

GESTURESPACE: Gestische Steuerung einer immersiven Projektion dreidimensionaler Inhalte

Kai Jauslin, Max Rheiner, Christian Weber, Gerhard M. Buurman

Zürcher Hochschule der Künste, Departement Design

Interaction Design Programm

Ausstellungsstrasse 60, 8005 Zürich, Switzerland

{kai.jauslin,max.rheiner,christian.weber,gerhard.buurman}@zhdk.ch

Mai 2009

ÜBERSICHT

In dieser Arbeit geht es um die berührungslose Steuerung interaktiver Anwendungen mittels Körpergesten eines Benutzers. Es wird ein neuartiges Konzept für das Tracking und die Erkennung von Hand- und Körpergesten entwickelt. Dieses basiert auf der kombinierten Auswertung von 3D-Distanzdaten und 2D-Bilddaten und benötigt so keine weiteren Hilfsmittel auf Seiten des Benutzers. Es wird eine einfache Gestensprache aufgebaut und experimentell anhand verschiedener Anwendungen in einem Gesamtprototyp umgesetzt und evaluiert. Eine Projektion auf den Boden stellt die virtuellen Inhalte direkt im Handlungs- und Bildraum des Benutzers dar. Diese sind dreidimensionaler Natur und richten sich perspektivisch auf die Position des Benutzers im Raum aus, wodurch ein hoher Grad an Immersion erreicht werden kann.

1. EINFÜHRUNG

Die Erwartungen an Benutzerschnittstellen interaktiver Systeme werden immer höher. Die Gestaltung unseres Lebensraums über und mit diesen Systemen, macht diese allgegenwärtig und durchdringt alle Bereiche unseres Alltags. Diese Ubiquität führt zu einer wechselseitigen Anpassung von Mensch und Maschine. Wir erwarten in dieser symbiotischen Beziehung, dass sich die Maschine ebenfalls der menschlichen Natur annähert und dadurch geeigneter auf unsere kommunikativen Äusserungen reagieren kann.

Insbesondere gilt dies in Bezug auf die Hilfsmittel, welche zur Kommunikation mit der Maschine benötigt werden. Die meisten Systeme beschränken die Kommunikation auf einen visuellen Kanal. Doch auch unsere anthropologisch motivierten angeborenen Werkzeuge, der Körper im Allgemeinen und die Hände im Speziellen, sollten sinnvolle Aufgaben erfüllen können. Dabei gilt es zu beachten: unser Körper ist auf Bewegung ausgelegt – wir sind keine „sitzenden Büroaffen“. Wir agieren mit unserem ganzen Körper in einem persönlichen Handlungsraum als „Bewegungstiere“. Unser Blickpunkt und Aufmerksamkeit ist in diesen Raum hinaus orientiert und erwartet von dort sinnliche Rückmeldungen in den unterschiedlichsten Formen. Die Fenster-Metapher des klassischen Bildschirms befriedigt nur einen kleinen Teil unseres Selbst.

Aktuelle Entwicklungen im Bereich der gestischen Schnittstellen – wie das berührungssensitive Display des Apple iPhone oder spezialisierte Multitouch-Tische – binden die Hände in die Interaktion mit ein. Dabei wird einerseits der

motorische Trieb aktiviert, andererseits das grösste Sinnesorgan – die Haut – in die Interaktion mit eingebunden. Um die Natürlichkeit zu steigern ist es wichtig, dass diese Interaktion ohne die Zuhilfenahme von zusätzlichen Werkzeugen, wie z.B. Handschuhe, erfolgen kann. Die gestische Sprache der Berührung befriedigt zusammenfassend eine sinnliche Komponente, mit deren Hilfe wir unseren Gedanken auf analoger Ebene mehr und besser Ausdruck verschaffen können. Im Gegenzug sind wir gerne auch bereit, einfache gestische Sprachformeln des Systems zu erlernen.

Meine Arbeit greift dieses Konzept der sinnlichen Handlung der Berührung auf und überträgt es auf den berührungslosen Raum. Dieser lässt sich gestisch erfahren und erleben. Der Benutzer kann Hände und Körper frei bewegen und somit seinem Bedürfnis nach körperlicher Aktivität implizit und explizit Ausdruck verleihen. Dieser berührungslose Handlungsraum wird um eine immersiven Komponente erweitert. Inhalte aus dem virtuellen Bezugssystem der Anwendung werden im Handlungsraum auf die Bodenfläche rund um den Benutzer projiziert. Die dreidimensionalen Inhaltsstrukturen richten ihre Perspektive so auf den Benutzer aus, dass von seinem Standpunkt die Illusion räumlicher Strukturen im Handlungsraum wahrnehmbar wird. Dadurch erreicht die Arbeit einen hohen Grad an natürlicher Immersion und wird räumlich erlebbar.

2. KUNSTGESCHICHTLICHER HINTERGRUND

Das Untersuchen des Verhältnisses von Bild und Betrachter, Bildraum und Realraum hat in der Kunstgeschichte eine lange Tradition. Die Vorstellungen von Raum haben sich über die Jahrhunderte stark verändert. Der Bildraum kann dabei als „plastische Manifestation der Wirklichkeitsauffassung eines Künstlers“ angesehen werden [2]. Ein Bild vermittelt mit den unterschiedlichsten Methoden die Illusion von Raum. Dabei soll der Gesamteindruck beim Betrachter immer eine gewisse Einheitlichkeit des Raumes erhalten.

Die Vorstellungen über diese Einheitlichkeit sind eng mit dem geschichtlichen und kulturellen Kontext verbunden. Ein ikonographischer religiöser Raum in der Grandval-Bibel aus dem 9. Jh [4] wird zu einem Erzählraum mit Vorder- und Hintergrund erweitert, wie in der Vertreibung aus dem Paradies von Masaccio [5]. Die Geschichte der perspektivischen Darstellung des Raumes kann vom Standpunkt betrachtet werden, dass der Bildraum „die Natur in sich aufnimmt und dem Mass und Anforderungen der Gesellschaft unterwirft“



Abb. 1: Grandval Bibel [4]

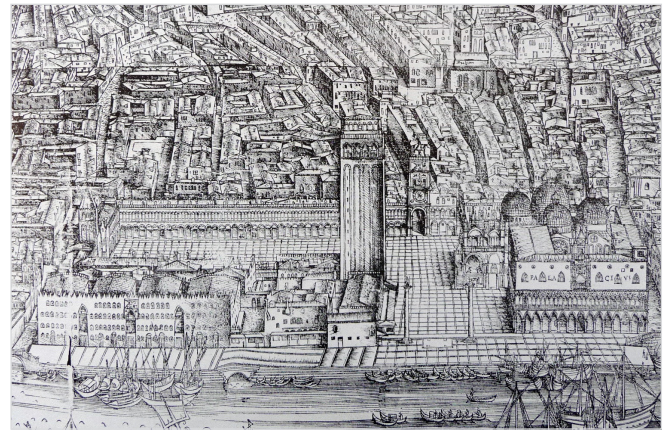


Abb. 2: Karte von Venedig [10]



Abb. 3: Deckenmalerei Kirche St. Ignazio in Rom [14]



Abb. 4: Dinamismo de um cão pela trela [7]



Abb. 5: Las Meninas after Velasquez [8]



Abb. 6: Las Meninas [6]

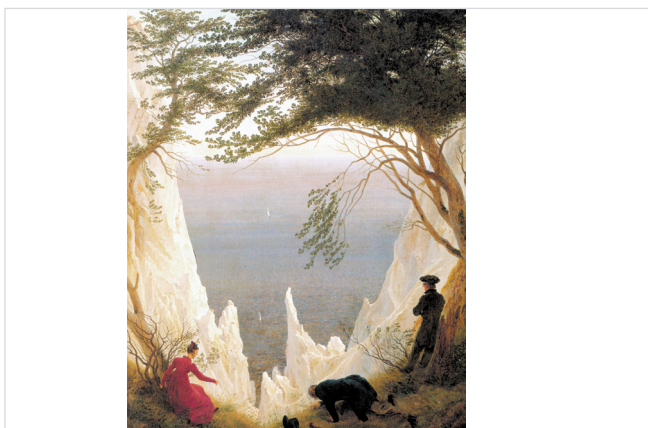


Abb. 7: Kreidefels auf Rügen [12]



Abb. 8: La Condition humaine [17]

[1]. Die Abbildungen des Raumes werden deshalb nicht nur von den technischen Möglichkeiten bestimmt, sondern auch von den geltenden Machtverhältnissen. Im Bild *Las Meninas* (1656) von Diego Velasquez [6] repräsentieren mehrere verschachtelte Fluchtpunkte die unterschiedlichen Ebenen der Macht – das Königspaar im Bild an der Wand erhält die eigene Perspektive.

Im Kubismus wird die Einheitlichkeit dadurch definiert, unterschiedliche räumliche Ansichten in einem Bildraum zu kombinieren und dadurch dem Betrachter einen gesamthafter Realitätsgewinn zu vermitteln: der Betrachter kann sich im Raum um das dargestellte Subjekt bewegen, vgl. *Las Meninas* von Pablo Picasso [8]. Ähnlich wird im Futurismus die Bewegung und damit das Zeitliche in einem Bildraum vereinheitlicht [7]. George Braques erzeugt im Bild *Violine und Krug* [9] eine taktile Wirkung durch eine scheinbar reliefartige Hervorhebung. Beim surrealistischen Maler René Magritte wird die semantische Korrektheit der Bildebene im Bild *La condition humaine* [17] durch bewusste Irreführung durchbrochen.

In meiner Arbeit entspricht die Bodenprojektion dem Bildraum. Die Perspektive der Inhalte richtet sich auf natürliche und direkte Weise nach der Position des Betrachters im realen Raum aus. Dieser gestaltet und betrachtet in gleichem Masse.

Neben dem Verhältnis von Betrachter und Bild ist die Ausrichtung des Blicks auf den Boden entscheidend. In der Kunstgeschichte findet zeitlich einige Zeit vor der französische Revolution ein Umbruch und „Loslösung vom allseitig begrenzten Raum“ [3] der Renaissance statt. Es verändern sich die Sehgewohnheiten indem der Blick in die Tiefe, Höher oder Ferne gelenkt wird. Nicht zuletzt durch die aufkommende Kartographie. Bekannt hierfür ist die Venedig-Karte von Jacopo de' Barbaris [10]. Die neue Raumauffassung und der Drang, neue Räume zu schaffen, widerspiegelt sich auch in den Deckenmalereien der Kirche St. Ignazio Rom [14]. Die Kirche ebnet und öffnet den Weg in den Himmel. Die Religions- und Literaturgeschichte hat sich schon immer klar nach oben orientiert [11]. Im Weltbild von Dante müssen die Seelen auf dem Weg nach oben zuerst die sieben Stufen des Fegefeuers überwinden um sich möglichst weit von der Hölle tief unten zu entfernen.

Mit dem Fortschritt der Industrialisierung löst sich der Blick und schweift freier herum. Der Blick „findet keinen Halt mehr“, „verliert den Bodenbezug“ [?]. Der „Raumsinn“ löst sich von der fixen horizontalen Perspektive. Es entsteht der „Blick von oben“, wie z.B. in Caspar Friedrichs *Kreidefels auf Rügen* [12]. Der Blick von oben hat eine aufklärerische Komponente. In den Weltkriegen wird das ausgenutzt in der ganzen Luftaufklärung [?].

Auch in der Geschichte der interaktiven Displays sind Veränderungen im Gange. Seit den 70er Jahren dominiert die Metapher des Fensters, der Blick hinaus, sowie der Blick auf den Schreibtisch. Mit den neuen Möglichkeiten portabler Projektionen oder mit portablen Displays können neue Sichtweisen geöffnet werden. In meiner Arbeit wird der Boden als Display aktiviert.

3. VERWANDTE ARBEITEN

Werke aus den unterschiedlichsten Bereichen zeitgenössischer Kunst und Technik beschäftigen sich mit den Themen von Raum, Projektionen, Perspektive, Tracking oder Gesten. In diesem Kapitel fasse ich meine Inspirationsquellen, ausserhalb der klassischen Kunstgeschichte, zusammen, und gebe einen Überblick über Arbeiten, welche sich mit ähnlichen Fragestellungen auseinandersetzen.

Medien, Kunst und Konzeptarbeiten

Zunächst stellt sich die Frage der Projektion von Inhalten in und auf den Raum eines Betrachters. Die Visual Jockeys (VJs) und die Mapping-Szene [15] beschäftigen sich allgemein gefasst mit der Projektion von Bildinhalten auf bestehende Strukturen im Raum. Die Projektion, als zweidimensionales Konstrukt, erhält durch die Reflektion im Raum eine dreidimensional erfahrbare Form.

Bei den *Augmented Sculptures* von Pablo Valbuena [13] entsteht durch die Projektion auf eine Anordnung von Kuben im Ausstellungsraum ein dreidimensionales Zusammenpiel von Licht, Schatten und Farbe. Das erzeugt die Illusion eines transformierten Raumes. In eine ähnliche Kategorie wie die Deckenmalereien der St. Ignazio gehören auch die *Pavement Drawings* des Strassenkünstlers Julian Beever [16]. Weitere Beispiele für Illusionen bezogen auf den Standpunkt des Betrachters finden sich unter der Bezeichnung *Forced-Perspective* auch in der zeitgenössischen Architektur oder der Werbung [18, 19, 20].

Im Gegensatz zu den perspektivischen Bildräumen bezieht sich das Konzept von Ivan Tihienko [21] mehr auf die Person und deren Interaktion mit ihrem Umraum. Er vermittelt die Idee einer persönlichen holografischen Projektion im urbanen Raum. Diese dient etwa der Steuerung von Geräten, welche der Benutzer trägt, oder als Navigationshilfe. Die Person steht optisch gesehen im Zentrum eines Bildes in Form eines Halbkreises.

Forschung und interaktive Displays

Eine Reihe von Forschungsprojekten beschäftigen sich mit der gezielten Positionierung von Projektionen im Raum und der Lokalisierung der damit verbundenen Interaktion. Beim *IBM Everywhere Display* [22] wird eine Projektion mittels motorisierter Steuerung eines Spiegels auf ausgewählte planare Oberflächen im Raum gelegt. Eine Kamera überwacht diesen Bereich durch die Erkennung von Bewegung (Motion Tracking). Die optische Erkennung der Fingerspitze dient der genauen Lokalisierung und Steuerung der Interaktion. Vorgeschlagene Anwendungsbeispiele sind Verkaufsläden oder Büroräume.

Eine ähnliche Richtung findet sich auch bei *Microsofts Play-Anywhere* [24] – wobei hier die Projektion hier auf die Fläche eines Schreibtisches beschränkt ist und der Fokus auf der Portabilität liegt. Auch hier werden die Interaktionen über die Bewegung erkannt. In der Dissertation von Daniel Cotting [23] wird eine Technik beschrieben, um in Innenräumen auf beliebigen Oberflächen positionierbare Displays zu generieren. Die dreidimensionalen Strukturen des Raumes werden über eine optische Hardware-Lösung mit der Hilfe von struk-

turiertem Licht erkannt. Dieses macht eine spezielle 3D-Kamera überflüssig.

Direkt auf den Körper einer Person und damit auch auf deren Position bezogene Projektionen sind der *Living Canvas* [25], wo das Bild in einer Lichtprojektion auf die weisse Kleidung von sich bewegendem Tänzern projiziert wird. Bei Sixth Sense [26] trägt die Person selber einen portablen Projektor, welcher Inhalte direkt in den Umraum oder auf den eigenen Körper projizieren kann.

Auf den Boden als Projektionsfläche fokussierte Produkte sind z.B. die interaktiven Bodenprojektionen von ArcStream [30] oder der interaktive Fussboden von Gesturetek [28]. Beide konzentrieren sich die Erkennung von Bewegung auf der Bodenfläche. Was sich dort bewegt, interagiert auch.

Im Bereich der interaktiven Bildschirme gibt es eine weitere Reihe von Produkten und Projekten. Diese können mehr oder weniger einfach mit den Händen gesteuert werden. Vertreter solcher Systems sind z.B. das Fraunhofer Institut mit dem i-Point 3d [27], GestureTek mit einer ganzen Serie von Produkten [28] oder einzelne Firmen wie Linden Labs mit dem SecondLife Body Controller [29]. Letzteres benutzt die Körperhaltung des Benutzers um den virtuellen Avatar zu steuern.

Microsoft Research beschreibt mit DepthTouch [31] ein interaktives Multitouch Display, welches ebenfalls mit einer 3D-Kamera arbeitet. Das Auswertungsprinzip basiert auf ähnlichen Grundlagen wie in dieser Arbeit beschrieben, ist aber auf eine fixe Oberfläche beschränkt. Über dieser können in einem begrenzten Raum, dreidimensionale Gesten durchgeführt werden.

Die Firma Kickerstudio hat für Canesta, einen der Hersteller von 3D-Kameras, eine gestische TV-Fernbedienung im Rahmen eines Interaction Design Prozesses entwickelt und beschreibt diesen Prozess in einer Fallstudie [32]. Darin wurden u.a. Benutzerstudien von Gesten durchgeführt. Die Autoren kommen zum Schluss, dass Gesten für die Steuerung eines Fernsehgerätes schnell emotional oder sozial interpretiert werden. Sie empfehlen deshalb eine Produktegestaltung basierend auf einfachen, kreisförmigen Handgesten, welche sie dementsprechend für Canesta umgesetzt haben.

Die recherchierten Bodenprojektionen fokussieren sich alle auf die Erkennung von Bewegung im Bodenraum bzw. in einem Raum, welcher von oben als Fläche mit einer Kamera überwacht wird. Deshalb ist nur eine sehr eingeschränkte und tendenziell eher unruhige Interaktion möglich, wie etwa das Aufdecken von Karten oder das eher symbolisch zu interpretierende Laufen über Wasser. Die mit den Händen steuerbaren Systeme wiederum ignorieren wiederum die Position des Körpers des Benutzers und sind im Prinzip primär als Spiegel zu bedienen. Interessant ist auf jeden Fall das DepthTouch-System von Microsoft, da hier auf inhaltlicher Ebene dreidimensionale Objekte zur Anwendung kommen.

4. KONZEPT

Der Gesturespace definiert einen Interaktionsraum zur experimentellen Erprobung der im Folgenden vorgestellten

Konzepte anhand verschiedener Beispielanwendungen. Er besteht aus drei Hauptkomponenten:

1. einer Projektion auf den Boden für die Abbildung eines virtuellen Bildraums
2. einer 3D-Kamera zur Erfassung der Interaktion von Körper und Händen des Benutzers
3. einer Anwendung, welche die Informationen der Kamera auswertet, interpretiert und die Ausgabe entsprechend gestaltet

Dieser Aufbau ist identisch mit dem eines klassischen Computersystems mit einer Von-Neumann-Architektur. Die Typisierung der Eingabe, Ausgabe und Verarbeitung ist derart gestaltet, dass kontinuierlich dreidimensionale Daten verwendet werden. Die Basiskoordinaten werden in einer mehrstufigen Pyramide als Hände, Körper und Gesten interpretiert.

Abbildung 1 illustriert den möglichen Aufbau eines solchen Systems für eine Ausstellung. Die Projektion wird über einen in flachem Winkel angelegten Projektionsspiegel umgeleitet, so dass der Schattenwurf von einer Person im Bildraum möglichst gering und einseitig erfolgt.

Die Kamera muss den Aktionsraum des Benutzers so überwachen können, dass sie Körper und Hände sehen kann. Für dieses Konzept und die Umsetzung gehe ich von einer einzelnen 3D-Videokamera aus. Diese liefert Distanzbilder des Raumes zurück. Das heisst für jeden einzelnen Punkt in einem zweidimensionalen Raster, wird eine ungefähre Distanz zurückgeliefert.

Die Basis des Interaktionsverlaufs ist eine abgekürzte Variante der von Dan Saffers beschriebenen *Three Zones of Engagement* [33]: Observation, Attraction, Interaction. Ein Benutzer bemerkt den Aktionsraum und begibt sich dort hinein. Der Aktionsraum ist als Bühne inszeniert. Sobald er in diesem Bereich steht, interagiert er bereits durch seine körperliche Präsenz mit dem System. Die Interaktion mit den Händen stellt schliesslich eine zweite Stufe der Interaktion dar. Hierdurch wird eine erweiterte Steuerung im Sinne von *Watch, Move, Interact* möglich. Man kann auch sagen, dass sich die Interaktion in eine eher orientierungsbezogene Primärphase

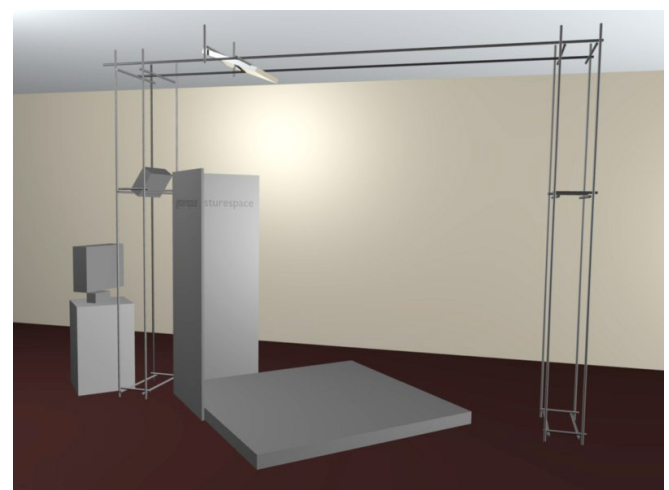


Abbildung 1. Mögliche Installation des Gesturespace

und eine steuerungsbezogene Sekundärphase aufteilt. Dieser Ablauf ist in Abbildung 2 dargestellt.

Der Begriff der *Geste* sei an dieser Stelle nochmals ausdrücklich definiert. In dieser Arbeit verstehe ich darunter eine *zeichenhafte Bewegung des Körpers oder bestimmter Körperteile*.

Das Auswertungs-Konzept gliedert sich in verschiedene Bereiche. Als Basis dient eine zwischen Körper und Hände gelegte, virtuelle Trennebene, welche sich der Position des Benutzers anpasst. Aufbauend auf dieser werden die Hände erkannt und in einem nächsten Schritt die Hände verfolgt (Hand Tracking). Dieses dient dann wiederum als Basis zur Erkennung von einfachen Gesten.

Dynamic Virtual Multitouch Layer

Die Grundlage für die spätere Erkennung von Gesten ist die zuverlässige Erkennung von Händen und Körper. Die niedrigauflösenden Punktwolke der 3D-Kamera, muss für die Verarbeitung auf semantisch höherer Ebene dementsprechend aufbereitet werden. Bisherige Systeme verwenden hier ausschliesslich die Erkennung von Bewegung (Motion Tracking), was die Interaktionsmöglichkeiten stark einschränkt, da z.B. Bewegungen des Körpers und der Hände nicht voneinander unterschieden werden können.

In dieser Arbeit wird ein einfaches, stabiles, fehlertolerantes und für die Verarbeitung in Echtzeit geeignetes Modell entwickelt, welches diese Beschränkungen aufhebt und so Körper und Hände separat, und von deren Bewegungen entkoppelt, verarbeiten kann.

Aufgrund der Distanzinformationen der Kamera wird, zu jedem Zeitpunkt der Messung, eine virtuelle Ebene in den Raum gelegt, welche Körper und Hände voneinander trennt. Dies ist vergleichbar mit einer berührungsempfindlichen Multitouch-Oberfläche, weshalb sie hier als *Dynamic Virtual Multitouch Layer* eingeführt wird.

Zunächst stellt sich die Frage, wie überhaupt ein einzelner Körper im Kamerabereich erkannt werden kann, damit dessen Position in den drei Raumdimensionen x, y und z bestimmt sind? Der Anspruch ist, dass die Erkennung jederzeit stabil funktioniert und gegebenenfalls zu einem späteren Zeitpunkt, auch auf mehrere Personen ausgedehnt werden kann. Abbildung 4 zeigt ein Distanz- oder Tiefenbild der Kamera mit einer Person im Blickfeld. Die Punktwolke wurde in ein Graustufenbild umgerechnet (je heller desto weiter vorne). Aufgrund des geringen Öffnungswinkels der Kamera (das

Field of View der verwendeten Swissranger 3000 3D-Kamera ist $47.5 \times 39.6^\circ$) entfällt die Möglichkeit einer zweidimensionalen Erkennung der Person rein aufgrund von Konturen, da die Person nie ganz im Bild zu sehen ist.

Die Verwendung der Distanzinformation der 3D-Kamera liegt nahe. Ein zu lösendes Problem besteht darin, dass diese Informationen ein starkes Rauschen aufweisen, stark distanzabhängig sind und bereits durch kleinste Körperbewegungen Abweichungen auftreten können. Ausserdem soll ein Schritt zurück oder zur Seite grundsätzlich unterscheidbar sein vom Ausstrecken einer Hand bei gleichbleibender Körperposition.

Das Dynamic Virtual Multitouch-Verfahren verwendet deshalb eine geeignete statistische Auswertung. Es definiert zunächst die Position des Körpers als die Position des Schwerpunktes der Punktwolke des sichtbaren Bereiches. Da die Kamera alles abschneidet, was über eine definierten Parameter der Distanz hinausgeht, erhält man dadurch eine, gegenüber äusseren Einflüssen, stabile Position des Schwerpunktes des Körpers. Die Haltungen der Hände und Arme haben, aufgrund ihrer kleinen Fläche im Bild, keinen nennenswerten Einfluss auf die Position des Körperschwerpunktes.

Die Position der dynamischen Ebene wird nun so berechnet, dass Körper und Hände optimal voneinander getrennt werden. Das Resultat sind zwei voneinander unabhängige zweidimensionale Bilder, welche Front- und Hintergrund bzw. die charakteristischen Eigenschaften von Händen und Körper darstellen.

Dies wird erreicht, indem die statistische Verteilung der Distanzen im Tiefenbuffer ausgewertet, und mittels *p-Quantilen* aufgeteilt wird. Durch die laufende Überprüfung dieser Aufteilung in der Praxis, konnte festgestellt werden, dass der Wert p mit der Gesamtdistanz variiert. Als gute Näherungsformel für die Trennebene hat sich deshalb ein Wert von $p = 4 * (z_{front} + 1)^2$ ergeben (das sind zwischen 1/4 und 1/30 der Verteilung). Der Wert von z_{front} wiederum entspricht der gemittelten Distanz des vordersten 1% der Verteilung (Perzentile).

Die Ebene des Körpers kann wiederum zahlenmässig über die letzten 10% der Verteilung berechnet werden – das sind die am weitesten entfernten Punkte. Abbildung 3 zeigt illustrativ das Vorgehen zur Aufteilung. Abbildung 5 zeigt exemplarisch die resultierenden Hand- und Körperebenen. Die Körperebene enthält hier störende Hintergrundelemente. Um diese herauszufiltern, wird die Kameraamplitude und Distanzmessung manuell auf ein bestimmtes Mass beschränkt.

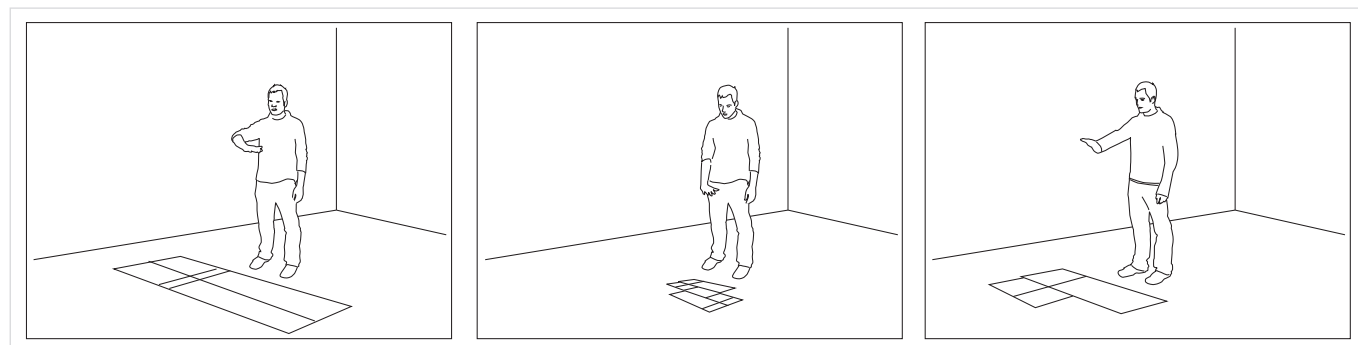


Abbildung 2. Aufmerksamkeit lenken, Interaktion mit dem Körper, Interaktion mit den Händen.

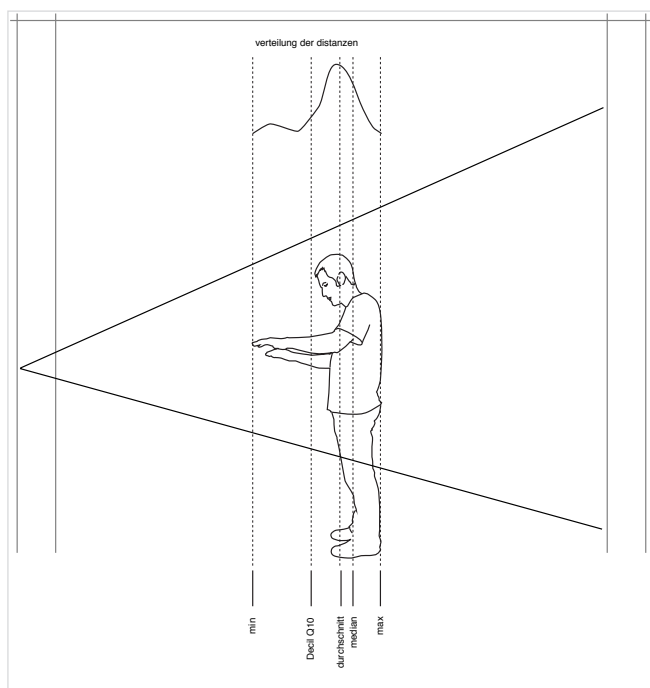


Abbildung 3: Verteilung der Distanzinformation bei ausgestreckten Händen (Handfläche sichtbar)

In einem letzten Schritt werden die aufgeteilten zweidimensionalen Informationen wieder auf Distanzen zurückgerechnet. Der genaue Schwerpunkt des Körpers ist nun bekannt.

Das praktische an diesem Ansatz ist nun, dass zu jedem einzelnen Zeitpunkt diese Aufteilung in die zwei Ebenen gewährleistet ist, unabhängig von der ablaufenden Bewegung. Sowohl Körper wie auch Hand können sich in allen drei Dimensionen unabhängig voneinander bewegen – die Trennebene wird immer dynamisch mitgeführt. Die erhaltenen 2D-Bildebenen können mit bekannten Algorithmen aus dem Bereich der Computer Vision aufbereitet und ausgewertet werden. Dies wird insbesondere für die Erkennung der Hände verwendet (s.u.). Die Bildebenen werden nach dem Splitting mittels einer Histogramm-Normalisierung optimiert.

Durch die Verwendung von Auswertungen statistischer Natur, ist das vorgestellte Verfahren nicht auf eine Form der Hände oder die Silhouette des Körpers fixiert, sondern reagiert auf beliebige Objekte, die in der Front-Ebene auftauchen. Auch mit der Bierflasche und dem Apérogebäck in der Hand oder dem Hut auf dem Kopf lässt sich das System noch steuern.

5. UMSETZUNG

Die Umsetzung des Konzeptes als Prototyp umfasst verschiedene Aspekte. Diese stehen in enger Abhängigkeit zueinander. So sind z.B. die Maximalgrösse der Projektion, die Genauigkeit und Reichweite der Kamera sowie der optische Öffnungswinkel direkt miteinander korreliert.

Randbedingungen Hardware, Software, Ausstellung

Die Basisinfrastruktur des Prototyps besteht aus einem Beamer Acer PD527W, einem Projektionsspiegel sowie einer 3D-Videokamera SR-3100 der Firma MESA Imaging. Diese konnte ich freundlicherweise als Leihgabe von der Computer Vision Group der ETH Zürich erhalten. Die Kamera sendet

codiertes Infrarotlicht aus, wertet die optischen Reflektionen aus und ermöglicht es so, einen definierten Raumbereich zu überwachen und ein Distanzbild in Form einer Punktwolke zu erhalten.

Im Gegensatz zu rein zweidimensionalen Kameralösungen hat dies den Vorteil, dass Entfernungen von Punkten direkt abgelesen werden können. Das Hand- und Körpertracking sind desweiteren nicht farbsensitiv und können mit beliebiger Kleidung und von Menschen beliebiger Hautfarbe bedient werden. Ebenso gibt es keine Probleme wegen schlechten Lichtverhältnissen – es funktioniert auch im Dunkeln – und es entfallen fehleranfällige Background-Subtraction Mechanismen. Je nach Parametrisierung und Fokussierung liefert die Kamera zwischen vier und 30 frames/s in einer QCIF-Auflösung von 176x144 Pixel.

Die technische Ausgangslage definiert die Möglichkeiten. Abbildung 4 zeigt eine von der Kamera erhaltene Punktwolke. Die Farben markieren die erkannten Distanzen. Das Rauschverhalten der Kamera ist abhängig von der eingestellten Genauigkeit, was aber einen negativen Einfluss auf die Framerate haben kann. Es kann auch passieren, dass Reflektionen ausserhalb des Bereiches fälschlicherweise mit aufgenommen werden. In verschiedenen Durchläufen wurden diese Parameter für den Prototyp optimiert.

Da der Prototyp an der Jahresausstellung 2009 der ZHDK ausgestellt wird, müssen die Masse der Installation für die Ausstellung geeignet sein. In der Vorphase habe ich deshalb zunächst eine Analyse der Möglichkeiten der involvierten Geräte vorgenommen, verschiedene Beamer evaluiert, Berechnungen angestellt, Experimente aufgebaut und Entscheidungen getroffen:

1. Die Projektion an der Ausstellung erfolgt auf eine 1.8m x 1.8m grosse quadratische Fläche. Die maximale Höhe des Projektionsspiegels ist die Raumhöhe im Gang des Schulhauses, was ca. 3.40 m entspricht.
2. Der Beamer wird um 90° gedreht, um das 4:3 Bild bis zu den Besuchern zu bringen.
3. Die Messdistanz der Kamera wird auf einen Kreis von 2.6m beschränkt und die Genauigkeit so optimiert, dass ein 3D-Bild mit etwa 17 frames/s generiert wird.
4. Die 3D-Kamera wird leicht von unten schauend montiert, so dass die Handfläche in den meisten Fällen möglichst gross ist.



Abbildung 4: Aus der Distanzinformation berechnetes 2D Graustufenbild und 3D Punktwolke



Abbildung 5: Distanzbild (Graustufen 2D), Hintergrundebene (Körper), Vordergrundebene (Hände)

Abbildung 2 zeigt ein Rendering der Ausstellungsplanung aus der Vorphase. Für die Auswertungssoftware steht ein separater iMac-Computer zur Verfügung.

Als Software- und Entwicklungsplattform habe ich mich für Windows und Microsoft Visual C++ entschieden, da momentan nur für diese Kombination aktuelle Treiber und Beispielsoftware der 3D-Kamera verfügbar sind. Von Max Rheiner erhielt ich zum Start ein Basis-OpenGL Framework zur Verarbeitung und Darstellung von 3D-Daten. Alle 2D Computer Vision Auswertungen (siehe unten) werden über die weitverbreitete OpenCV Bibliothek [34] erledigt.

Handerkennung

Das Auswertungskonzept liefert eine separate Handebene und eine Körperebene. Die vordere, der Kamera zugewandte Handebene, kann viele Elemente enthalten, z.B. den Kopf des Benutzers oder Teile der Beine. Der Teil der Handerkennung beschäftigt sich damit, in der Handebene möglichst zuverlässig eine einzelne, oder zwei unabhängige Hände zu erkennen.

Hierfür durchläuft das Handbild (in Graustufen umgewandelt) zuerst einige Optimierungen zur Verminderung des Rauschens. Danach folgt ein Thresholding mit anschließender Konturerkennung. Die Schwierigkeit hierbei ist, dass je nach Lage der Hand-Ebene eine einzelne Hand aus mehreren Teilbereichen, im Extremfall inkl. Arm, bestehen kann. Diese Einzelteile müssen dann wieder als ganze Hand erkannt werden, resp. die einzelnen Teile der richtigen Hand, von zweien, zugeordnet werden.

Die gewonnen Konturen werden zuerst gefiltert (zu klein, zu gross) und danach mittels einem k-Means Clustering-Verfahren in zwei Cluster aufgeteilt. Auf diesen Clustern werden dann wiederum die Schwerpunkte ausgerechnet. Nahe zusammenliegende Schwerpunkte werden zur gleichen Hand gezählt. Mit diesem Verfahren kann sowohl die Anzahl Hände, wie auch deren Position weitgehend stabil für einen einzelnen Zeitpunkt ermittelt werden.

Verfolgung der Hände

Die Handerkennung liefert eine mögliche Position von Händen für einen ganz bestimmten, einzelnen Zeitpunkt. Fehlerhafte Erkennung aufgrund starken Rauschens oder aufgrund von Körperbewegungen, muss auf andere Weise ausgeglichen werden. Schaut der Kopf auf den Boden, kann es sein, dass die Stirne ebenfalls bis in Handebene hinein reicht, was zu einer physisch nicht vorhandenen Hand führt. Die Stabili-

tät der Handerkennung über Zeit wird durch die Verfolgung der Hände (Hand-Tracking) gewährleistet. Die Hände dürfen wackeln oder kurzfristig verschwinden, ebenso muss ein sinnvoller Umgang mit temporären drei Händen gefunden werden.

Um diese Probleme zu lösen, wird ein zeitliches Vertrauensintervall eingeführt. Eine neue Hand muss zuerst eine bestimmte Zeit präsent sein, um als gültige Hand erkannt zu werden. Im Gegenzug darf sie kurzzeitig verschwinden und danach wieder erscheinen.

Bei zwei Händen stellt sich zusätzlich das Problem der über die Zeit konsistenten Zuordnung. Insbesondere bei Überkreuzungen oder schnellen Bewegungen kann das System die Hände vertauschen. Deshalb werden die neuen und vorangehenden Handpositionen eines Zeitraums verglichen und jeweils der nächstliegenden Hand zugeordnet. Dies limitiert zwar die Gesamtgeschwindigkeit der Handbewegung, erlaubt aber stabile Überkreuzungen.

Alle erhalten Werte für Körper und Hände durchlaufen eine einstellbare zeitliche Glättung wodurch Wackeln weitgehend vermieden wird.

5. GESTEN

Aufbauend auf den Ergebnissen aus den verschiedenen Stufen von Erkennung und Tracking von Körper und Händen, kann eine Erkennung von Gesten folgen.

Für die Auswahl und Einschränkung möglicher Gesten und die Umsetzung der Gesten-Erkennung für den Prototyp, habe ich mich von folgenden Grundsätzen leiten lassen:

1. *Einfachheit*: wie die Studie von KickerStudio [32] zeigt, werden Gesten schnell als sozial offensiv oder aggressiv wahrgenommen. Einfache, eher gleichförmige Bewegungen scheinen sich in der Praxis besser zu bewähren.
2. *Überprüfbarkeit*: es muss im Prototyp nachvollziehbar sein, wieso eine bestimmte Geste verarbeitet wurde, und wieso nicht. Nach Möglichkeit sollen sich Gesten nicht gegenseitig ausschliessen oder aufheben.
3. *Umsetzbarkeit*: die Erkennung der Gesten soll stabil sein und so schnell und einfach wie möglich erfolgen.
4. *Intuition*: soweit möglich sollen die Gesten vom Benutzer intuitiv erfasst werden könne, ohne dass eine spezi-

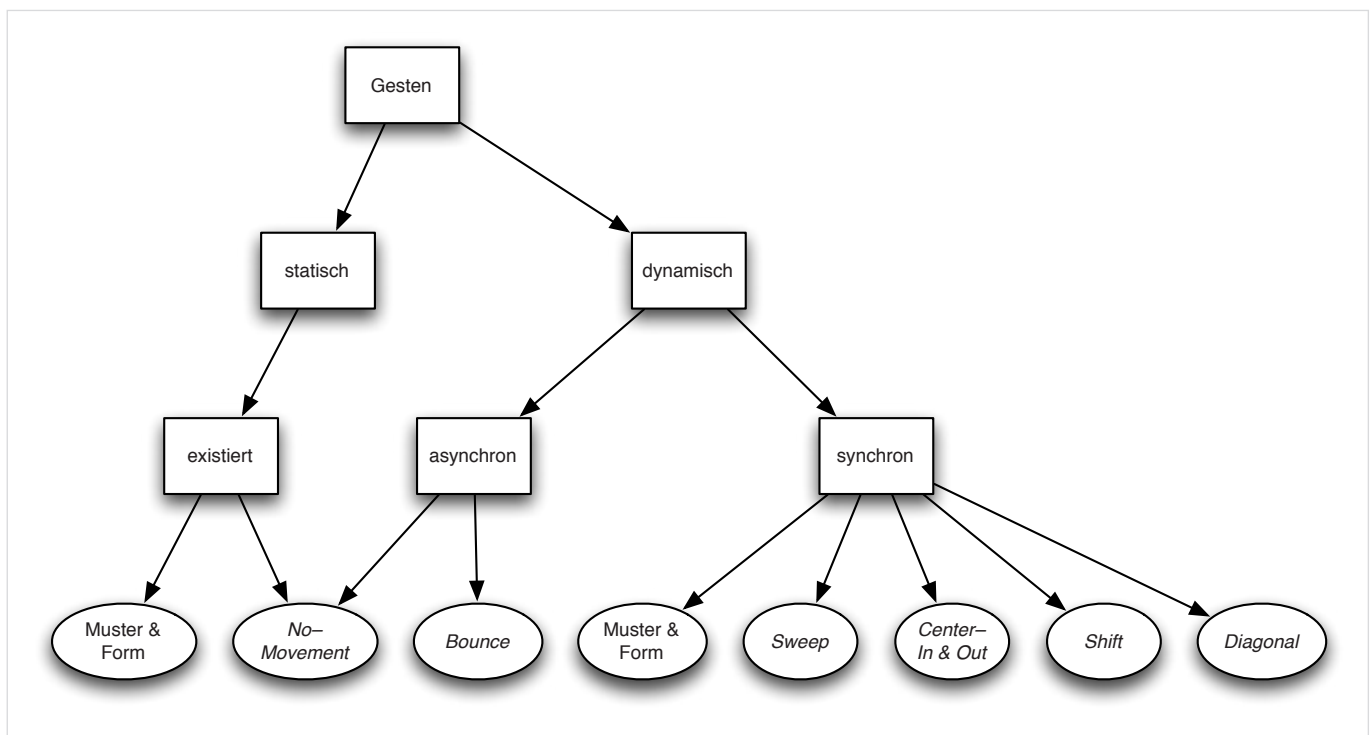


Abbildung 6: Klassifizierung der verschiedenen Gestenformen

elle Anleitung zwingend nötig wird. Dies bedingt eine entsprechende Gestaltung der Anwendung.

Diese Grundsätze führen zu einem einfachen Gestenalphabet für den Prototyp (vgl. Abbildung 6). Diese können als Grundlage für die weitere Erforschung gestischer Steuerung von Anwendungen dienen. Eine illustrative Darstellung der verschiedenen Gestenformen findet sich in Abbildung 7.

Gestenalphabet

In der Literatur [33] wird zwischen *statischen* und *dynamischen Gesten* unterschieden. Statische Gesten sind Symbole, welche aufgrund ihres Bildes resp. ihrer Postur erkannt werden können – sie sind entweder vorhanden, oder nicht – es gibt keinen zeitlichen Ablauf. Ein Beispiel für eine statische Geste ist die ausgestreckte Handfläche, welche als *Stop* interpretiert wird, oder die Symbole der Zeichensprache.

Die einfachsten statischen Gesten in diesem System sind im Folgenden aufgeführt. Da sie gleichzeitig Grundlage für die Erkennung komplexerer dynamischer Gesten sind, werden sie hier als *Trivialgesten* bezeichnet. Als solche gelten:

1. Körper vorhanden, Körper nicht vorhanden
2. Hand vorhanden, Hand nicht vorhanden (jeweils separat für beide Hände)

Für alle statischen Gesten ist zusätzlich die Information über die Position im Raum (x, y, z) bekannt.

Der nächste Schritt ist die Erweiterung dieser Trivialgesten über die Zeit:

1. Körper steht still (wenig Bewegung über die Zeit)
2. Hand wird still gehalten

Diese gelten streng genommen bereits als *dynamische Gesten*, obwohl hier explizit keine Bewegung gefordert wird. Das Grundproblem hier ist dass der Benutzer zunächst einmal nicht weiss, wie lange er die Hand still halten muss, damit eine Aktion ausgelöst wird. Umgekehrt kann eine natürliche Handposition dazu führen, dass ungewollt eine Aktion ausgelöst wird. Für die Umsetzung dieser *No Movement*-Gesten, muss die Anwendung entsprechende visuelle oder akustische Hinweise zum zeitlichen Stand der Geste bringen. Beispielsweise ein Klang, welcher in der Intensität zunimmt, oder das zunehmende visuelle Pulsieren, oder die Grössenveränderung eines Objektes.

Für die weitere Unterscheidung von Gesten, reichen die Begriffe statisch und dynamisch nicht aus. Für die Steuerung einer Anwendung ist es wichtig, zu welchem Zeitpunkt eine Aktion effektiv ausgeführt wird. Eine Handbewegung von links nach rechts könnte bereits noch während der Bewegung zu einer nachvollziehbaren Aktion führen, z.B. der Drehung eines 3D-Modells. Umgekehrt kann eine Geste definiert sein durch eine Handbewegung von links nach rechts und wieder zurück. Diese kann dann erst bei der Rückwärtsbewegung als gültige Geste interpretiert werden. Um diese Fälle verbal zu trennen, führe ich den Begriff der *Synchrongeste* ein. Eine Synchrongeste führt bereits während ihrer Bewegung zur Ausführung einer Aktion in der Anwendung.

Synchrongesten sind das direkte Gegenstück der *No Movement*-Gesten. Nur wenn sich die Hand bewegt, wird eine Aktion ausgeführt. Für die prototypische Umsetzung wird die Erkennung der Bewegung auf die drei Hauptachsen x,y,z eingeschränkt. Um diese Gestenform von natürlichen Handbewegungen zu unterscheiden, muss die Hand mit einer definierten Mindestgeschwindigkeit und Mindeststrecke in eine Richtung bewegt werden. Dieser Typus von Gesten wird hier

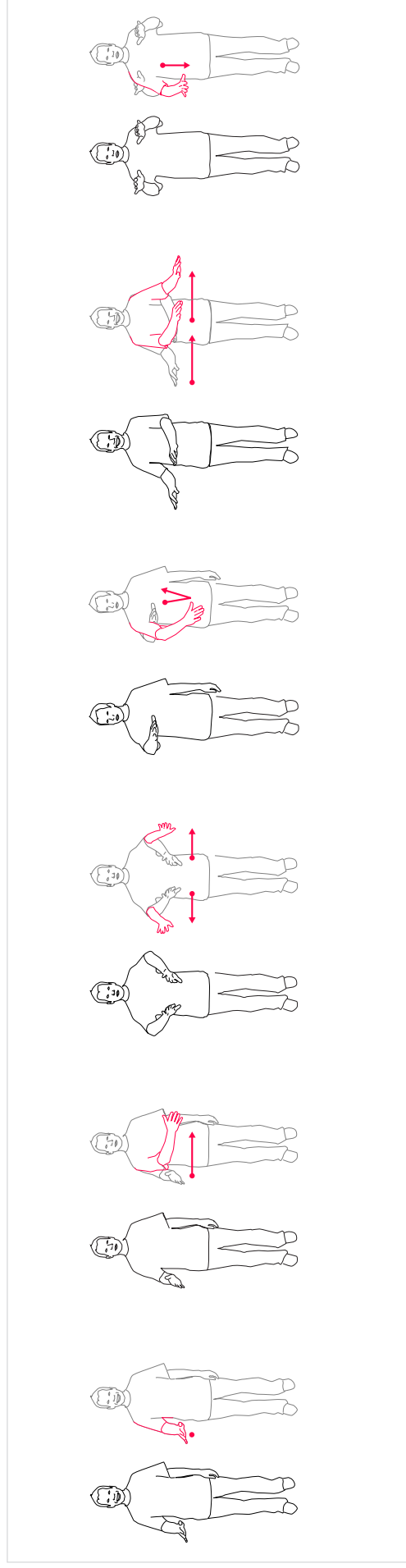


Abbildung 7. Illustration der wichtigsten Klassen des Gestenalphabets. Von links nach rechts: *No-Movement, Sweep, Center-Out, Bounce, Shift*.

als *Sweep-Geste* bezeichnet, da die Bewegung im Horizontalen wie ein natürliches Wischen aussieht.

Die Schwierigkeit der Interpretation von Sweep-Gesten liegt darin, dass wir aufgrund unseres Körperbaus, unserer Muskulatur, meistens eine anschliessende Gegenbewegung durchführen (Reflex). Dies kann man leicht nachvollziehen, wenn man versucht, einen virtuellen Knopf in der Luft zu drücken – die Hand schnell wieder ein wenig nach oben, da der erwartete Widerstand nicht auftritt. Zur reinen Erkennung von Sweep-Gesten, wird deshalb ein Zeitintervall definiert, in welchem die Gegenbewegung auftreten darf, ohne dass eine Gegenaktion ausgelöst wird.

Die Kombination zweier Sweep-Gesten durch die Gegenbewegung bezeichne ich als *Bounce-Geste*, da die Hand wie an ihren Ursprung zurückspringt. Der Knopfdruck ist eine solche Geste.

Wenn man sich die räumliche Situation veranschaulicht, so ist jede Bounce-Geste eigentlich elliptischer Natur. Eine Hand legt nie eine gerade Strecke zurück sondern fährt immer auf einer elliptischen Bahn. Die von KickerStudio beschriebenen Kreisgesten [32] können dann als Bounce-Gesten verstanden werden, und umgekehrt. Bounce-Gesten sind mehrstufige Prozesse und arbeiten mit Statuszuständen, vergleichbar mit einem Doppelklick bei der Maus.

Nach den einfachen Gesten einer einzelnen Hand, können wir Gesten betrachten, welche mit zwei Händen parallel ausgeführt werden. Diese wären z.B.

- *Center-Out*: von der Mitte ausgehend, ziehen beide Hände gleichmässig nach aussen
- *Center-In*: von aussen kommend stossen beide Hände gleichmässig in die Mitte
- *Linear Shift*: beide Hände in gleichmässigem Tempo und parallel nach oben, unten, rechts oder links
- *Diagonal*: beide Hände in diagonal unterschiedliche Richtungen, also z.B. links hoch, rechts runter

Diese hier beschriebenen einfachen Gesten lassen sich als Baukasten oder Basisalphabet für komplexere Gesten einsetzen. Sie stehen also für einzelne Zeichen, welche sich zu bedeutsamen Wörtern und Symbolen kombinieren lassen. Die Wörter lassen sich dann zur Gestensprache kombinieren.

Umsetzung Gesten-Erkennung

Für die Umsetzung der Gesten-Erkennung im Prototyp war die einfache Umsetzung massgebend. In der Theorie existieren zahlreiche Modelle und Softwarepakete zur Erkennung von Abläufen und Mustern. Diese umfassen z.B. selbstlernende Hidden-Markov-Ketten oder Neuronale Netzwerke.

Sämtliche hier vorgestellten Gesten wurden mittels einfachen, nachvollziehbaren linearen Algorithmen und Finite State Machines (FSM) implementiert. Gerade durch diese Einschränkung, können gezielte Aussagen zum Timing und der Wirksamkeit von Gesten gemacht werden.

Allerdings würde deren Verwendung Grundsatz 2 und 3 missachten und schlichtweg den zeitlichen Rahmen dieser Arbeit bei weitem sprengen. Gerade aber diese Einschränkung er-

möglicht auch, genaue Aussagen über Parametrisierung und Timings zu machen. Das Timing aller Gesten ist stark parametrisiert und die Werte konnten auf allen Ebenen laufend experimentell optimiert werden.

6. PROTOTYP UND APPLIKATIONEN

Der entwickelte Prototyp setzt alle erwähnten Konzepte in die Praxis um, und wurde während der Entwicklungsphase intensiv in einem iterativen Designverfahren zur kontinuierlichen Evaluation eingesetzt. Zur Demonstration der verschiedenen Arten von definierten Gesten wurden verschiedene Applikationen entwickelt:

1. *Bälle*: demonstriert das statische Ein- und Zweihand-Tracking, inkl. der Geschwindigkeits- und Beschleunigungsmessung. Der Benutzer kann einen oder zwei abstrakten Bälle mit der Hand durch das Feld prellen.
2. *Luftbild-Viewer*: demonstriert die synchronen horizontalen Sweep-Gesten. Der Inhalt der Bilder bezieht sich auf den kunstgeschichtlichen Bezug des *Blick von oben*.
3. *3D-Modell Viewer für Architektur*: demonstriert das Körper-Tracking und die kombinierten Synchron-Gesten (Zoom-In, Zoom-Out, Shift Up/Down, Shift Left/Right).

Die Applikationen werden mittels einer übergeordneten Menüstruktur zusammengehalten. Die Schwierigkeit dabei ist, dass das Menü aus jeder Applikation mit einer gleichartigen Geste aufgerufen werden kann. Zunächst wurde hierfür eine Geste auf komplexerer Ebene definiert: der *Double-Sweep*. Dieser entspricht einem zweifachen horizontalen Sweep, ähnlich einem Wegwischen der dargestellten Inhalte. In der Praxis hat sich aber herausgestellt, dass eine so elementare Funktion wie der Menüaufruf, über eine intuitivere Geste abgebildet werden muss. Deshalb wird nun im Prototyp eine vertikale Bounce-Geste verwendet – auf *Knopfdruck* erscheint das Menü.

7. ERGEBNISSE UND DISKUSSION

Der Gesturespace ist eine Installation zur Erforschung von Gesten zur Steuerung interaktiver Anwendungen. Er kombiniert das Prinzip einer auf den Betrachter bezogenen immersiven, perspektivischen Projektion, mit einer körperhaften, gestischen Interaktion des Benutzers im Raum. Dabei ist eine gleichzeitige Interaktion von Körper und Händen möglich. Die Arbeit wurde von der Basis auf mit State-of-the-Art Technologie konzipiert und iterativ in einem Prototyp umgesetzt. Er ermöglicht erstmals eine gezielte systematische Erforschung von Gesten im freien Raum. Die erzielten Beiträge für den Bereich des Interaction Designs gliedern sich in die folgenden Bereiche:

- *Virtual Dynamic Multitouch Layer*: Entwicklung eines neuartigen Konzeptes zur stabilen und gleichzeitigen Auswertung und Tracking von Körper- und Handgesten im freien Raum. Dieses ist simpel, stabil und effizient. Es werden keine Hilfsmittel (Marker) auf Seiten des Benutzers benötigt.

- *Aktivierung des Bodens* als interaktives immersives Medium für dreidimensionale Inhalte.
- *Definition eines Gestenalphabets* zur einfachen Erstellung von Anwendung und als Basis für inhaltlich komplexe Gestenformen von Händen und Körper.
- *Bau eines erfahrbaren Prototyps*, welcher alle Aspekte kombiniert und eine gezielte Evaluation und Erforschung der Gesten anhand einfacher illustrativer Anwendungen erlaubt.

Im Experiment zeigt sich, dass durch die Projektion auf den Boden und durch den Einbezug dreidimensionaler Inhalte ein hoher Grad an Immersion erreicht werden kann. Ein vergleichbares System mit dieser Art von Tracking und Display konnte in der Recherche nicht gefunden werden.

Diese Arbeit ist als initiales Fundament für die experimentelle Erforschung von Gesten zu verstehen. Die Resultate zeigen, dass:

1. die Art und Weise der Interaktion und Projektion in den verschiedenen Experimenten grundsätzlich als natürlich wahrgenommen wird
2. je einfacher die ausgeführte Geste ist, desto natürlicher die Interaktion erscheint
3. das optische und/oder akustische Feedback viele Ungenauigkeiten im Tracking oder der Erkennung von Gesten ausgleicht, sofern die Gesten nicht zu komplex sind
4. komplexe Gesten immer der Erklärung bedürfen
5. die visuelle Gestaltung der Anwendung die als intuitiv empfundenen Gesten direkt und stark beeinflusst.

Insbesondere die Applikation des architektonischen Viewers funktioniert aufgrund der Direktheit der Interaktion hervorragend. Die Steuerung des Blickpunktes aufgrund der Körperposition ist intuitiv. Die Gesten des Center-In und Center-Outs sind bereits von Multitouch-Applikationen als Zoom-In und Zoom-Out bekannt und benötigen nur wenig Erklärungen.

Aus der Erfahrung mit der Entwicklung des Prototyps lässt sich sagen, dass die genaue Parametrisierung und die Abstimmung des Timings der einzelnen Gesten, absolut entscheidend für ein angenehmes Gesamterlebnis sind.

8. AUSBLICK

Im jetzigen Prototyp ist das System auf eine Person und eine relativ kleine Fläche beschränkt. Das Prinzip lässt sich aber durch die Verwendung mehrerer 3D-Kameras auch auf mehrere Personen und grössere Flächen erweitern. Die Projektionsgrösse selbst ist abhängig von der direkten oder indirekten Distanz des Projektors von der Bodenfläche. Auch hier liessen sich mehrere Projektoren kombinieren.

Das Tracking und die Erkennung von Bewegung könnte man noch weiter justieren. Zur Rauschreduktion wäre der Einsatz der neuen Kamerageneration (SR-4000) und weiterer Software-Filter sinnvoll. Das Gestenalphabet könnte durch den Einsatz von State-of-the-Art Mustererkennung erweitert

und generalisiert werden. Ebenso kann eine Erkennung von statischen Gesten eingebaut werden.

Abschliessend kann festgehalten werden, dass die Schwierigkeiten bei der Gestaltung gestischer Schnittstellen primär darin liegen, wie sie dem Benutzer intuitiv erfahrbar gemacht werden können. Die Frage, was eine intuitive Geste auszeichnet, hängt zudem stark von der zu steuernden Anwendung ab. Mit dieser Arbeit wird eine Türe geöffnet für ein breites Feld an Forschung im Bereich der körpergestenbasierten Interaktion.

DANK

An die Mentoren meiner BA-Arbeit: Max, Christian, Gerhard. Der Computer Vision Group der ETH Zürich (Prof. Dr. Marc Pollefeys für die Leihgabe der 3D-Kamera). Jonas, Michael, Tobias – 4. Semester Interaction Design für die Mithilfe bei der Gestaltung der Ausstellung und das Kameragehäuse. Gabi für alles andere.

REFERENZEN

1. Levèbre, Henri (1991): The Production of Space. Oxford (UK), Cambridge. ISBN 0-63118177-6.
2. Rothko, Mark (2005): Die Wirklichkeit des Künstlers: Texte zur Malerei. 1. Auflage, Beck, München. ISBN 9783406528804.
3. Christa Maar, Hubert Burda (2006): Iconic Worlds: Neue Bilderwelten und Wissensräume. Bewegliche Fluchtpunkte: Der Blick von oben und die moderne Raumanschauung. Dumont Buchverlag. ISBN 3-83217995-X.
4. Grandval-Bibel (um 840 n. Chr.). Ausgestellt in der British Library, London. Informationen Online verfügbar unter: <http://www.moutier.ch/new/navig/general/histoire/biblemg.htm>, zuletzt geprüft am 9.5.2009.
5. Masaccio (15. Jh): The Expulsion from the Garden of Eden (Bildmaterial). Online einsehbar unter [http://en.wikipedia.org/wiki/The_Expulsion_from_the_Garden_of_Eden_\(Masaccio\)](http://en.wikipedia.org/wiki/The_Expulsion_from_the_Garden_of_Eden_(Masaccio)), zuletzt geprüft am 9.5.2009.
6. Diego Velasquez (16. Jh): Las meninas (Bildmaterial). Online einsehbar unter http://de.wikipedia.org/wiki/Las_Meninas, zuletzt geprüft am 9.5.2009.
7. Giacomo Balla (1912): Dinamismo de um cão pela trela. Albright-Knox Art Gallery, Buffalo, New York. Online einsehbar z.B. unter http://www.jerryandmartha.com/yourdailyart/uploaded_images/Balla-705712.jpg, zuletzt geprüft am 9.5.2009.
8. Picasso, Pablo (1957): Las Meninas (after Velasquez). Museo Picasso, Barcelona. Online einsehbar z.B. unter <http://www.artchive.com/artchive/P/picasso/meninas.jpg.html>, zuletzt geprüft am 9.5.2009.
9. Braques, George (1910): Violine und Krug. Kunstmuseum Basel. Online verfügbar z.B. unter <http://www.unterricht.kunstbrowser.de/images/braquegeigeundkrug1910originalgro.jpg>, zuletzt geprüft am 9.5.2009.

10. De Barbari, Jacopo (um 1500): Karte von Venedig. Online verfügbar unter http://commons.wikimedia.org/wiki/File:DeBarbari_pducale.JPG, zuletzt geprüft am 9.5.2009.
11. Wertheim, Margaret (2000): Die Himmelstür zum Cyberspace. Ammann, Zürich. ISBN 3250104175.
12. Friedrich, Caspar David (1818): Kreidefels auf Rügen. Museum Oskar Reinhart am Stadtgarten, Winterthur. Online einsehbar unter http://de.wikipedia.org/wiki/Kreidefelsenauf_R%C3%BCgen, zuletzt geprüft am 9.5.2009.
13. Valbuena, Pablo (2007): Augmented sculpture series. Ars Electronica. Online verfügbar unter <http://www.pablo-valbuena.com>, zuletzt geprüft am 9.5.2009.
14. RomaCulta (2001): Die illusionistische Deckenmalerei von Sant'Ignazio in Rom. Andrea Pozzo, um 1690. Online verfügbar unter <http://www.romaculta.it/det/decke-ignazio-pozzo.html>, zuletzt geprüft am 2.4.2009.
15. Mapping Festival 2009. Online verfügbar unter <http://www.mappingfestival.com>, zuletzt geprüft am 9.5.2009.
16. Beever, Julian (2008): Pavement drawings. Online verfügbar unter <http://users.skynet.be/J.Beever/pave.htm>, zuletzt geprüft am 9.5.2009.
17. Magritte, René (1933): La Condition humaine. National Gallery of Art, Washington USA. Online verfügbar unter http://www.nga.gov/cgi-bin/timage_f?object=70170&image=17567&c=, zuletzt geprüft am 9.5.2009.
18. Clean Cut Media (2008): Distorted Typography. Verfügbar online unter <http://blog.cleancutmedia.com/art-design/creativity-distorted-typography>, zuletzt geprüft am 17.5.2009.
19. Break-Thru Designs: Forced Perspective Graphics. Verfügbar online unter <http://www.breakthrudesigns.com/concrete-graphics-decals.html>, zuletzt geprüft am 17.5.2009.
20. 2Loop (2006): 3D painted rooms. Verfügbar online unter <http://www.2loop.com/3drooms.html>, zuletzt geprüft am 17.5.2009.
21. Tihienko, Ivan (2008): Holographic Interface – Round interface – Ringo. Bezalel Academy of Arts and Design, Jerusalem. Online verfügbar unter <http://www.vimeo.com/1416530>, zuletzt geprüft am 9.5.2009.
22. Rick Kjeldsen, Anthony Levas, Claudio Pinhanez (2003): Dynamically Reconfigurable Vision-Based User Interfaces. In: 3rd International Conference on Vision Systems (ICVS'03). Graz, Austria. Online verfügbar unter <http://www.research.ibm.com/ed/publications/icvs03.pdf>, zuletzt geprüft am 2.4.2009.
23. Cotting, Daniel (2007): Smart displays in interactive visual workspaces. Diss. ETH Nr. 17346. Online verfügbar unter <http://e-collection.ethbib.ethz.ch/view/eth:29957>, zuletzt geprüft am 2.4.2009.
24. Wilson, Andrew D. (2005): PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System. Online verfügbar unter <http://research.microsoft.com/en-us/um/people/awilson/papers/wilson%20playanywhere%20uist%202005.pdf>, zuletzt geprüft am 9.5.2009.
25. Naef, M. und Boyd, C. (2008): Feasibility of the living canvas: restricting projection to a performer on stage. In: Proceedings of the 16th ACM international Conference on Multimedia (Vancouver, British Columbia, Canada, October 26 - 31, 2008). MM '08. ACM, New York, NY, 589-598. Online verfügbar unter doi:10.1145/1459359.1459438, zuletzt geprüft am 2.4.2009.
26. Mistry P., Maes P., Chang L. WUW - Wear Ur World - A Wearable Gestural Interface. In: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems. Online verfügbar unter doi: 10.1145/1520340.1520626, zuletzt geprüft am 9.5.2009.
27. Fraunhofer Heinrich-Hertz-Institut (2009): iPoint 3D. Online verfügbar unter <http://www.hhi.fraunhofer.de/de/departments/interactive-media-human-factors/overview/ipoint3d/>, zuletzt geprüft am 9.5.2009.
28. GestureTek (2009): Cube. Online verfügbar unter <http://www.gesturetek.com/cube/introduction.php>, zuletzt geprüft am 9.5.2009.
29. Linden Labs (2008): Hands Free 3D (Video). Online verfügbar unter http://www.youtube.com/watch?v=KqwUn_KgrDQ, zuletzt geprüft am 9.5.2009.
30. ArcStream (2008): Living Image. Online verfügbar unter: <http://www.arcstreamav.com/living-image/>, zuletzt geprüft am 9.5.2009.
31. Benko Hrvoje, Wilson Andrew (2009): DepthTouch: Using Depth-Sensing Camera to Enable Freehand Interactions On and Above the Interactive Surface. Online verfügbar unter <http://research.microsoft.com/apps/pubs/default.aspx?id=79848>, zuletzt geprüft am 10.5.2009.
32. Kicker Studio (2009): Case Study: Gestural Entertainment Center for Canesta. Online verfügbar unter <http://www.kickerstudio.com/blog/2009/03/case-study-gestural-entertainment-center-for-canesta/>, zuletzt geprüft am 2.4.2009.
33. Saffer, Dan (2008): Designing Gestural Interfaces. O'Reilly Media, Inc. ISBN 0596518390.
34. OpenCV: Open Source Computer Vision library. Online verfügbar unter: <http://opencv.willowgarage.com/>, zuletzt geprüft am 9.5.2009.