trotzdem auf den Patienten konzentrieren können.

Home Control. Wie bereits in [Ber11] beschrieben, können Gesten sehr gut für den Einsatz in einem intelligenten Haus genutzt werden. Man kann mit einfachen Bewegungen das Licht dimmen, den Thermostat regeln oder den Fernsehgerät stumm schalten. Dabei muss man noch nicht einmal im selben Raum wie das Gerät sein. Auch umfangreichere Aufgaben könnten mit einer einzigen Handbewegung erledigt werden, die sonst viel Zeit in Anspruch nehmen würden. So könnte man vor dem Verlassen des Hauses alle Fenster schließen, die Heizungen und das Licht abschalten und alle nicht relevanten Geräte vom Strom nehmen, ohne durch das gesamte Haus laufen zu müssen.

Computer Supported Cooperative Work. Gesten machen einen wichtigen Teil der Kommunikation zwischen Menschen aus. Wenn wir uns einen virtuellen Arbeitsplatz vorstellen, an dem ein Meeting mit Personen, die über große Entfernungen verteilt sind, stattfindet, dann begeben wir uns in die virtuelle Realität. Jede Person in dem Meeting (virtuelle Welt) wird durch einen Avatar dargestellt. Wenn dieser Avatar die Gesten eins zu eins wiedergeben kann, dann wird die Immersion der virtuellen Welt wesentlich erhöht. Für dieses Szenario muss der Computer noch nicht verstehen, was der Nutzer mit der Geste meint. Er muss sie einfach nur wiederholen. Wenn aber die Nutzer zusammen an einem virtuellen Produkt arbeiten, dann muss der Computer wissen, was der Nutzer mit einer Bewegung ausdrücken möchte.

Es fallen sicherlich noch weitere Einsatzmöglichkeiten an. Beispielsweise können Rapid Prototyping, Musik- oder Filmproduktionen durch den Einsatz von einem neuen Interface, das mithilfe von dreidimensionalen Gesten gesteuert wird, profitieren.

5 Fazit

Diese Arbeit gibt einen Einblick in die Mensch-Computer-Interaktion und bietet einen Überblick über verschiedene Ansätze zum Erkennen von dreidimensionalen Gesten. Diese Arbeit hat weder den Anspruch auf Vollständigkeit noch wurden alle Verfahren komplett durchdrungen. Es wurde bewusst darauf verzichtet, die Verfahren miteinander zu vergleichen oder eine Wertung abzugeben. Die vorliegende Arbeit ist im Rahmen von Anwendungen 1 (AW1) an der HAW Hamburg entstanden. AW1 dient den Studierenden als eine Möglichkeit, sich einen Überblick über das Anwendungsgebiet zu erarbeiten und sich die Grundlagen anzueignen.

Nachfolgend wird ein Ausblick über die Schritte und Ziele gegeben, die im Laufe des nächsten Studienjahrs im Hinblick auf die Masterarbeit anstehen. Abschließend werden mögliche Risiken genannt und es werden Vorschläge gemacht, wie man die Risiken nach Möglichkeit minimieren kann.

5.1 Ausblick

Die Literatur, die in Abschnitt 3.4 aufgelistet wird und ebenso die Konferenzen und Ergebnisse der Arbeitsgruppen, die im selben Abschnitt zu finden sind, sollen in AW2 durchdrungen werden. Es muss sich in die verschiedenen Verfahren eingearbeitet werden, die für die Lösung der Probleme zur Verfügung stehen und es sollten die geeignetsten Verfahren gefunden werden.

In Projekt 1 wird zusammen mit anderen Studierenden des Projektes I²E an einem Prototypen gearbeitet, an dem Mixed Reality *erlebt* werden kann⁷. Das erste Ziel ist es, einen Tisch zu bauen, an dem eine Person mithilfe einer VR-Brille 3D-Modelle betrachten kann. In Abbildung 5 wird ein Konzept des Prototyps gezeigt, auf dem ein Nutzer (auf dem Bild als Avatar dargestellt) mit einem virtuellen Modell interagiert. Die Modelle können von dem Nutzer durch Gesten gewechselt werden. Betreut wird das Projekt durch Frau Prof. Dr. Birgit Wendholt.

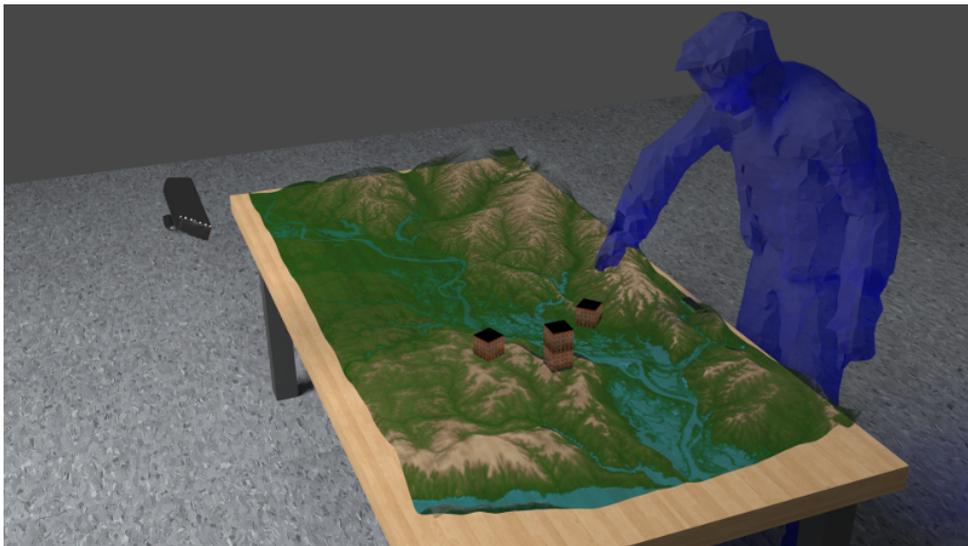


Abb. 5. Virtuelle Landschaft, Häuser und Avatar auf dem Tisch-Prototypen

Neben den allgemeinen Arbeiten an dem Prototypen wird sich dabei speziell mit der Gestensteuerung beschäftigt und zunächst werden einige einfache Gesten implementiert. Es sind für die Umsetzung folgende Arbeitsschritte identifiziert worden:

- Erstellung eines Konzeptes für die Softwarearchitektur und den Tisch
- Umsetzung der beiden Konzepte
- Einarbeitung in verschiedene Verfahren und von Vergleichen der Verfahren
- Auswahl der passenden Verfahren
- Erstellung einer Architektur für Gestensteuerung und prototypische Umsetzung

Weitere Arbeitsschritte sollten dann in Anwendungen 2 identifiziert und im Rahmen von Projekt 2 umgesetzt werden.

5.2 Risiken

Es gibt mehrere Risiken die den Erfolg der Masterarbeit verhindern oder beeinträchtigen können.

Die Gruppe muss einen Prototypen bauen, an dem dann die verschiedenen Arbeiten realisiert werden können, unter anderem auch die Steuerung mittels dreidimensionaler Gesten. Es kann passieren, dass zu dem benötigten Zeitpunkt noch kein Prototyp zum Testen der Ergebnisse vorhanden ist. Um sich dagegen abzusichern bietet es sich an, die Umsetzung so unabhängig wie möglich von dem Prototypen zu gestalten.

⁷ Projektwebsite: <http://i2e.informatik.haw-hamburg.de>

Für den Prototyp wird Hardware benötigt, die derzeit noch mit großen Lieferschwierigkeiten behaftet ist. Es soll deshalb schon von Anfang an darauf geachtet werden, eine Alternative für den Fall zu haben, dass Komponenten nicht lieferbar sind.

Da zur gleichen Zeit mehrere Studenten mit ihrer Masterarbeit an dem Tisch arbeiten, kann es zu unterschiedlichen Ansichten und Prioritäten kommen. Neben den fachlichen Fähigkeiten muss auch auf ein gutes Arbeitsklima innerhalb der Gruppe geachtet werden. Von großer Wichtigkeit ist es, dass Entscheidungen getroffen werden, mit denen sich alle Beteiligten identifizieren können.

Es sind einige technische Probleme zu lösen, die nicht direkt mit dem Thema in Verbindung stehen. Dadurch kann man Gefahr laufen, sich zu sehr mit diesen aufzuhalten und wichtige Zeit zu verlieren. Als Beispiele wären Sicherheit von Daten oder schnelle Übertragung von Informationen über das Netzwerk zu nennen. Um dem entgegenzuwirken, sollten solche Punkte von Anfang an mit Forschungsgruppen aus der HAW Hamburg und anderen Einrichtungen abgestimmt und ausgetauscht werden.

Zudem kann es passieren, dass eine ähnliche Entwicklung vor Beendigung der Masterarbeit veröffentlicht wird. Es würde die Arbeit obsolet machen.

Literatur

- [App07] APPLE: *Apple erfindet mit dem iPhone das Mobiltelefon neu @ONLINE*. <https://www.apple.com/de/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html>. Version: Januar 2007
- [Ber11] BERNIN, Arne: *Einsatz von 3D-Kameras zur Interpretation von räumlichen Gesten im Smart Home Kontext*. Berliner Tor 5, 20099 Hamburg, HAW Hamburg, Diplomarbeit, 2011
- [Bou08] BOUSTANI, Christian E.: *Human-Computer-Interaction (HCI)*. http://www.se.uni-hannover.de/priv/lehre_2008winter_seminar/Christian_El_Boustani-Human-Computer-Interaction-Folien.pdf. Version: 2008
- [DBG13] DIONISIO, John David N. ; BURNS, William G. ; GILBERT, Richard: 3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities. In: *ACM Comput. Surv.* 45 (2013), Juli, Nr. 3, 34:1–34:38. <http://dx.doi.org/10.1145/2480741.2480751>. – DOI 10.1145/2480741.2480751. – ISSN 0360–0300
- [Die13] DIECK, Hannes: *Modellbasierte Handposenerkennung*. (2013)
- [Goo13] GOOGLE: *ATAP Project Tango – Google @ONLINE*. <https://www.google.com/atap/projecttango/>. Version: März 2013
- [Hei12] HEINECKE, Andreas M.: *Mensch-Computer-Interaktion: Basiswissen für Entwickler und Gestalter*. Springer-Verlag New York Incorporated, 2012
- [Hel08] HELLIGE, Hans D.: *Mensch-Computer-Interface*. In: *Zur Geschichte und Zukunft* (2008)
- [HKI⁺12] HILLIGES, Otmar ; KIM, David ; IZADI, Shahram ; WEISS, Malte ; WILSON, Andrew: HoloDesk: direct 3d interactions with a situated see-through display. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* ACM, 2012, S. 2421–2430
- [Hod13] HODSON, Hal: Leap Motion hacks show potential of new gesture tech. In: *New Scientist* 218 (2013), Nr. 2911, S. 21
- [JBGO] JIANG, Xiaoyi ; BELLON, Olga Regina P. ; GOLDFOG, Dmitry ; OISHI, Takeshi: *Advances in Depth Image Analysis and Applications*.
- [KB13] KING, Nicholas V. ; BENJAMIN, Todd: *United States Patent: 8514221 - Working with 3D objects*. August 2013
- [Koc13] KOCH, Wolfgang: *Tracking and Sensor Data Fusion*. Springer, 2013
- [KS94] KOONS, David B. ; SPARRELL, Carlton J.: Iconic: speech and depictive gestures at the human-machine interface. In: *Conference Companion on Human Factors in Computing Systems*. New York, NY, USA : ACM, 1994 (CHI '94). – ISBN 0–89791–651–4, 453–454
- [LaV13] LAVIOLA, Joseph J.: 3D Gestural Interaction: The State of the Field. In: *ISRN Artificial Intelligence* (2013)
- [LOIB13] LEE, Jinha ; OLWAL, Alex ; ISHII, Hiroshi ; BOULANGER, Cati: SpaceTop: integrating 2D and spatial 3D interactions in a see-through desktop environment. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ACM, 2013, S. 189–192
- [Man01] MANN, Steve: *Intelligent Image Processing*. New York, NY, USA : John Wiley & Sons, Inc., 2001. – ISBN 0471406376
- [Met13] META: *Meta - SpaceGlasses @ONLINE*. <https://www.spaceglasses.com/>. Version: März 2013
- [Nit07] NITSCHKE, Christian: *3D Reconstruction*. Saarbrücken, Germany, Germany : VDM Verlag, 2007. – ISBN 3836410621, 9783836410625
- [NKA10] NAIK, Ganesh ; KUMAR, Dinesh ; ARJUNAN, Sridhar: Pattern classification of Myo-Electrical signal during different Maximum Voluntary Contractions: A study using BSS techniques. In: *Measurement Science Review* 10 (2010), Nr. 1, S. 1–6
- [Nuw13] NUWER, Rachel: Armband adds a twitch to gesture control. In: *New Scientist* 217 (2013), Nr. 2906, S. 21
- [Ovi99] OVIATT, Sharon: Ten myths of multimodal interaction. In: *Communications of the ACM* 42 (1999), Nr. 11, S. 74–81

- [Pot11] POTRATZ, Olaf: *Ein System zur physikbasierten Interpretation von Gesten im 3D-Raum*. 2011
- [Suc87] SUCHMAN, Lucille A.: *Plans and situated actions: the problem of human-machine communication*. Cambridge university press, 1987
- [Suc07] SUCHMAN, Lucy: *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press, 2007
- [SZ94] STURMAN, David J. ; ZELTZER, David: A survey of glove-based input. In: *Computer Graphics and Applications, IEEE* 14 (1994), Nr. 1, S. 30–39
- [TML⁺79] THACKER, Charles P. ; MCCREIGHT, Edward M. ; LAMPSON, Butler W. ; SPROULL, Robert F. ; BOGGS, David R.: *Alto: A personal computer*. Xerox, Palo Alto Research Center, 1979
- [WA12] WEBB, Jarrett ; ASHLEY, James: *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress, 2012
- [Wei94] WEISER, Mark: Creating the invisible interface:(invited talk). In: *Proceedings of the 7th annual ACM symposium on User interface software and technology ACM*, 1994, S. 1
- [WF98] WACHSMUTH, Ipke ; FRÖHLICH, Martin: *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop, Bielefeld, Germany, September 17-19, 1997, Proceedings*. Bd. 1371. Springer, 1998
- [Woe07] WOEBKEN, Chris: *New Sensual Interfaces*. (2007)

Abbildungsverzeichnis

1	www.omekinteractive.com (abgerufen: 03.03.2014).....	4
2	Die Verarbeitungsschritte zum Erkennen einer Geste.	5
3	leejinha.com (abgerufen: 05.03.2014)	6
4	thalmic.com (abgerufen: 05.03.2014)	6
5	i2e.informatik.haw-hamburg.de (abgerufen: 06.03.2014).....	9