



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Masterarbeit

Christian Fischer

Entwicklung eines multimodalen
Interaktionssystems für computergestützte
Umgebungen

Christian Fischer
Entwicklung eines multimodalen
Interaktionssystems für computergestützte
Umgebungen

Masterarbeit eingereicht im Rahmen der Masterprüfung
im Studiengang Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer : Prof. Dr. rer. nat. Kai v. Luck
Zweitgutachter : Prof. Dr. rer. nat. Gunter Klemke

Abgegeben am 22. November 2007

Christian Fischer

Thema der Masterarbeit

Entwicklung eines multimodalen Interaktionssystems für computergestützte Umgebungen

Stichworte

Multimodale Interaktion, Computer Supported Collaborative Work, Mensch-Maschine-Interaktion, Ubiquitäres Computing

Kurzzusammenfassung

Um die Zusammenarbeit einer Personengruppe in Hinblick auf Effizienz und Ergonomie zu steigern, wird eine Arbeitsumgebung mit technischen Systemen ausgestattet. Die Integration dieser Systeme in das alltägliche Umfeld geschieht für den Benutzer transparent, er nimmt sie nicht aktiv wahr. Ein Gesichtspunkt ist die intuitive Handhabung des Systems: was nicht erlernt werden muss, wird einfach benutzt.

Basierend auf diesem Leitgedanken wird in dieser Arbeit die Idee einer geeigneten Interaktion mit einer solchen Umgebung vorgestellt. Im Speziellen soll die direkte Manipulation von Objekten auf Anzeigeflächen durch die Kombination von Zeigegestik und Sprache erreicht werden.

Bei der praktischen Entwicklung wird die Zeigehandlung in den Vordergrund gestellt und ein neuartiger Ansatz für die technische Umsetzung konzipiert. Die Entwicklung eines Prototypen belegt die Plausibilität des Konzeptes und offeriert das Potential dieser neuen Eingabemöglichkeit.

Christian Fischer

Title of the master thesis

Development of a multimodal interaction system for computer supported environments

Keywords

multimodal interaction, computer supported collaborative work, human computer interaction, ubiquitous computing

Abstract

The performance enhancement of collaboration groups is possible by integrating computational technology in their every day working environment. It has to be realized in a transparent way, the user may not be disturbed by the technology. One aspect is the intuitive handling of the whole environment: "If something has not to be learned, it will just be used"

Based on that idea in this thesis a suited interaction technique for collaborative workspaces will be introduced. In particular the direct manipulation of objects on display areas shall be realized via the combination of pointing gesture and speech. The practical work concentrates on the evaluation of a new approach to realize the pointing gesture. The development of a prototype proves the plausibility of that concept and demonstrates the potential of the new input device.

Inhaltsverzeichnis

1. Einführung	7
1.1. Motivation	7
1.2. Ziel dieser Arbeit	8
1.3. Gliederung der Arbeit	8
2. Grundlagen	10
2.1. Mensch-Maschine-Interaktion	10
2.2. Multimodale Interaktion	12
2.3. Computergestützte Gruppenarbeit	14
3. Analyse	17
3.1. Szenario	17
3.1.1. Beschreibung	17
3.1.2. Szenario-Komponenten	18
3.1.3. Konkretisierung: Interaktion	19
3.1.4. Klassifizierung	21
3.2. Anforderung an Interaktion	23
3.3. Festlegung der Modalitäten	27
3.4. Überblick vorhandener Technologien	30
3.4.1. Optisches, statisches Tracking	30
3.4.2. Kamera als Eingabegerät	33
3.4.3. Interaktion durch Bewegung	35
3.4.4. Interaktion durch Sprache	36
3.5. Auswertung	37
4. Entwicklung eines Zeigegerätes	40
4.1. Der interaktive Raum	40
4.2. Logischer Entwurf	43
4.2.1. Controller: Technologische Ebene	47
4.3. Machbarkeitsstudie	49
4.3.1. Vorgehensweise	50
4.3.2. Hardware und Entwicklungsumgebung	50
4.3.3. Systemaufbau	52

4.3.4. Auswertung	58
4.4. Übertragbarkeit auf Standardgeräte	61
4.4.1. Beschreibung	62
4.4.2. Aufgabenstellung und Umsetzung	63
4.4.3. Auswertung	66
4.5. Optimierung	66
4.5.1. Bestimmung der Zeigeposition	66
4.5.2. Lokalisierung	71
4.5.3. Portierung in ein Multi-Monitor-System	73
4.6. Auswertung	74
5. Schluss	77
5.1. Zusammenfassung	77
5.2. Ausblick	78
5.2.1. Weitere Einsatzmöglichkeiten des Zeigestocks	78
5.2.2. Verbesserung der Zeigehandlung	80
5.2.3. Weitere Ideen	82
Literaturverzeichnis	83
A. Anhang	89
A.1. Photogrammetrie	89

1. Einführung

1.1. Motivation

Im September 1991 prägte [Weiser \(1991\)](#) in seinem Aufsatz: „The Computer for the 21st Century“ den Begriff „Ubiquitous Computing“ und damit die Idee der allgegenwärtigen Computertechnik, die den Menschen unmerklich zu jeder Zeit bei seiner natürlichen Handlung unterstützt. Hierbei muss eine Integration von Technik in alltägliche Handlungen so nahtlos erfolgen, dass dem Nutzer die Interaktion mit einem Computer nicht bewusst wird ([Slay und Thomas \(2006\)](#)).

Die vielfältigen Möglichkeiten der Interaktion mit dem System über verschiedene Anzeigekomponenten und Eingabemöglichkeiten sollen für den Benutzer besonders im Arbeitsumfeld von Bedeutung sein und effizienzsteigernd wirken. Abhängig vom Anwendungsfall kann die dafür am Besten geeignete Technologie ausgewählt werden, die Verwendung des Arbeitsplatzes ist nicht auf statische und herkömmliche Konzepte (z.B. Maus und Tastatur am klassischen Bildschirmarbeitsplatz) beschränkt. Dienste sollen zur richtigen Zeit am richtigen Ort in der richtigen Form für den vorgesehenen Benutzer und Zweck verfügbar sein (Bill Buxton in [Paley u. a. \(1998\)](#)).

Das UbiComp-Projekt der HAW Hamburg entspringt diesem Forschungsgebiet und befasst sich in zahlreichen Themengebieten mit der Fragestellung nach Unterstützbarkeit alltäglicher Handlungen durch Computertechnologie. Aktuell liegt der Fokus mit „Ambient Intelligence“ auf der Entwicklung einer intelligenten Umgebung.

Das hierfür gewählte Szenario ist die Umsetzung eines Gruppenarbeitsplatzes (Computer Supported Collaborative Workspace - CSCW). Ziel dieses Arbeitsumfeldes ist die Steigerung der Effektivität und Effizienz von Teamarbeit durch den Einsatz von Informations- und Kommunikationstechnologie.

Die Umsetzung dieser Idee erfordert eine Anreicherung der Umgebung mit neuartigen, dem Konzept entsprechenden Interaktionsmöglichkeiten. Bedingt durch die Heterogenität des Arbeitsbereiches in Bezug auf die verschiedenen Interaktionsbereiche ist ein nahtloser Übergang von einem Systems zum nächsten aus ergonomischer und technologischer Sicht zwingend notwendig („Seamless Interaction“).

Der Benutzer rückt bei den Betrachtungen in den Vordergrund (Ishii u. a. (1994)), der Computer ist ein Instrument – hieraus ergibt sich die Basis für ein Interaktionskonzept: Der Ursprung muss in der Natur liegen – die Wechselwirkung mit der Umgebung soll intuitiv geschehen, die Entwicklung orientiert sich an dem natürlichen bzw. herkömmlichen (bereits erlernten und damit als natürlich empfundenen) Verhalten.

1.2. Ziel dieser Arbeit

In dem oben aufgeführten Forschungsgebiet ist die zentrale Fragestellung: Wie wird Computertechnologie unsichtbar?

Inhalt dieser Arbeit ist der Themenbereich Interaktion und demnach die Frage: Wie kann die Interaktion mit einem Computersystem stattfinden, ohne dass dem Benutzer offensichtlich bewusst ist, dass er mit einem Computer kommuniziert?

Analysiert werden die Anforderungen an notwendige Interaktionsmöglichkeiten in einer computergestützten Umgebung. Darauf und auf aktuelle Arbeiten basierend wird ein geeignetes Konzept erstellt, um die Interaktion nach dem Ansatz des „Ubiquitous Computing“ zu ermöglichen. Angelehnt an die natürliche Kommunikation und Interaktion zwischen Menschen liegt der Schwerpunkt auf der multimodalen Interaktion mit der Prämisse „Intuitivität“.

Praktisches Ziel dieser Arbeit ist die Validierung dieses Interaktionskonzeptes durch die Entwicklung eines prototypischen Eingabesystems.

1.3. Gliederung der Arbeit

Kapitel 2 vermittelt Grundlagen zu Forschungsbereichen in denen sich das UbiComp-Projekt der HAW Hamburg und damit diese Arbeit bewegt. Es werden die historische Entwicklung der Mensch-Maschine-Interaktion (Kapitel 2.1) und Ansätze für eine neue Generation der Interaktion (Kapitel 2.2) dargestellt.

Im Analysekapitel 3 wird das Grobkonzept der zu entwickelnden Interaktionskomponente erarbeitet. Aus dem Szenario (3.1) gehen die Rahmenbedingungen der Anwendung hervor, aus denen sich die Anforderungen an das Eingabegerät erheben (Kapitel 3.2) und genauer spezifizieren (Kapitel 3.3) lassen. Mit der Recherche nach bestehenden Interaktionskonzepten wird das neuartige Konzept festgelegt.

Dieser theoretische Ansatz wird in Kapitel 4 umgesetzt. Dazu gehört die Prüfung auf Plausibilität in der Machbarkeitsstudie (4.3) und die Weiterentwicklung zu einem Prototyp. Nach

der konzeptionellen Erstellung des Entwurfs vom Gesamtsystem (Kapitel 4.2) erfolgt die Prüfung auf Plausibilität in Form einer Machbarkeitsstudie (Kapitel 4.3). Durch Optimierung der Technik (4.4) und der Algorithmen (4.5) wird der Prototyp erstellt.

Das Schlusskapitel (5) gibt einen Überblick über die erfolgten Schritte dieser Arbeit. Der Abschnitt 5.2 offeriert Fragestellungen, mit denen Ansatzmöglichkeiten für weiterführende Arbeiten bestehen.

2. Grundlagen

2.1. Mensch-Maschine-Interaktion

1963 hat Ivan Sutherland im Rahmen seiner Doktorarbeit am MIT¹ das Grafikprogramm „Sketchpad“ (vgl. [Sutherland \(1964\)](#)) entwickelt (siehe Abbildung 2.1). Es gilt als eine der

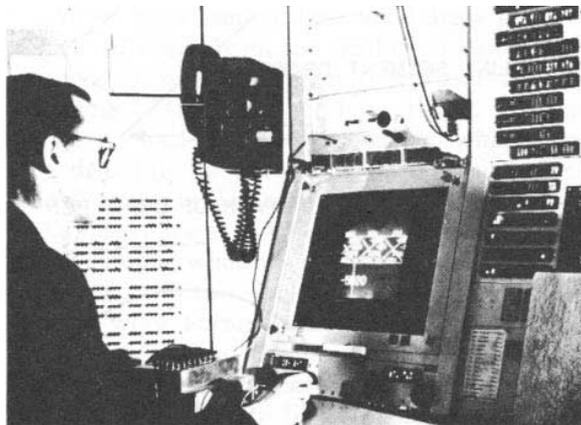


Abbildung 2.1.: Sketchpad

ersten interaktiven Grafikanwendungen und setzte den Startschuss frei für die praktische Mensch-Maschine-Interaktion. Großen Anklang fand die Technologie der grafischen Interaktion im Bereich der CAD/CAM-Systeme, zusammen mit Lichtgriffeln konnten grafische Objekte direkt am CRT-Monitor manipuliert werden. Zeitnah waren viele Entwicklungen in diese Richtung zu erkennen: Kurz zuvor schrieb [Licklider \(1960\)](#), ein weiterer Pionier der Informationstechnologie, den Artikel „Man-Computer-Symbiosis“, in welchem die Notwendigkeit der Interaktion zwischen Computer und Computernutzer hervorgehoben wurde. Ebenfalls 1963 veröffentlichte D. C. Engelbart (Erfinder der Computermaus) „Augmentation of Human Intellect“, eine bahnbrechende Arbeit über die Erweiterung von Intelligenz durch die Entwicklung von Netzwerken. Dem folgte eine Reihe von Entwicklungen, darunter z.B. das „Dynabook“.

¹MIT: Das Massachusetts Institute of Technology ist eine weltweit führende Universität in Cambridge im Bereich der technologischen Forschung und Lehre

Die Idee von Xerox PARC, als Hauptverfechter Alan Kay, war eine intuitive Bedienung, verbunden mit einer leistungsfähigen Hardware, die grafisch anspruchsvolle Anwendungen unterstützt. Kay stellte bei der Entwicklung die Orientierung an menschlichen Bedürfnissen in den Vordergrund.

Dass die Mensch-Computer-Interaktion in den Anfängen nicht auf die Bedürfnisse jedes beliebigen Benutzers angepasst wurde, stellte ebenfalls [van Dam \(1997\)](#) fest. Bis in die 1980er wurde der Fokus der Entwicklung hauptsächlich auf Funktionalität und Performanz gelegt. Mit der enormen Popularität der Desktop-Programme bei unerfahrenen Benutzern gingen die Anforderungen an die Benutzbarkeit dieser Tools vielmehr in Richtung „einfache Bedienbarkeit“ als ausschließlich Effizienz, die Entwicklung fand nun in zwei konträren Richtungen statt. „Einfache Bedienbarkeit“ bezeichnet hier, wie schnell der Umgang mit dem System erlernbar ist und wie gut man sich diesen merken kann.

So klassifiziert van Dam in seinem Artikel "Post-WIMP User Interfaces" vier stilistisch und rückblickend auf die bisherige Entwicklung chronologisch voneinander unterscheidbare Interaktionsschemata: Während in der ersten Epoche (1950er bis 1960er) Computer noch im Stapelbetrieb liefen, Eingaben mit Lochkarten und Ausgaben per Zeilendruckern funktionierten, somit also keine, nach heutigem Verständnis bekannte Benutzerinteraktion zu Stande kommen konnte, erlaubte die zweite Epoche (1960er bis 1980er) die kommandobasierende Interaktion und Mehrbenutzerfähigkeit der Recheneinheit nach dem Timesharing-Verfahren auf Mainframes. In den 1970ern wurde die nächste, bis heute anhaltende Generation des Xerox Palo Alto Research Center ins Leben gerufen. Umgesetzt wurden grafik- sowie netzwerkbasierende Arbeitsstationen und vor allem grafische Benutzeroberflächen nach dem WIMP-Prinzip (windows, icons, menus, pointing devices), das erste System war der „Xerox Alto Computer“, beschrieben durch [Wadlow \(1981\)](#). Die Entwicklung ist bis heute selbstverständlich weiter fortgeschritten, jedoch mehr was den Realismus der Interaktionswidgets angeht, als Veränderungen des Konzepts als solches.

[van Dam \(1997\)](#) beschreibt dies als Schwäche und fordert die Notwendigkeit einer vierten Generation der Mensch-Computer-Interaktion. Mit der Komplexität des Funktionsumfangs einer Anwendung steigt der Umfang der Benutzerschnittstelle nichtlinear. Das bedeutet, einzelne Funktionen oder Widgets einer Anwendung sind separat leicht zu handhaben, die Anhäufung dieser Komponenten in einem Programm führt jedoch zur oben beschriebener Komplexität.

Im Gegensatz zur herkömmlichen Oberflächenbedienung sollen idealerweise alle vorhandenen Sinne und die natürliche Sprache eingebunden und zusätzlich die Gruppenkommunikation gefördert werden.

2.2. Multimodale Interaktion

Multimodale Interaktion steht hier für die Interaktion zwischen Mensch und Computer, wobei die klassische sehr begrenzte Interaktion um die Verwendung möglichst vieler, dem Menschen zur Verfügung stehende Sinne erweitert wird. Aus sinnesphysiologischer Sicht gibt es sieben Modalitäten: Sehen, Hören, taktile Wahrnehmung, Gleichgewichtssinn, die Tiefensensibilität und die chemischen Sinne (Riechen und Schmecken).

Das Forschungsgebiet der multimodalen Interaktion leitet die, wie im vorherigen Kapitel geforderte, vierte Generation der Mensch-Maschine-Interaktion ein. Beispielhaft für solche Entwicklung neuer Strategien ist das Leitprojekt „SmartKom²“ des Bundesforschungsministeriums. Unter der Koordination des DFKI in Saarbrücken arbeiten Wissenschaftler und Experten aus der Wirtschaft fachübergreifend in den Bereichen (vgl. [Wahlster \(WS 2000/2001\)](#)):

Mustererkennung Sprach-, Bild-, Gestik, und Mimikerkennung, Stifteingabe, Handschrifterkennung

Musterverstehen und Semiotik Sprach-, Bild-, Gestik, und Mimikverstehen, Kommunikationswissenschaft Dialogforschung

Wissensverarbeitung Intentionserkennung, Präsentationsplanung, Graphik- und Sprachgenerierung

Kognitionswissenschaft Benutzermodellierung, Evaluation

Software und Applikationsengineering Systemdesign und Systemintegration, Mobile Systeme

Projektmanagement Planung, Steuerung, Organisation

Das Projekt hat zur Aufgabe die „Dialogische Mensch-Maschinen-Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten“.

SmartKom kritisiert bei der bisherigen Entwicklung von Interaktionskonzepten zwischen Mensch und Technik die Orientierung nach dem Produkt. Da sich der Benutzer dem System anpassen bzw. den Umgang damit erlernen muss, bauen sich Hemmungen auf, die dazu führen können, dass die Funktionalität des System nur teilweise oder gar nicht erschlossen wird.

Die Benutzerschnittstelle soll durch Förderung der Nutzerfreundlichkeit besonders Computernutzer die Interaktion mit Technik erleichtern. Wie bei [van Dam \(1997\)](#) (beschrieben in Kapitel 2.1) tritt hierbei der Benutzer in den Mittelpunkt der Betrachtungen (Nutzerzentrierung). Die resultierende Schnittstelle lässt sich durch zwei Eigenschaften beschreiben:

²SmartKom: <http://www.smartkom.org>

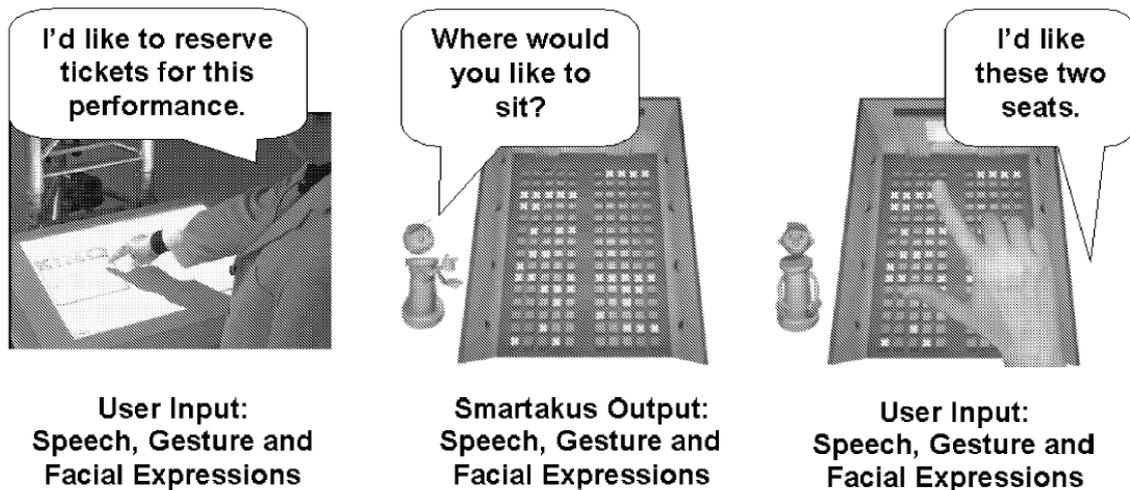


Abbildung 2.2.: Beispieldialog in SmartKom

Selbsterklärend Durch den Einfluss der menschlichen Sinne wie Sprache, Gestik, Mimik und Augenbewegungen und der Synergie dieser natürlichen Eingabemodalitäten bildet sich eine intuitive Schnittstelle.

Benutzeradaptiv Das System passt sich dem einzelnen Benutzer individuell an. SmartKom setzt hierfür Benutzermodellierung, Planerkennung und Lernverfahren ein, wodurch eine Eigendynamik des Systems möglich wird. Kontextabhängig können unaufgefordert Dialoge stattfinden (Beispieldialog s. Abb. 2.2) und fehlerhafte, unvollständige und gestörte Eingaben sinnvoll interpretiert werden.

SmartKom beschreibt folgenden Mehrwert dieses Konzeptes gegenüber herkömmlichen Interaktionssystemen:

- Nahtlose Integration und koordinierte semantische Verarbeitung sich wechselseitig ergänzender Eingabemodalitäten wie Sprache, Gestik/Mimik, Stifteingabe/Graphik sowie Biometrie.
- Robuste Verarbeitung von möglicherweise ungenauer, mehrdeutiger oder teilweise inkorrekt eingetragener Eingabe.
- Kontextsensitive Interpretation der dialogischen Interaktion auf dem Hintergrund dynamischer Diskurs-, Aufgaben- und Benutzermodelle.
- Adaptive Generierung von koordinierten, multimodalen Präsentationen.
- Halb- oder vollautomatische Komplettierung von delegierten Aufgaben durch Integration von Informationsdiensten.

- Intuitive Erscheinung der Schnittstelle als personalisierter Präsentationsagent.
- Umfangreiche und intuitive Hilfe- und Erklärungsfunktionen im Dialogkontext sowie zahlreiche metakommunikative Funktionen.
- Adaptive Steuerung der Dialoges durch benutzer- und situationsangepaßte Interaktionsstile.

Abbildung 2.3 zeigt die Grobarchitektur von SmartKom.

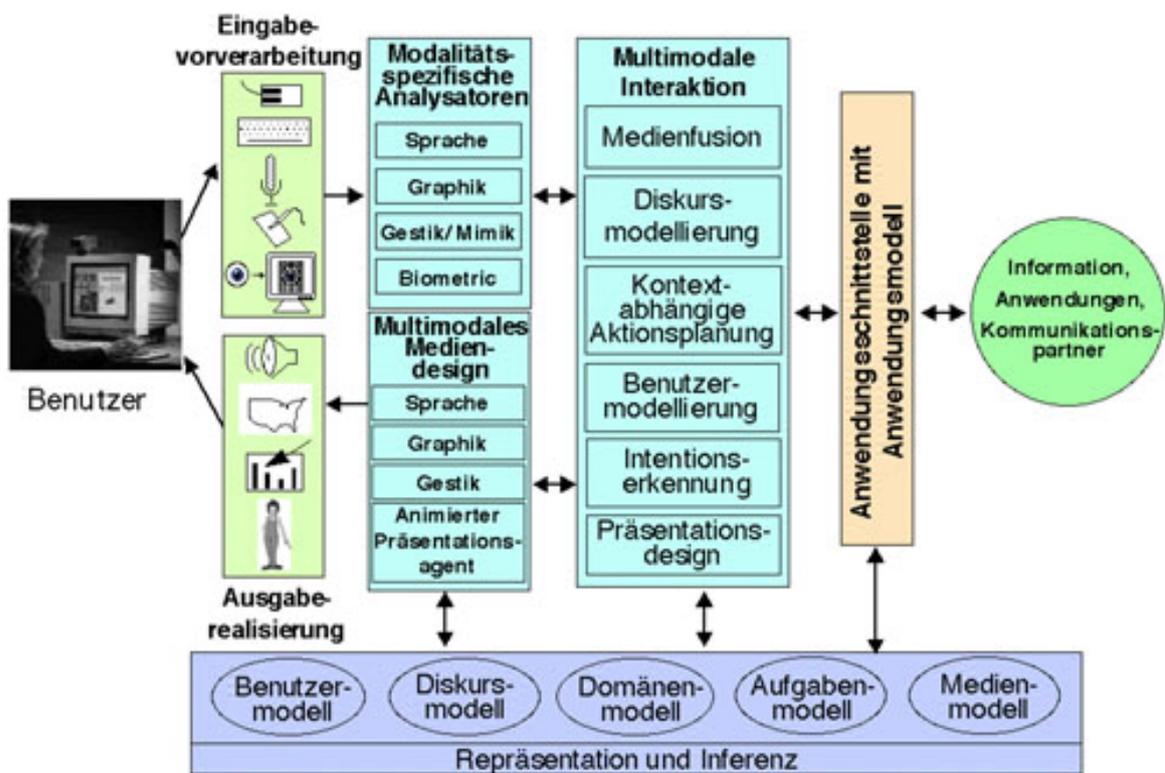


Abbildung 2.3.: Grobarchitektur von SmartKom

2.3. Computergestützte Gruppenarbeit

Das aktuell verwendete Szenario des in Kapitel 1.1 beschriebenen UbiComp-Projektes an der HAW Hamburg ordnet sich im Forschungsgebiet „Ubiquitous Computing“ ein. Spezieller Fokus wird dabei auf Szenarien gerichtet, die sich in Arbeitsumgebungen abspielen. Mehrere Personen arbeiten miteinander und werden durch die technischen Möglichkeiten in dem

Raum unterstützt. Diese spezielle Art des „Ubiquitous Computing„ ist ein eigenständiges Forschungsgebiet und heisst „Computer Supported Collaborative Workspace“(CSCW).

Computer Supported Collaborative Work ist kein in sich abgegrenztes Thema, sondern ein Wissenschaftsfeld, es umfasst neben weiteren Disziplinen Informatik, Soziologie, Psychologie, Arbeits- und Organisationswissenschaften, Management, Anthropologie, Wirtschaftsinformatik, Wirtschaftswissenschaften und Medienwissenschaft. Der Fokus der HAW Hamburg beschränkt sich auf den Bereich der Informatik und verfolgt, gemäß [Burger \(1997\)](#), als Ziel die Kooperation zwischen Menschen und deren Unterstützbarkeit durch Rechner.

Geprägt wird der Begriff CSCW 1984 durch Irene Greif (MIT, jetzt bei Lotus) und Paul Cashman (Digital). Sie organisieren einen Workshop³, der sich mit den theoretischen Grundlagen und Methodologien für Gruppenarbeit und deren Unterstützung durch Rechner auseinandersetzt ([Schlichter u. a. \(2001\)](#)). Die erste Konferenz findet 1986 (CSCW'86) statt. Eine allgemeingültige Definition von CSCW gibt es nicht. [Wilson \(1991\)](#) schreibt:

„CSCW is a generic term which combines the understanding of the way people work in groups with the enabling technologies of computer networking, and associated hardware, software, services and techniques.“

Selbst der Ansatz von CSCW ist nicht eindeutig: [Schlichter u. a. \(2001\)](#) verwenden die Abkürzung CSCW als Akronym. Eine Vorwärtsanalyse betrachtet den Begriff aus Sicht der Informatik:

- C** Ausgangspunkt ist der Rechner
- S** Er soll als Unterstützungsmedium eingesetzt werden.
- C** Bisher beschränkte sich die Unterstützung auf Individualumgebungen. Neue technische Möglichkeiten legen nun eine Auseinandersetzung mit Unterstützungsformen für kooperative Aktivitäten nahe.
- W** Die zu bewältigende Arbeitsaufgabe (sie steht bei dieser Betrachtungsweise meist hinten an).

Nach [Möslein \(1993\)](#) wird bei der Rückwärtsanalyse vom organisationstheoretischen Hintergrund ausgegangen, ergänzend erfolgt die Betrachtung sowie Einbindung von Kooperation und Unterstützung durch Rechner:

- W** Im Zentrum der Behandlung von CSCW steht die zu bewältigende Arbeit selbst.
- C** Die Bewältigung der Arbeitsaufgaben aber erfolgt in der Regel arbeitsteilig und damit im Zusammenspiel mehrerer Kooperationspartner.
- S** Diese Aufgabenbewältigung ist zu unterstützen.

³CSCW Workshop 1984 im Endicott House, Massachusetts

- C** Dafür sind insbesondere Möglichkeiten des Rechnereinsatzes zu berücksichtigen und weiterzuentwickeln.

Im Rahmen des UbiComp-Projektes wird dem Ansatz der Rückwärtsanalyse nachgegangen, obwohl die Betrachtung aus Sicht der Informatik geschieht. Diese Sichtweise entspricht der des SmartKom-Projektes (vgl. 2.2), die Frage ist nicht „Wie kann man den Rechner am Besten verwenden?“, sondern „Wie können Aufgaben am besten mit IT-Background bewältigt werden?“.

Das nachfolgende Szenario beschreibt die Optimierung bei der Handhabung bzw. dem Ausführen von Aktivitäten in einer gemeinschaftlichen Arbeitsumgebung, die mit Mitteln der Informatik unterstützt wird.

Vollständigkeitshalber ist es an dieser Stelle notwendig, den von [Johnson-Lenz \(1981\)](#) beschriebenen Begriff „Groupware“ als „An intentional group process plus software to support them.“ zu benennen. Groupware bezeichnet Systeme, denen die theoretischen Grundlagen und Ausarbeitungen gemäß CSCW zugrunde liegen. Eine Beschreibung von Groupware liefert [Johansen \(1988\)](#):

„Groupware is a generic term for specialised computer aids that are designed for the use of collaborative work groups. Typically, these groups are small project-oriented teams that have important tasks and tight deadlines. Groupware can involve software, hardware, services and/or group process support.“

In den beiden Beschreibungen von computergestützter Kollaboration von Gruppen sowohl mit- als auch untereinander werden die notwendigen Bereiche definiert, über die sich die Bildung eines CSCW bzw. Groupware erstrecken sollte: Kommunikation, Hardware, Software und Dienste bzw. Services. Es ist also notwendig, dass die Konzipierung einer Interaktionskomponente übergreifend in allen drei Bereichen stattfinden muss, hierzu siehe Entwurfskapitel 4.2.

3. Analyse

3.1. Szenario

Für das Forschungsgebiet Collaborative Workspace lassen sich diverse Anwendungsfälle generieren. Im Rahmen der Entwicklung eines multimodalen Interaktionssystems ist die Beschränkung des Verwendungs- bzw. Anwendungsfreiraums unumgänglich.

3.1.1. Beschreibung

Der Themenrahmen „Collaborative Workspace“ des UbiComp-Projekt der HAW Hamburg definiert ein aktuelles Szenario, das Rescue-Szenario. Aktuell beschäftigt sich eine Forschungsgruppe mit den notwendigen Bestandteilen für diesen Anwendungsfall. Schauplatz ist beispielhaft der Frankfurter Flughafen, der anwendungsspezifische Schwerpunkt liegt auf der Überwachung und Koordinierung von Einsätzen Rettungskräfte in Desastersituationen, wie z.B. einem Grossbrand und der anschliessenden Brandbekämpfung durch die Werksfeuerwehr.

Eine derart konkrete Umgebung lässt die Definition von Randbedingungen zu, hier aus [Pie-ning \(2007\)](#):

1. Das Gelände ist bekannt, endlich und überschaubar.
2. Es existieren detaillierte Karten der Umgebung und der Gebäude.
3. Der Flughafen wird von einer Werksfeuerwehr betreut, die mit einem üppigen finanziellen Etat ausgestattet ist.
4. Die Installation von Sensorik auf dem Gelände und Vernetzung von Komponenten ist möglich.
5. Ein stationärer Leitstand kann auf dem Gelände eingerichtet und für die Anwendungsfälle des Flughafens optimiert werden.

3.1.2. Szenario-Komponenten

Das Szenario erfordert eine Aufteilung in Arbeitspakete, aktuell sind diese (vgl. Abbildung 3.1):

- Ad-hoc Sensornetzwerk zur Gebäudeüberwachung und Navigation, [Davids \(2007\)](#)
- Wearable Computer in Disaster Szenarien, [Hinck \(2007\)](#)
- Leitstand für Disaster Szenarien, [Piening \(2007\)](#)
- Darüberhinaus beschäftigt sich [Möller-Freuck \(2007\)](#) mit dem Thema Kartendarstellung in den Bereichen Feuerwehr- und Polizeieinsätze.

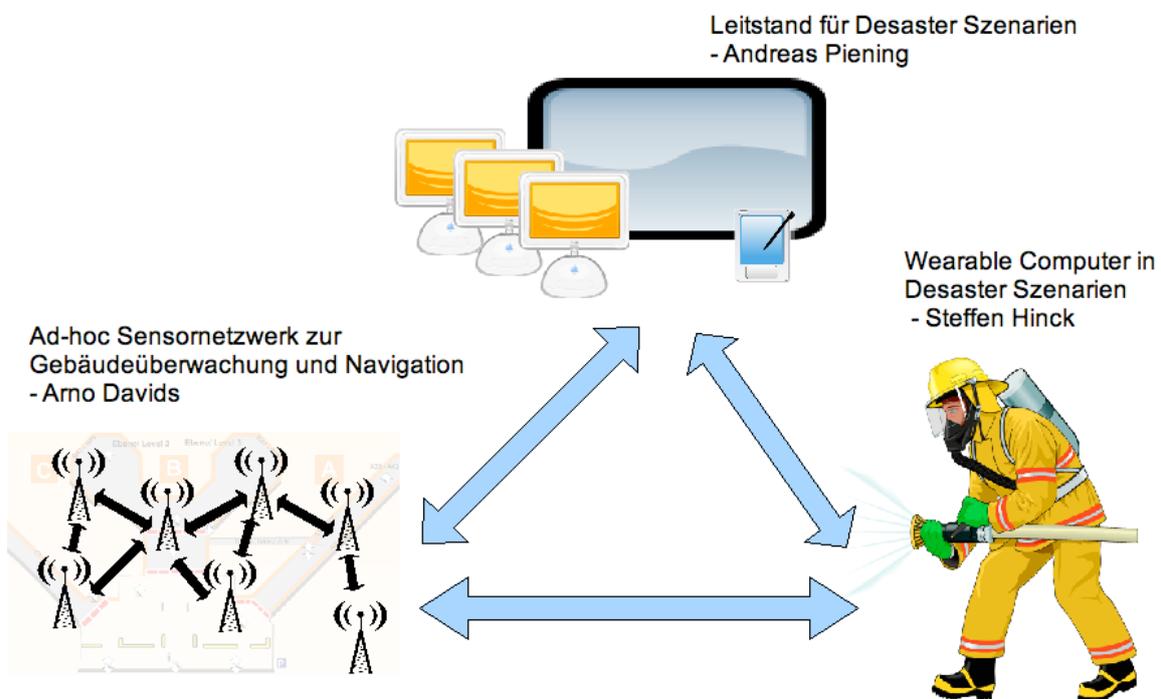


Abbildung 3.1.: RESCUE-Projekt: Zusammenspiel der Komponenten

Das Themengebiet von Andreas Piening ([Piening \(2007\)](#)) ist eine geeignete Plattform für Fragen der multimodalen Interaktion und wird hier für das eigentliche Szenario verwendet. Aus der Perspektive dieser Arbeit steht nicht die Rescue-Metapher im Vordergrund, sondern die funktionelle sowie technologische Umsetzung der Interaktion mit bzw. innerhalb eines Leitstandes.

Bevor Anforderungen an das Interaktionssystem erhoben werden können, müssen die Bedienungsmöglichkeiten des Leitstandes charakterisiert werden. Piening kennzeichnet den Leitstand als die Komponente mit Schlüsselfunktion und definiert den Aufgabenbereich:

- Koordination des Einsatzes
- Sammeln, Bewerten und Sondieren aller Informationen, die während der Rettungsmission seitens des Rettungspersonals und der beteiligten Komponenten gewonnen werden
- Verteilung von Warnungen und Hinweisen an das Rettungspersonal, Delegation von Befehlen und Überwachung des Geschehens

Die Idee des Leitstandes als Koordinations-, Kommunikations- und Delegationsplattform entspricht einem repräsentativen Beispielszenario des umfassenden Forschungsgebiets computergestützte kollaborative Arbeitsumgebungen (Computer Supported Collaborative Workspace – CSCW, vgl. Kapitel 2.3), das daraus resultierende konkrete System fällt in den Bereich Groupware (vgl. Kapitel 2.3).

Dieser Themenrahmen ist ein eigenes Forschungsgebiet im Rahmen des UbiComp-Projekts¹. Hier werden diverse Aspekte wie z.B. die Infrastruktur, die Darstellung oder die Interaktion in einem CSCW behandelt.

3.1.3. Konkretisierung: Interaktion

Interaktion bezeichnet die bidirektionale Kommunikation mit einem System bzw. in diesem Fall mit einer Umgebung, das heisst es besteht nicht nur die Möglichkeit der Eingabe in das System, sondern ebenfalls die Ausgabe, das Feedback an den Nutzer.

Ausgabe

In (Piening, 2007, "Konzeption") werden die wesentlichen Informationen benannt, welche dem Leitstand zur Verfügung gestellt werden müssen. Zur Darstellung gehören geografische Geländeübersichten, Positionen von Rettungskräften, Einsatzfahrzeugen und missionsstrategisch wichtigen Punkten, wie Brandherde oder Evakuierungsbereiche, sowie Statusinformationen und Priorisierungsannotationen.

Die Personen des Leitstandes befinden sich in einem Raum und sind idealerweise permanent von relevanten Informationen umgeben. Eine Möglichkeit der Umsetzung ist die Präsentation der Informationen auf überwiegend grossflächigen Anzeigeflächen. Grundsätzlich kann zwischen zwei Konzepten der Darstellung unterschieden werden, stets mit dem Aspekt der Mehrbenutzerfähigkeit:

¹UbiComp-Projekt der HAW Hamburg: <http://users.informatik.haw-hamburg.de/~ubicomp/>

TableTop Horizontal installierte Monitore berücksichtigen den gewohnten Aspekt der gemeinschaftlichen Arbeit an einem Konferenztisch. Die beteiligten Personen können, wie von herkömmlich bekannten, z.B. strategischen Besprechungen, gegenüber oder nebeneinander stehen. Unterstützend für die Kollaboration mit anderen Personen können Objekte² bzw. Bereiche direkt mit den Händen fokussiert bzw. berührt werden. Ermöglicht ist der direkte Sichtkontakt mit dem Kollegen, was Mimik und Gestik einschliesst.

Wall-Mounted Vertikal angebrachte Displays haben den klassischen Präsentationscharakter. Die beteiligten Personen stehen üblicherweise nebeneinander bzw. hintereinander. Zeigehandlungen können nachvollzogen werden, der direkte Sichtkontakt mit dem Partner ist eingeschränkt. Aufgrund der großen Anzeigefläche ist das direkte Markieren in Form von Berührung von Objekten nicht in jedem Bereich möglich.

Eingabe

Das Interaktionskonzept muss dem Nutzer die Navigation, das Stellen von Anfragen an Objekte und die Option der Eingabe erlauben. Es werden konkrete Eingabemetaphern benannt:

Annotation Jedes Objekt kann mit Informationen versehen und diese auf beliebige Art angezeigt werden. Dies kann automatisch erfolgen, wenn z.B. Objekte mit Datenquellen verknüpft sind und die Daten bestimmte Grenzbereiche erreichen oder verlassen. Alternativ ist die manuelle Eingabe möglich, um beispielsweise Stati von Objekten bei Ausfall der Sensorik von Hand zu ändern.

Distanzbestimmung Mit dem Markieren von zwei Objekten kann bei gleichzeitiger Modusumschaltung eine Entfernungsberechnung via Luftlinie bzw. über die kürzeste oder schnellste Route stattfinden.

Selektion und Detailabruf Bei Anwahl bzw. Markierung eines Objektes werden dem Nutzer weitere objektspezifische Informationen angezeigt. Zusätzlich erfolgt eine temporäre Zuweisung des Objektes an den Nutzer.

Befehlsdelegation Das Selektieren von Objekten eröffnet dem Benutzer die Möglichkeit, Anweisungen an das Objekt zu senden. Ein Kontextmenü stellt mögliche Aktionen dar, welche durch oder auf das Objekt ausführbar sind.

Interaktionskonzepte für konventionelle Bildschirmarbeitsplätze wurden unter der Voraussetzung entwickelt, dass ein einzelner Benutzer mit dem System arbeitet. Die Erhebung der

²Personen, Fahrzeuge, strategische Punkte, etc. werden im Folgenden als Objekte bezeichnet

Eingabeanforderungen muss sich von diesen herkömmlichen Eingabeschemata der klassischen Desktopmetapher distanzieren. Eine genaue Argumentationsreihe für ein neues Interaktionsbewusstsein liefert [van Dam \(1997\)](#), wie bereits in Kapitel 2.1 erörtert. Die wesentlichen aufgaben- sowie umgebungsbedingten Unterschiede zu bisherigen Verfahren sind die Notwendigkeit der Mehrbenutzerfähigkeit und die technologische Entwicklung bezüglich Größe und Positionierung aktueller Anzeigegeräte.

Im Anforderungskapitel muss dafür gesorgt werden, dass diese Funktionalität durch das Interaktionssystem gewährleistet wird.

3.1.4. Klassifizierung

Bevor konkrete Anforderungen erhoben werden, ist die systematische Einordnung des vorliegenden Szenarios sinnvoll. In der Forschung gibt es keine allgemeingültige Klassifizierung von CSCW-Systemen, jedoch mehrere Ansätze der Strukturierung aus unterschiedlichen Perspektiven:

Gemäß **Anytime/Anywhere-Matrix** (s. Abb. 3.2) nach [Kienle \(2005\)](#) in Anlehnung an die **Raum-Zeit-Matrix** von [O'Hara-Devereaux und Johansen \(1994\)](#) lassen sich Anwendungen im CSCW in vier Kategorien unterteilen. Zu beachten ist hierbei, dass lediglich Komponenten eines CSCW auf eine Kategorie beschränkt werden können, die gesamte Arbeitsumgebung muss allen vier Bereichen genüge tun.

Wird das gesamte Rescue-Szenario mit verteilter Sensorik, Leitstand etc. betrachtet, bewegt es sich auch in den Bereichen „Same Time / Different Place“ und „Different Time / Different Place“. Das gewählte Teilszenario für multimodale Interaktion begrenzt diese Kategorisierung. Da davon auszugehen ist, dass es nur einen Leitstand gibt, erfolgt die Einordnung dieser Arbeit in die Kategorie „Same Time / Same Place“ – die Entwicklung fokussiert Szenarien, in denen Teilnehmer an ein- und demselben Ort und zum gleichen Zeitpunkt, also zeitlich synchron und räumlich benachbart miteinander arbeiten.

[Teufel u. a. \(1995\)](#) unterscheiden im **3K-Modell** (siehe Abbildung 3.3) nach drei Typen der Interaktion bzw. Zusammenarbeit der Gruppe und stellt den jeweiligen Unterstützungsaspekt in den Vordergrund: **Kommunikation** ist die Verständigung von Personen, **Koordination** bezeichnet die Abstimmung aufgabenbezogener Aktivitäten und Ressourcen und **Kooperation** meint die gemeinschaftliche Verfolgung eines Ziels.

Diese Arbeit lässt sich nicht einer dieser Attribute zuordnen sondern befindet sich etwa im Zentrum des Dreiecks, daher ist eine Verwendung dieses Modells zur Klassifizierung nicht geeignet. Vielmehr kommt die Frage auf, ob sich hieraus nicht Anforderungen an das Eingabegerät stellen lassen. Die drei Punkte können als Aspekte auf Anwendungsebene fungieren, hierzu in Kapitel 3.2 mehr.

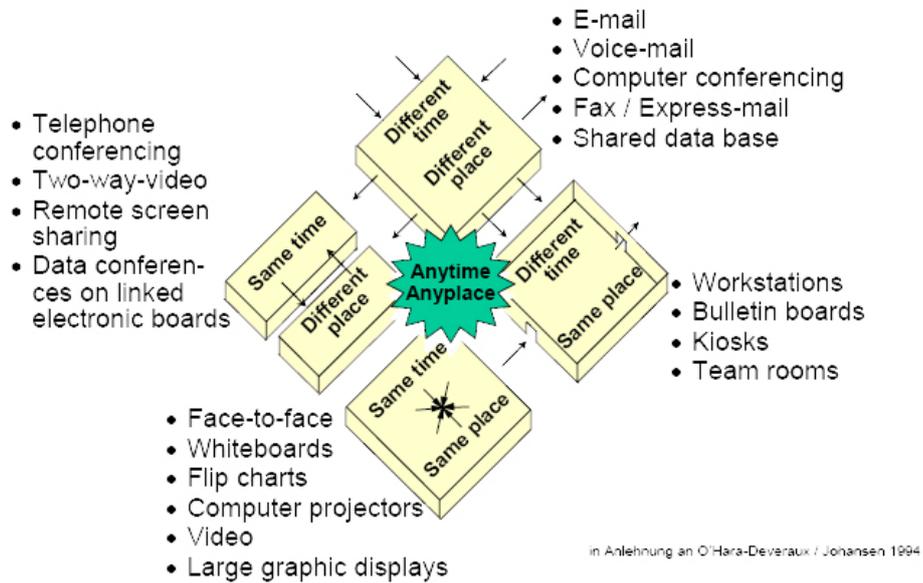


Abbildung 3.2.: Anytime/Anyplace-Matrix

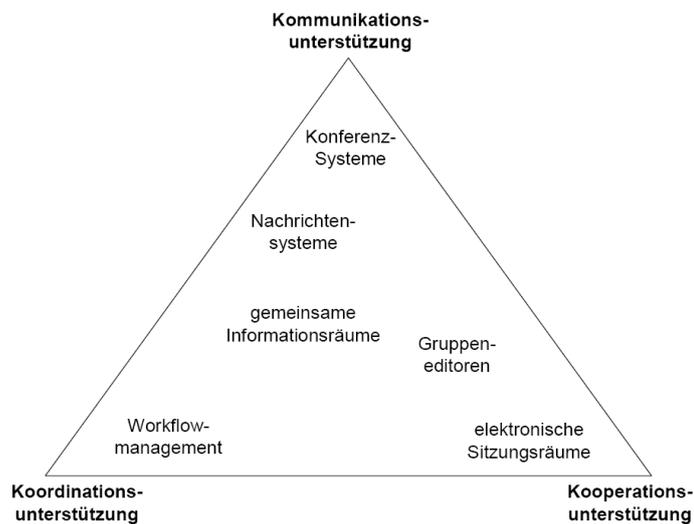


Abbildung 3.3.: Klassifikation nach 3K-Modell und Anwendungsklassen (aus [Schlichter u. a. \(2001\)](#))

Eine Klassifizierung nach Anwendungsklassen liefert [Ellis u. a. \(1991\)](#). Themenbereiche werden nach deren Funktionalität in das 3K-Modell eingeordnet, diese sind zum Beispiel: Nachrichtensysteme, Gruppeneeditoren, elektronische Sitzungsräume, Rechnerkonferenzen, Agentensysteme und Koordinationssysteme. Da sich das Szenario Leitstand nicht in das 3K-Modell einordnen lässt und die Klassifizierung nach Anwendungsklassen auf dem 3K-Modell aufbaut, ist hier folglich ebenfalls keine Einordnung möglich. Vorstellbar ist der Einsatz in Gruppeneeditoren zur direkten, mehrbenutzerfähigen Manipulation, in elektronischen Sitzungsräumen zur Präsentation und in Koordinationssystemen, um beispielsweise die Verteilung von Arbeitspaketen zu steuern.

3.2. Anforderung an Interaktion

Wie bereits in Kapitel [3.1.4](#) erwähnt, können die drei Komponenten des 3K-Modells (siehe [3.3](#)) verwendet werden, um Anforderungen an ein Eingabegerät zu erheben. In diesem Fall wird die Zielanwendung nicht nach diesen Punkten charakterisiert, sondern versucht, die Anwendung auf alle drei Punkte auszudehnen. Es handelt sich um die Aspekte Kommunikation, Koordination und Kooperation.

Dass dieser „Einsatz“ des 3K-Modell vertretbar ist, zeigt ein Vergleich mit einer Beschreibung von Kollaboration von Mike Gotta (META Group). In einer Gesprächsrunde von KMWorld ([Lamont u. a. \(2004\)](#)) unter Judith Lamont (KMWorld) zum Thema Kollaboration antwortet er auf die Frage, wie Kollaboration definiert ist:

„The industry has struggled to define collaboration concisely. We use a functional definition – tools and services that help groups of people share information, communicate, coordinate and work together as a team. From a technology perspective, it includes e-mail, calendaring, forums, instant messaging, Web conferencing and virtual workspaces.[...]“

Gotta benennt keine eindeutige Definition, sondern stellt den Bezug zur Perspektive her und definiert zwei Sichtweisen: funktional und technologisch. Eine **funktionale Betrachtung** erfolgt auf einer vollkommen abstrakten Ebene ohne Bezug auf den technischen Hintergrund. Er beschreibt soziale Aspekte, welche notwendig für einer Kollaboration sind. Die **technologische Schicht** setzt sich mit sämtlichen Fragestellungen bezüglich Auswahl und Verwendung von Hard-, Software sowie verwendeten Diensten auseinander.

Eine derartige Unterteilung ist ebenfalls sinnvoll für die Entwicklung eines Eingabegerätes im Sinne dieser Arbeit. Die beiden Sichtweisen können konzeptionell unabhängig voneinander betrachtet werden, bei der Entwicklung und Implementation jedoch ist eine Synergie der beiden Bestandteile unbedingt notwendig. Auf funktionaler Ebene werden Anforderungen an

das Gesamtsystem definiert, dies ist hier in Kapitel 3.2 die Zielsetzung. Diese müssen von den Komponenten auf technologischen Ebene erfüllt werden, was im Kapitel 4 erarbeitet wird.

Zusammenfassend ergeben sich aus funktionaler Sicht vier Attribute der Kollaboration, die durch das Interaktionskonzept abgedeckt werden müssen. Der Vergleich zum 3K-Modell und Gotta ist in Tabelle 3.1 dargestellt.

Attribut	nach Gotta	3K-Modell
Teilen von Wissen	share information	Kooperation
Kommunikation	communicate	Kommunikation
Koordination	coordinate	Koordination
Teamarbeit	work together as a team	Kooperation

Tabelle 3.1.: Vergleich funktioneller Attribute

Dem gegenübergestellt werden die nichtfunktionalen Anforderungen an die Anwendungsbedien-
 dung des Leitstandes nach [Piening \(2007\)](#). Diese sind Annotation, Distanzbestimmung,
 Selektion und Detailabruf sowie Befehlsdelegation. Hier eine nähere Erläuterung der Attri-
 bute, welche ein Interaktionsschema im CSCW unterstützen muss, und die jeweilige Einord-
 nung der nichtfunktionalen Anforderungen:

Teilen von Wissen Die Betrachtung von Wissen schliesst gemäß DIKW-Modell nach [Ack-off \(1989\)](#) neben dem eigentlichen Wissen ebenfalls Informationen und Daten ein. Im
 folgenden stehen diese drei Begriffe für ein und dasselbe, sie an dieser Stelle bzw.
 in dieser Arbeit zu unterscheiden, ist nicht notwendig. Interessant wird die Klassifizie-
 rung beim späteren Entwurf des Feinkonzeptes auf der funktionalen Ebene, wie zum
 Beispiel im Fall einer Synergie der drei Komponenten.

Um die Verbreitung von Wissen anschaulich zu gestalten, wird es im vorliegenden Sze-
 nario durch Objekte, „greifbare“ Symbole dargestellt. Beteiligte können Informationen
 zeigen bzw. präsentieren und austauschen. Im Leitstand gehört hierzu die Selektion
 mit Detailabruf und im weitesten Sinne die Annotation von Objekten sowie die Distanz-
 bestimmung.

Kommunikation ist, neben dem Austausch von Wissen, Grundvoraussetzung für das Ent-
 stehen von diesem. Gemäß Klassifizierung des Szenarios und somit des „Collaborati-
 ve Workspcae“ nach zeitlich synchronen und räumlich benachbarten Abläufen ist die
 Mehrbenutzerfähigkeit des Systems zwingend notwendig.

Koordination Das System unterstützt die Benutzer bei der Koordination von Arbeitsabläu-
 fen, beispielhaft ist eine Aufgabenverteilung bei einer Projektbesprechung oder bezüg-
 lich des Leitstand-Szenarios die Befehlsdelegation.

Teamarbeit Aufbauend auf die Kommunikation muss das gemeinschaftliche Handeln als solches unterstützt werden. Wichtig ist hier, dass alle aktiv teilnehmenden Benutzer gleichzeitig miteinander arbeiten, in diesem Fall interagieren können.

Der Einsatzleitstand ist die zentrale Koordinations- und Kommunikationsplattform und wird folglich von mehreren Personen besetzt. Gleichzeitig ist er die Informationssenk vor Ort, alle in der Einsatzumgebung gesammelten Informationen laufen hier zentral zusammen. Aus zwei Gründen muss hier die lokale Zusammenarbeit uneingeschränkt gewährleistet sein:

- Jeder Entscheidungsträger muss beliebig auf den gemeinsamen Datenbestand zugreifen können, ohne dabei andere Mitarbeiter zu behindern.
- Die gleichzeitige Kooperation muss möglich sein, das heißt die beteiligten Personen müssen miteinander arbeiten können. Das Benutzerkonzept bezieht sich auf das Gesamtsystem und somit verschiedene Themengebiete, wie z.B. die Netz-Infrastruktur, die Betrachtung hier beschränkt sich auf den Gesichtspunkt der Interaktion.

Kollaboration bezeichnet neben der Zusammenarbeit mehrerer Einzelpersonen bzw. mehrerer Gruppen die Zusammenarbeit mit verschiedensten Systemen. Gemäß der in Kapitel 3.1.3 beschriebenen Anzeigeszenarien sind zwei Interaktionskonzepte notwendig, um die Anforderungen zu erfüllen und überhaupt das Potential einer solchen Umgebung auszuschöpfen: Interaktion mit TableTop- und mit Wall-Mounted-Displays.

Die besondere Herausforderung ist hierbei die Gewährleistung des nahtlosen Wechsels bzw. Übergangs von einer Schnittstelle zur anderen innerhalb des Gesamtsystems. Mit dieser Fragestellung wird das Forschungsgebiet der „nahtlosen Interaktion“ („Seamless Interaction“) betreten. Ishii u. a. (1994) bezeichnen die notwendige Grundmotivation, aus der heraus Interaktionskonzepte entstehen sollten, und beschreiben zwei wesentliche Gesichtspunkte der Metapher „nahtlos“ im Kontext der Interaktion in computergestützten Umgebungen.

Wie aus dem Grundlagenkapitel 2.3 bekannt, bestimmt die Richtung der Analyse des Akronymes CSCW die eigentliche Motivation für eine kollaborative Umgebung. Entweder bildet der technologische Aspekt (Computer im Vordergrund – Vorwärtsanalyse) oder die eigentliche Aufgabe (organisationstheoretischer Hintergrund – Rückwärtsanalyse) die Grundlage für einen Systementwurf. Bei der Beschreibung erster Erfahrungen zur Entwicklung multimedialer Infrastrukturen im Rahmen eines Forschungsprojektes am (damaligen) Rank Xerox's EuroPARC benannten Buxton und Moran (1990) ihre klare Tendenz:

[...] we wanted to learn to do smart things with stupid technology today, rather than wait and do stupid things with smart technology tomorrow. [...]

Diese Aussage entspricht der oben genannten Rückwärtsanalyse und unterstreicht damit die Richtung, in der hier gegangen werden soll. Ishii u. a. (1994) zeigen sich als Verfechter dieses

Leitfadens und verdeutlichen diese Aussage mit zwei Thesen, dargelegt als Folgerungen aus der Begrifflichkeit „nahtlos“:

Kontinuität Ein Benutzer entwickelt im Laufe seiner Arbeit und der damit verbundenen Kommunikation mit anderen Personen und Systemen, für sich optimale Methoden der Bearbeitung seiner Aufgaben. Diese Erfahrungen sollten als gewinnbringend angesehen und dementsprechend in neue Systemstrukturen übernommen werden. Kontraproduktiv ist Groupware, welche von diesem Benutzer eine konzeptionell völlig neue Bedienung verlangt.

reibungsloser Übergang Im Rahmen eines kollaborativen Arbeitsumfeldes besteht die Notwendigkeit, an verschiedenen Arbeitsplätzen und in unterschiedlichen Modi zu arbeiten. Ziel in dieser Umgebung muss die Minimierung der kognitiven Anforderungen an den Nutzer bei der freien Bewegung und möglichem Wechsel zwischen den Systemen sein.

In [Fischer \(2006\)](#) wurden Gesichtspunkte herausgearbeitet, mit denen eine Entwicklung getreu der nahtlosen Interaktion möglich ist:

- **Intuitivität:** Ein Benutzer möchte ein Objekt hervorheben bzw. markieren und zeigt darauf – für ihn und jeden Beteiligten ist eindeutig verständlich, um welches Objekt es sich handelt. Die Interaktion erfolgt selbstverständlich, wird z.B. etwas gegriffen, ändert sich der Objektzustand, bei gewollter Bewegung verändert sich die Position des Objektes.
- **Natürlichkeit:** Mit Natürlichkeit sind an dieser Stelle keine Konzepte, Dinge, Abläufe oder Vorgänge gemeint, welche der Natur entspringen. In diesem Szenario passiert Interaktion wie erwartet, Bewegungen und Wechselwirkungen geschehen naheliegend bzw. selbstverständlich. Dies muss nicht zwangsläufig die Natur als Vorbild haben, sondern „gelernte“ Natürlichkeit. Ein Beispiel ist hier der Umgang mit einer Desktop-Oberfläche auf einem Computer – die Verwendung dieser ist antrainiert und wird als natürlich empfunden.

Beispiel: Ergänzend zur Intuitivität soll die Art und Weise, wie der Benutzer etwas bewegt, an Realismus gewinnen. Werden Objekte unter einer Geschwindigkeit losgelassen, bleiben sie nicht an der aktuellen Position stehen, sondern verzögern, ganz wie es aus der realen Welt erwartet wird.

- **Umfangreichtum:** Der Benutzer wird in den Möglichkeiten, mit seiner Umgebung in Interaktion zu treten, nicht beschnitten. Er hat einen nahezu uneingeschränkten Funktionsumfang.
- **Mehrbenutzerfähigkeit:** Mehrere Benutzer können gleichzeitig untereinander und mit Objekten interagieren

3.3. Festlegung der Modalitäten

Aus dem Grundlagenkapitel 2.2 ist bekannt, dass multimodale Interaktion im Idealfall sämtliche Sinne eines Menschen einbezieht. Die Gestaltung des Ausgabekanals ist bereits durch das Szenario gegeben und nicht Aufgabe dieser Arbeit: die Computer-zu-Mensch-Interaktion geschieht „nur“ visuell, es werden großflächige Anzeigen im ganzen Raum verteilt.

Die Mensch-zu-Computer-Interaktion bzw. Eingabe wird hier von dem Anspruch der Natürlichkeit bzw. Intuitivität aus Kapitel 3.2 bestimmt. Die reale Kommunikation zwischen zwei Personen geschieht im allgemeinen Fall audio-visuell, zur Anwendung kommen **Sprache als Medium, Gestik** und **Mimik**.

Gestik wird unter anderem mit Hilfe von Händen und Armen lautsprachersetzend und lautsprachbegleitend bzw. lautsprachunterstützend eingesetzt. Eine Einteilung der Gesten kann in drei Arten erfolgen: lexikalisierte Gesten, Zeigegesten und ikonische Gesten. Rückblickend auf die Anwendungsfälle des Szenarios sind objektbezogene Handlungen direkt auf der Anzeigefläche gefordert bzw. sinnvoll und durch direktes Zeigen oder Berührung sehr gut umsetzbar. Aus diesem Grund wird der Hauptaugenmerk auf Zeigegestik gelenkt.

Sprache kommt unterstützend zum Einsatz, verstärkt oder ergänzt somit Gestiken.

Mimik gibt grossen Aufschluss über den Gemütszustand einer Person. Bezogen auf das Rescue-Szenario kann die Darstellung von Informationen davon abhängig gemacht werden. Ist der Benutzer besonders aufgeregt, wird die Informationsdarstellung reduziert, um somit den Überblick in einer Stresssituation entgegen zu kommen. Dies ist nur ein Beispiel für ein mögliches adaptives Verhalten des Systems auf die Verfassung der involvierten Person.

Im Rahmen dieser Arbeit hat das System nicht die Aufgabe, die Gemütszustände der Benutzer zu erkennen, es wird vorausgesetzt, dass die Interaktion mit dem System direkt und entschieden erfolgen muss – die Auswertung von Mimik wird daher nicht als Eingabemodalität eingesetzt.

Die gesamten bisherigen Anforderungen können auf funktionaler Ebene durch folgende Aktionen abgedeckt werden:

Zeigerpositionierung Einem Kooperationspartner oder dem System muss erkenntlich gemacht werden können, welches Objekt bzw. welcher Punkt auf der Anzeigefläche aktuell von Interesse ist. Dies erfordert die Möglichkeit des „Zeigens“.

Wechselwirkung Sobald auf ein Objekt gezeigt wird, muss die Möglichkeit bestehen, eine Aktion darauf oder damit auszuführen. Gemeint ist hier z.B. die Anwahl, Auswahl, Markierung von Objekten oder Bereichen oder das Hervorrufen eines objektspezifischen Zustandwechsels.

Bewegung Ein selektiertes Objekt soll, soweit es dem Anwendungskonzept entspricht, bewegt werden können. Hierzu gehört neben der Translation ebenfalls die Rotation. Zu beachten ist, dass keine Beschränkung auf den zweidimensionalen Raum erfolgen soll – dreidimensionale Darstellung impliziert die Bewegung in drei Richtungen.

Mehrbenutzerfähigkeit Eine gleichzeitige Verwendung des Systems durch mehrere Benutzer erfordert neben der anwendungstechnischen Umsetzung des parallelen bzw. kooperierenden Arbeitens die technologische Möglichkeit der Eingabe von mehreren Geräten. Wesentlich ist hierfür die eindeutige Assoziation des erhaltenen Eingabeereignisses mit dem entsprechenden Benutzer.

Benutzerlokalisierung Besonders im Rahmen der Mehrbenutzerfähigkeit ist die Möglichkeit der Lokalisierung des Benutzers von Vorteil. Aus den Untersuchungen zur territorialen Einteilung der Interaktionsfläche für TableTop-Arbeitsplätze von [Scott \(2005\)](#) geht hervor, dass es natürlichen Ursprungs ist, sich den Anzeigebereich abhängig von seiner Position zu manipulieren (s. Abb. 3.4). Darauf basierend ist es sinnvoll die Anzeigefläche und Inhalte der anzuzeigenden Informationen der räumlichen Position des einzelnen Benutzers dynamisch anzupassen.



Abbildung 3.4.: Territoriale Aufteilung der Arbeitsfläche nach [Scott \(2005\)](#)

Die zu verwendenden Modalitäten sind Gestik mittels Hand und Sprache. Es muss festgelegt werden, wie die oben erarbeiteten Aktionen mittels Hand und Sprache umgesetzt werden können. Die Betrachtung hierbei muss für die beiden Zielsysteme individuell erfolgen bzw. es muss erörtert werden, inwiefern eine Unterscheidung bei der Umsetzung der Aktionen notwendig ist.

Mit Blick auf die Ansprüche Intuitivität und Natürlichkeit soll im Rahmen der Objektmanipulation das Gefühl vermittelt werden, das Objekt direkt mit der Hand greifen und bewegen zu können. Die **Zeigerpositionierung** wird idealerweise durch eine Zeigegestik, mit einem Finger bzw. der Hand, vollzogen. Hier kommt der wesentliche Unterschied zwischen TableTop- und Wall-Mounted-Systemen zu Tage: Distanz. Auf einem Tisch kann ein Objekt unmittelbar berührt werden, im Szenario verwendete Displays sind so gestaltet, dass jeder Punkt mit der Hand einer Person erreicht werden kann.

Tabelle 3.2 zeigt Funktionen der Hand, welche für die Interaktion in der Systemumgebung interessant sind.

Es müssen zwei Eingabeszenarien fokussiert werden. Die Fragestellung lautet: Wie kann am Besten mit einem TableTop und mit einer Wand gearbeitet werden?

Bezeichnung	Funktion der Hand	Funktion in der Systemumgebung
Zeigerpositionierung Orientierung Lokalisierung	Zeigerichtung dreidimensionale Haltung Position im dreidimensionalen Raum	Bestimmung des Zeigeortes Bewegung von Objekten Adaptives Systemverhalten (bzgl. Position des Betrachters)
Wechselwirkung	Greifen, Fingerdrücken	Hervorrufen einer Zustandsänderung von Objekten oder einer Wechselwirkung in der Systemumgebung (z.B. Klick)
Bewegung	Ausüben von Kraft in eine Richtung (Gestik)	Bewegung von Objekten
Haptik	Wahrnehmung von physikalischer Kraft (z.B. Druck, Vibration)	Wirkung von Objektkräften auf deren Ankerpunkt

Tabelle 3.2.: Assoziierung der Funktionen: Hand zu System

3.4. Überblick vorhandener Technologien

Mit potentiell geeigneten Eingabegeräten sind hier Geräte gemeint, welche annähernd den Anforderungen aus Kapitel 3.2 gerecht werden, diese sind: Zeigerpositionierung, Orientierung, Lokalisierung, Wechselwirkung, Bewegung und Haptik.

Wichtig ist bei der Wahl von Referenzprojekten die Beachtung von Beschränkungen der Interaktion aus der Seminararbeit. Im Hinblick auf die Natürlichkeit der Interaktion bilden die Hände das Eingabemedium, durch Gestiken werden Wechselwirkungen erreicht. Somit reduziert sich das Themenumfeld und lässt sich wie folgt klassifizieren:

3.4.1. Optisches, statisches Tracking

Mit optischem Tracking ist die Methodik gemeint, mit Hilfe eines statischen Kameraaufbaus in einer Interaktionsumgebung Bewegungen aufzunehmen und zu analysieren. Zu unterscheiden sind hier zwei Möglichkeiten der Umsetzung:

Tracking des Eingabemediums

Hierbei wird das Interaktionsobjekt an sich aufgenommen. Bezogen auf die Seminararbeit ist die Hand und damit Objekte in der Hand im Fokus der Auswertung. Es gibt diverse Projekte, in denen diese Technik verwendet und validiert wird. Beispielhaft werden hier Techniken präsentiert:

2003 entwickeln [Cao und Balakrishnan \(2003\)](#) VisionWand (vgl. Abb. 3.5), ein Stab, der als dreidimensionales Eingabegerät fungiert. Bewegungen in alle sechs Freiheitsgrade (Translation X, Y, Z und Rotation X, Y, Z) können erfasst werden. Von zwei Kameras aufgenommen, wird die Position und Orientierung des Gerätes anhand von farblich eindeutigen Markierungen an den beiden Enden errechnet.

[Miyaoku u. a. \(2004\)](#) verwenden in „C-Blink“, ein mobiles Endgerät mit Display als Eingabegerät. Eine auf der Leinwand befestigte Kamera erkennt das Endgerät, indem das Display als Lichtquelle fungiert. Über zugeordnete Farben können unterschiedliche Geräte erkannt werden (vgl. Abb. 3.6). Um sich von der Umgebungshelligkeit abzuheben, leuchtet die Lichtquelle nicht permanent, sondern blinkt z.B. mit 30 Hertz. Die Funktionalität beschränkt sich hierbei auf die Positionsbestimmung des Eingabegerätes, wodurch in Verbindung mit den Tasten auf dem Gerät eine Selektion von Objekten auf der Leinwand möglich ist.

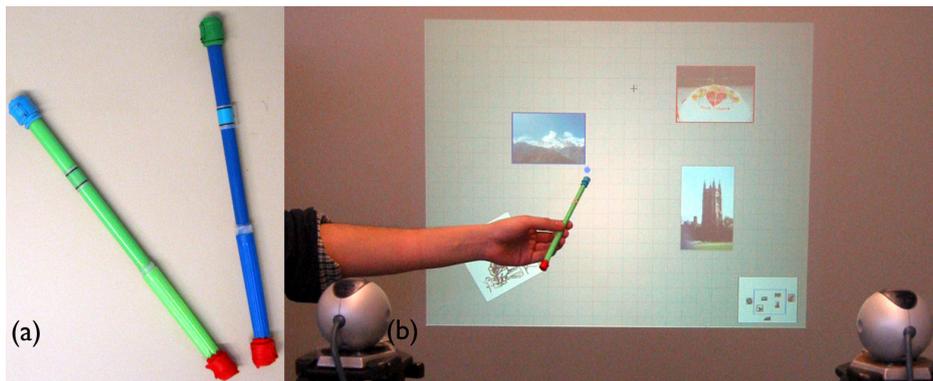


Abbildung 3.5.: VisionWand

Vogel und Balakrishnan (2005) verwenden einen mit Markierungen versehenen Handschuh (vgl. Abb. 3.7), der mit Hilfe eines Vicon-Tracking-Systems³ im Raum erkannt wird. Aktuell werden nur Funktionen umgesetzt, die aus der Interaktion mit einer Maus bekannt sind, der Fokus liegt auf Zeigen und Selektion.

Lösungen wie „VisionWand“ erfüllen zu einem Großteil die Anforderungen an die Interaktionskomponente, jedoch ist das optische, statische Tracking des Eingabegerätes nicht praktikabel, wenn auf die Szenarien in dem Collaborative Workspace zurückgeblickt wird. Die Aufbauten (Kameras, etc.) schränken den Bewegungsfreiraum erheblich ein, was sich besonders bei der Verwendung mit mehreren Personen nachteilig niederschlägt.

Tracking auf dem Ausgabemedium

Tracking des Ausgabemediums, in diesem Fall großer Display bzw. Leinwände hat den Vorteil, dass Kameraaufbauten zwar benötigt werden, aber besonders bei der Installation hinter der Leinwand keineswegs den Bewegungsfreiraum beeinträchtigen.

Einen klassischen Ansatz für Laserpointer-Interaktion liefern Dan R. Olsen und Nielsen (2001). Der Punkt eines Laserpointers auf einem Display wird als Cursor interpretiert und kann positionsabhängig einen von vier Modi einnehmen: Positionierung, Scrolling, Graffiti (Eingabe im Freistil) und als Interaktionselement auf dafür vorgesehene Schaltflächen.

Funktional weiter ausgeprägt ist eine Mehrpunkt-Lasertechnik von Matveyev und Göbel (2003) (vgl. Abb. 3.8), der IR-Stylus(optical tweezers). Der prägnante Unterschied ist die Verwendung bestimmter Optik, welche einen Laserpunkt in drei einzelne aufspaltet, wodurch funktional vier Freiheitsgrade der Bewegung erreicht werden. Die Interaktion mit Objekten kann über zwei Ansätze umgesetzt werden: Über Gestiken, die geometrisch interpretiert

³Vicon Tracking System: www.Vicon.com

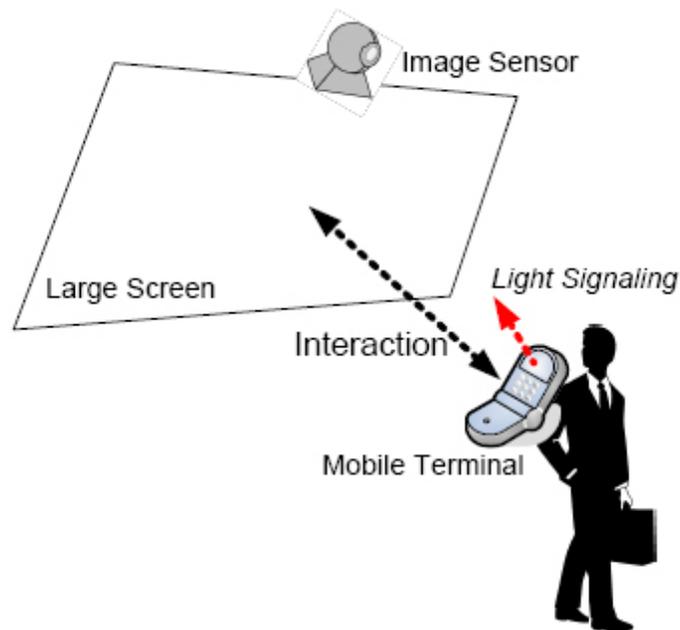


Abbildung 3.6.: C-Blink

werden oder durch gewollte mechanische Veränderung der Optik durch z.B. Drücken einer Taste.

Mehrbenutzerfähigkeit beim Einsatz von Laserstrahlen erreichen [Bi u. a. \(2005\)](#). Das Eingabegerät wird um einen Funksender erweitert, der neben der Betätigung des Laserpointers periodisch die Kennung des Gerätes unter Verwendung von Zeitschlitzverfahren überträgt. Somit ist eindeutig assoziierbar, welcher Benutzer mit welchem Zeigegerät arbeitet.

Um den Funktionsumfang zu erweitern, bietet sich der Einsatz von Gestenerkennung an, wie es beispielsweise [Shizuki u. a. \(2006\)](#) tun. Hierbei liegt der Fokus auf der Steuerung von Präsentationen oder Objekten auf der Leinwand. Durch das Ausführen von Bewegungsmus-

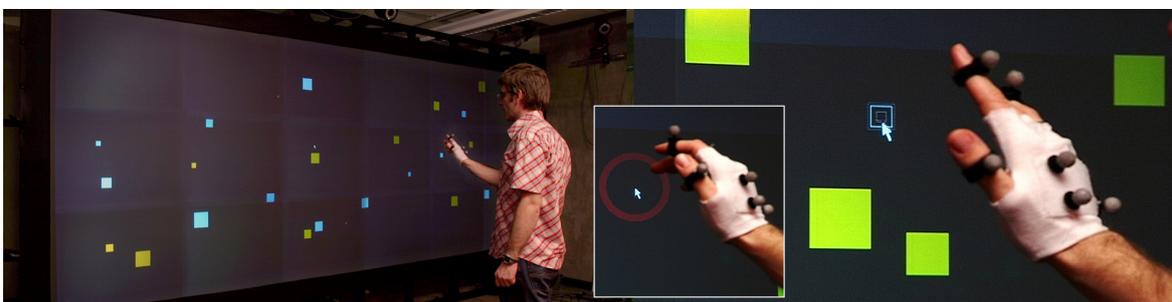


Abbildung 3.7.: Distant Freehand Pointing and Clicking

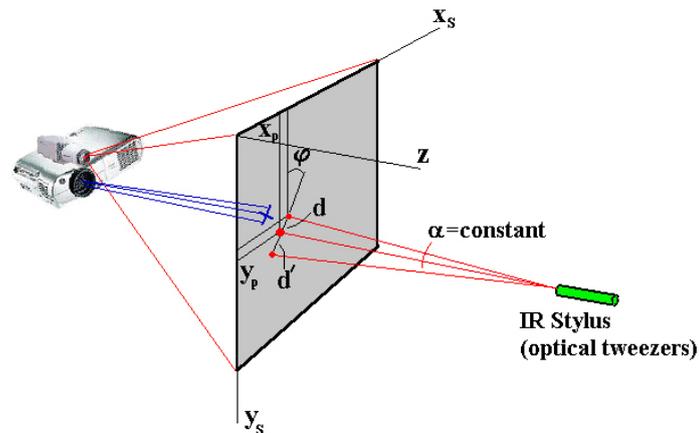


Abbildung 3.8.: IR-Stylus

tern (besonders mit Bewegungen über den Bildschirmrand hinaus) kann Interaktion erreicht werden, wie z.B. Seitenwechsel, Rotation oder Zoom.

3.4.2. Kamera als Eingabegerät

Dieser Abschnitt könnte zu Kapitel 3.4.1 gehören, denn eine Kamera ist auf das ausgegebene Bild gerichtet und markante Punkte im Bild werden analysiert. Die eigentliche Idee ist jedoch eine ganz andere: eine Kamera selbst wird als Interaktionskomponente verwendet und nimmt, im Gegensatz zu z.B. den Laserpointern, nicht die Auswirkung anderer Eingabegeräte auf, sondern es wird das eigene Bild interpretiert, wodurch eine Interaktion adaptiert werden kann.

Gemäß Ballagas u. a. (2006) können Applikationen anhand von zwei Attributen klassifiziert werden. Die **Positionierung bzw. Bewegung** und die **Ausrichtung**. Die Positionierung sagt aus, ob ein Objekt, wie z.B. ein Zeiger, relativ oder absolut bewegt wird. Die Ausrichtung der Kamera meint, ob auf den zu interagierenden Bereich oder einer Referenzfläche gezeigt wird oder in eine andere Richtung, also die direkte oder indirekte Verwendung. Dies erlaubt eine Differenzierung nach vier unterschiedlichen Ansätzen. Im Folgenden wird nach dem Attribut Ausrichtung unterschieden:

Direkte Manipulation

Eine Kamera wird direkt auf die zu interagierende Fläche oder den zu interagierenden Raum gerichtet.

Wie in vielen aktuellen Projekten wird bei [Ballagas u. a. \(2005\)](#) eine integrierte Kamera eines mobilen Endgerätes (in dem Fall ein Handy) verwendet. Mit „Point and Shoot“ kommen visuelle Codes zum Einsatz, um die Koordinaten des Zeigepunktes zu bestimmen (s. Abb. 3.9). Es gibt mehrere Ansätze zur Implementation, aktuell wird bei der Selektierung eines Objektes kurzzeitig ein Raster auf dem eigentlichen Bild eingeblendet. Eine Optimierung wäre die Visualisierung der Codes mit Infrarotlicht. Nachteilig ist hierbei, dass das optische Feedback nur zeitdiskret umsetzbar ist, der Benutzer würde lediglich auf dem Handy einen Cursor sehen, erst Interaktionen sind auf dem großen Display nachvollziehbar.



Abbildung 3.9.: Point and Shoot

Ein vielversprechender Ansatz ist die Umsetzung des Interaktionsgerätes „Zoom-and-Pick“, das gleichzeitig die Ausgabequelle ist. [Forlines u. a. \(2005\)](#) zeigen ein System, welches auf einen beliebigen, jedoch markierten Bereich das Ausgabebild projiziert, und gleichzeitig mit einer Kamera den Fokus der Halterichtung und damit die Position des Cursors berechnet. Wie in Abbildung 3.10 zu sehen ist, passt sich das Bild der Haltung der Hand bzw. des Arms an und erscheint stets perspektivisch korrekt. Über geometrische Algorithmen lässt sich die Zeigerposition berechnen. Theoretisch kann so jede Oberfläche zu einem Display werden, einzig die Frage der Kollaboration wird in der Arbeit nicht gestellt, wobei die Mehrbenutzerfähigkeit technisch umsetzbar ist.

Indirekte Manipulation

Das „CaMus“-System von [Rohs u. a. \(2006\)](#) hat als Ziel, eine musikalische Live-Performance zu gestalten. Es ähnelt stark dem „Point and Shoot“ von [Ballagas u. a. \(2005\)](#), jedoch ist die Positionierung hier relativ umgesetzt. Das Handy wird auf einen „virtuellen Arbeitsplatz“ gerichtet und über feste Code-Gitter-Strukturen ist eine konkrete, absolute Lokalisierung des Zeigers möglich (s. Abb. 3.11).

Ebenfalls unter Verwendung einer Gitterstruktur mit farbigen Codes geht das Eingabesystem von [Hachet u. a. \(2005\)](#) einen Schritt weiter. Es ermöglicht relative Bewegungen in alle sechs Freiheitsgrade (dreidimensionale Translation und Rotation). Die Philosophie hierbei ist, das zu

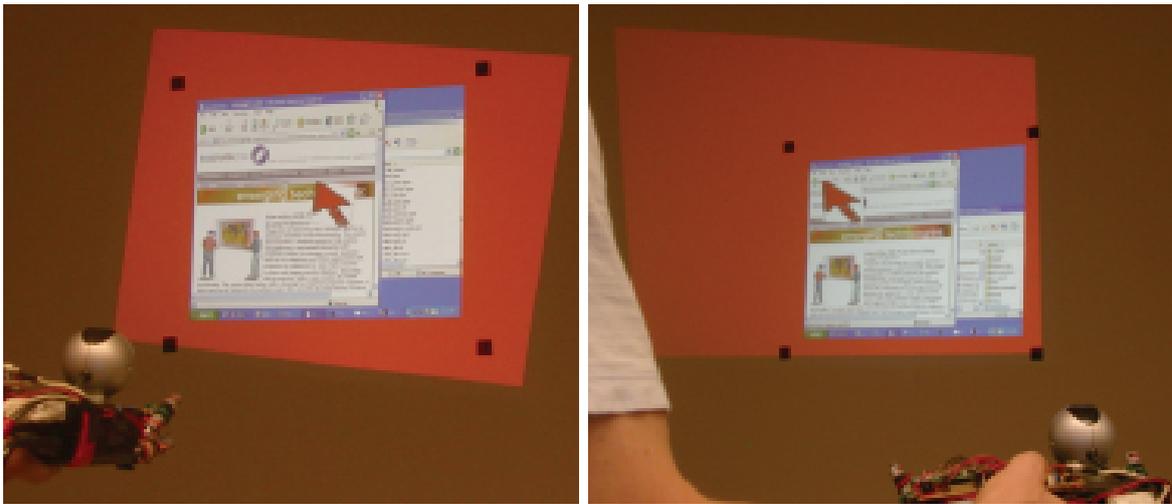


Abbildung 3.10.: Zoom-and-Pick: Projektor und Kamera

betrachtende Objekt in der Hand zu halten (das Gittermuster) und mit einer, auf einem PDA installierten, Kamera die Bewegungen auswerten zu können (s. Abb. 3.12).

Ungebunden an eine bestimmte Kameravorlage arbeiten die Systeme „Sweep“ von [Ballagas u. a. \(2005\)](#) und „TinyMotion“ von [Wang u. a. \(2006\)](#), jedoch ist hier die Positionierung relativ. Über Differenzen zwischen Bildern können Handbewegungen in vier Richtungen festgestellt werden (vgl. Sweep in Abb. 3.13).

3.4.3. Interaktion durch Bewegung

Interaktion bzw. Positionierung durch Bewegung ist bereits in Kapitel 3.4.2 mit den Systemen „Sweep“ und „tinyMotion“ vorgestellt worden. An dieser Stelle werden für die Aufnahme von Bewegungen keine Kameras verwendet, sondern Beschleunigungssensoren. Dass die Interaktion mit solchen Sensoren Potential hat, zeigt [Zhai u. a. \(1993\)](#) in einem Vergleich verschiedener Eingabetechniken im dreidimensionalen Raum, als versucht wird, zwei Tetraeder konkruent aufeinander zu bringen und dies in passabler Zeit gelingt. Festzuhalten ist hierbei, dass es sich um relative Bewegungen handelt. Darauf aufbauend gibt es viele Projekte, eines von [Rekimoto \(1996\)](#), in dem ein PDA durch Rotation um die eigenen Achsen bedient werden kann.



Abbildung 3.11.: CaMus

3.4.4. Interaktion durch Sprache

Sprache als Modalität für die Interaktion kommt hier unterstützend zum Einsatz. Generell wird bei Spracherkennungssystemen zwischen der **sprecherunabhängigen Spracherkennung** und der **sprecherabhängigen Spracherkennung** unterschieden. Das sprecherunabhängige System benötigt keine vorhergehende Trainingsphase der Benutzer, hat aber zur Folge, dass der Wortschatz auf einige tausend Wörter begrenzt ist. Wird das Erkennungssystem sprecherabhängig umgesetzt, muss ein Training erfolgen und ist damit an spezielle Personen gebunden. Der verwendbare Wortschatz ist signifikant höher.

Das Forschungsgebiet ist bereits weit erschlossen, es gibt viele kommerzielle Software-Produkte mit sehr hoher Treffergenauigkeit.

Tse u. a. (2006) verwenden Spracherkennung in einem multimodalen Interaktionssystem und bereichern dadurch Zeigereignisse und Gestiken an. Auf einem Tabletop-System werden die Einzelbenutzerprogramme Google Earth und Warcraft 3 durch Gestik und Sprache von mehreren Personen gesteuert (vgl. Abbildung 3.14). Für die Spracherkennung wird das „Speech Application Programmers Interface“⁴ von Microsoft eingesetzt. Um gesprochene Wörter eindeutig einem Benutzer zuzuordnen zu können, erfolgt die Auswertung der Sprache für jeden Benutzer individuell auf einem eigenen Rechner unter der Verwendung von Headsets. Dies ermöglicht Mehrbenutzerfähigkeit ist, aber aus technischer Sicht sehr aufwändig.

Eine technisch interessante Lösung verspricht das „In Silico Vox“-Projekt von Lin u. a. (2007). Eine wesentlich schnellere Dekodierung und die Reduktion des Leistungsverbrauches wird

⁴Microsoft SAPI: <http://www.microsoft.com/speech/speech2007/default.aspx>



Abbildung 3.12.: Hachet: Zweihändige 3D-Interaktion



Abbildung 3.13.: Sweep

durch die direkte Umsetzung der Spracherkennung in Hardware erreicht. Simulationsergebnisse einer FPGA-Lösung (Architektur siehe Abb. 3.15) zeigen, dass es mit einer 10- bis 20-fach geringeren Taktfrequenz möglich ist, gleiche Ergebnisse zu erreichen wie professionelle Anwendungen auf Softwarebasis, Referenz ist CUE Sphinx⁵ 3.0.

3.5. Auswertung

Im Rahmen der Analyse wurden, bezogen auf das Szenario, die benötigten Modalitäten für eine performante Interaktion bestimmt (vgl. Kapitel 3.3). Im nächsten Kapitel (4) findet der

⁵CMU Sphinx Open Source Speech Recognition Engines: <http://cmusphinx.sourceforge.net/html/cmusphinx.php>

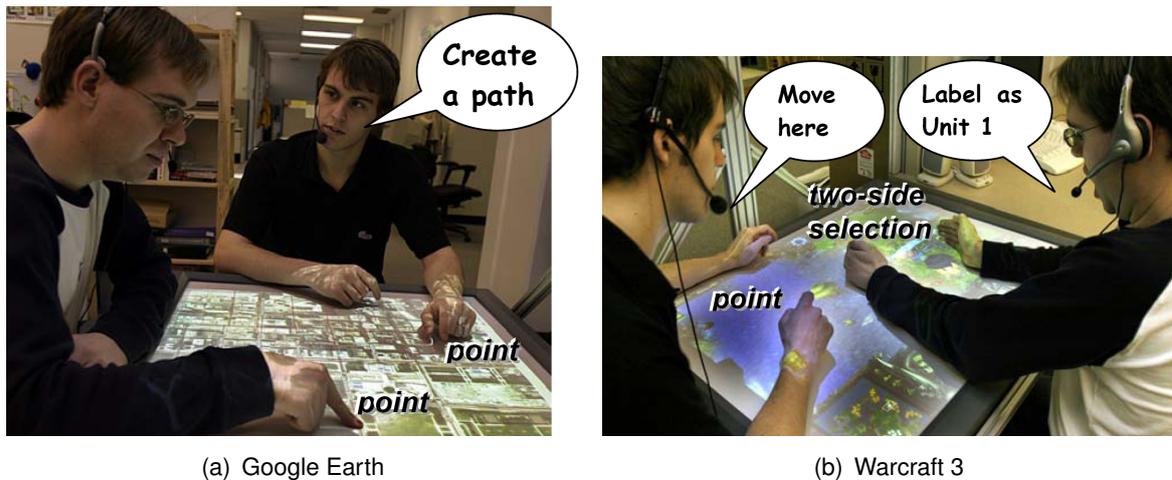


Abbildung 3.14.: Interaktion mit Sprache

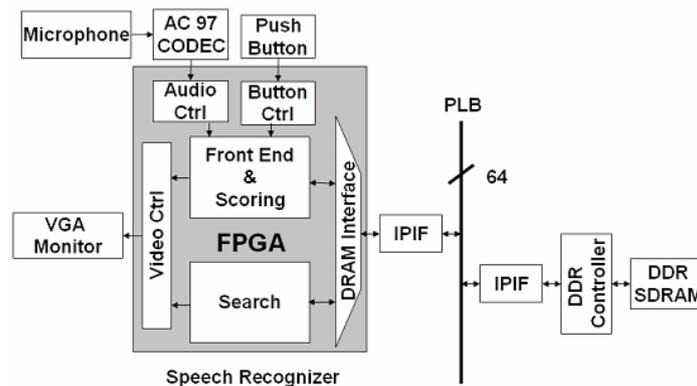


Abbildung 3.15.: FPGA-Architektur des „In Silico Vox“-Projektes

Entwurf des multimodalen Interaktionssystems statt und es werden zur Konzeptvalidierung benötigte Grundfunktionen für die multimodale Interaktion implementiert. Dies erfordert die Konkretisierung der Interaktionsmodalität bzw. dessen Umsetzung – Es erfolgte die Betrachtung vorhandener Eingabesysteme, welche die beiden unterschiedlichen Eingabekonzepte, basierend auf den Modalitäten Zeigegestik und Spracheingabe, anwenden.

Zeigegestik Es gibt Lösungen, um Zeigehandlung zu erkennen. In der Analyse konnten vier Lösungsansätze charakterisiert werden: Unter das optische, statische Tracking fallen Tracking des Eingabemediums und Tracking auf dem Ausgabemedium: Für statisches Tracking wird ein Kameraaufbau benötigt. Befindet sich dieser vor der Leinwand, wird der Bewegungsfreiraum stark eingeschränkt, die Umsetzung mit mehreren Eingabegeräten gestaltet sich schwierig. Ein Aufbau hinter der Anzeigefläche setzt grossflä-

chige Räume voraus. Eine statischer Aufbau beschränkt aufgrund der benötigten Infrastruktur die Portierbarkeit des Gesamtsystems.

Bei der Verwendung einer Kamera als Eingabegerät gibt es die direkte und die indirekte Manipulation der Anzeigefläche: Abgeleitet von der natürlichen Zeigegestik sollte die direkte Bestimmung eines gemeinten Objektes oder Bereiches Entwicklungsziel sein, eine indirekte Lösung entspricht nicht diesem Verhalten. Die Verwendung einer Kamera bedarf der direkten Sichtverbindung zur Anzeigefläche. Diese Anforderung ist annehmbar, da bei der realen Zeigegestik zwischen mehreren Personen ebenfalls ein direkter Sichtkontakt besteht.

Spracheingabe Die Unterstützung der Eingabe durch Sprache ist bereits in diversen Projekten umgesetzt worden und hat sich als praktikabel herausgestellt. Unter Verwendung eines definierten, eingeschränkten Wortschatzes ist die Anwendbarkeit ausreichend gegeben. Durch die vorhandenen Erfahrungen, Tools und Bibliotheken kann dieses Gebiet als erschlossen angesehen werden, im Folgenden wird die Interaktion durch Sprache als verwendbare Interaktionsmodalität behandelt und als bestehende Systemkomponente eingesetzt.

Im Realisierungskapitel (4) wird technologisch untersucht, wie auf Basis einer Kamera als Zeigeobjekt die Interpretation von Zeigegestiken umsetzbar ist.

4. Entwicklung eines Zeigegerätes

4.1. Der interaktive Raum

Der interaktive Raum ist die Umgebung in der sich die Benutzer bewegen und miteinander arbeiten. Die Beschreibung erklärt den Aufbau dieses Areals, um die Notwendigkeit des geforderten Interaktionsgerätes zu verdeutlichen. In Abbildung 4.1 ist der Grundriss des Computerlabors im Rahmen des UbiComp-Projektes an der HAW Hamburg dargestellt.

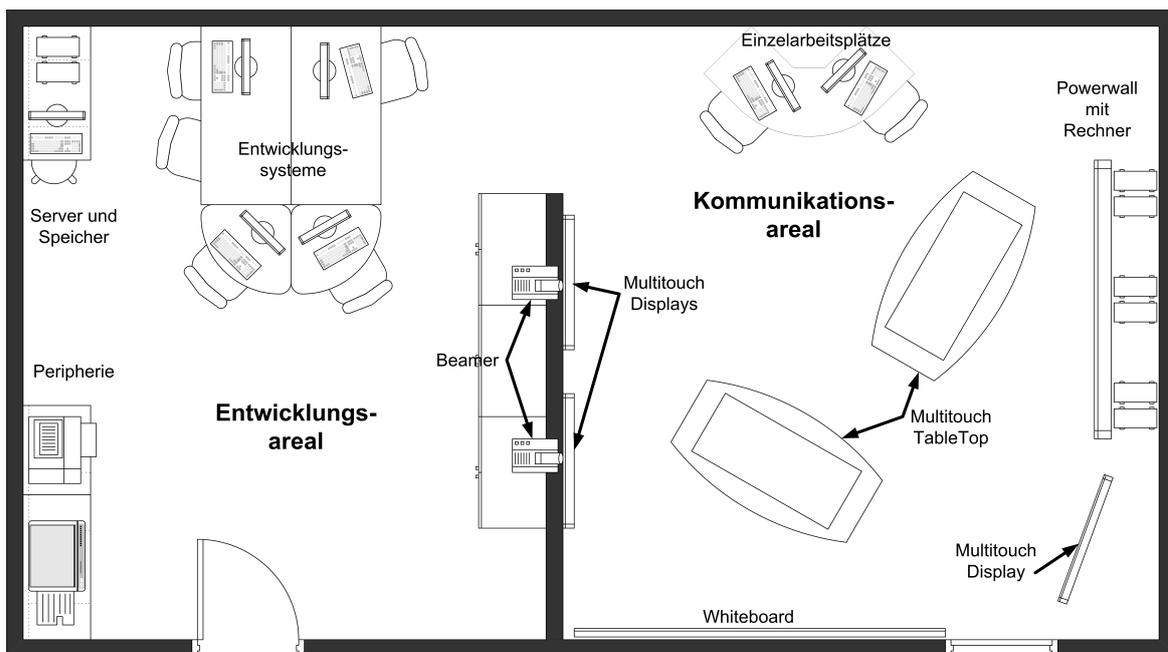


Abbildung 4.1.: Der Interaktive Raum, Grundriss

Grundsätzlich ist der Raum in zwei Areale aufgeteilt, die Entwicklungsarbeit wird in dem linken Bereich geleistet. Die örtliche Nähe zum unmittelbaren Anwendungsbereich ist gewollt und entspringt einer „ideologischen“ Einstellung, die sehr gut von [Johanson \(2003\)](#) wiedergegeben wird:

„**Practice what we preach**“ – Die Entwicklung in dem eigentlichen Anwendungsgebiet stattfinden zu lassen (Entwicklertreffen, Brainstorming, Implementierung, etc.) ist eine sehr effektive Möglichkeit, Ideen und Konzepte auf Gültigkeit zu prüfen. Fragestellungen der Art: „Ist die Anwendung überhaupt praktikabel?“ können umgehend durch Tests sehr konkret beantwortet werden.

Johanson (2003) benennt weitere Leitsätze für die Erstellung seines „iRoom“ (s. Abb. 4.2), von denen uns ein paar beim Aufbau motiviert haben:



Abbildung 4.2.: Zweite Version des „iRoom“

„**Mechanization, not-automation**“: Die in dem Umfeld befindlichen Entwicklungen und Werkzeuge sind keine statischen Komponenten bestimmter Anwendungen, sondern sollen dem Benutzer die Option geben, diese in Abhängigkeit von seiner Aufgabe selbst zu wählen. Dieser Ansatz wird ebenfalls im SmartKom-Projekt gefordert und als „Benutzeradaptiv“ bezeichnet (vgl. Kapitel 2.2).

Eine Grundlage für dieses Vorgehen ist die Entwicklung mit dem Gesichtspunkt „**Wide applicability**“. Dies bedeutet, erstellte Anwendungen sollen nicht nur an einem bestimmten Szenario gebunden werden, sondern auch portierbar in andere Anwendungsfälle sein.

Ubiquitäres Computing beschreibt die Allgegenwärtigkeit von Rechenleistung, kann aber unter zwei Gesichtspunkten realisiert werden, wie bereits im Grundlagenkapitel 2.2 zum Thema „Nutzerzentrierung“ erörtert. Der Benutzer steht im Mittelpunkt und bestimmt die zu bewältigende Aufgabe – der Raum ermöglicht durch die integrierte Technik die Realisierung und nicht anders herum, also dass durch die Ausstattung der Umgebung die Aufgabe festgelegt wird. Johanson (2003) nennt diesen Gedanken „**Reliance on social conventions**“.

Sämtliche Ein- und Ausgabekomponenten sind in einem gemeinsamen Netzwerk miteinander verknüpft und können somit von allen Systemen und Personen genutzt werden.

Das gleiche gilt für Dienste: Im folgenden Kapitel 4.2 werden alle aktuell in dieser Projektumgebung stattfindenden Arbeiten aufgezeigt. Ein paar diese Arbeiten haben die Umsetzung von Diensten zur Etablierung einer geeigneten Infrastruktur als Ziel.

Peripheriegeräte und Server befinden sich im Entwicklungsbereich, da eine unmittelbare physikalische Interaktion mit ihnen nicht notwendig ist und das Konzept der „unsichtbaren“ Computer im Kommunikationsbereich (rechts)gewahrt bleibt.

Serverseitig sind (über RAID¹) redundant ausgelegter Netzwerkspeicher vorhanden, ein Blade-Server² „IBM BladeCenter HS20³“ sorgt mit fünf Rechereinheiten für die benötigte Rechenleistung.

Das Kommunikationsareal liefert diverse Interaktionsbereiche und bietet somit genügend Spielraum für verschiedene Gruppenaktivitäten:

Powerwall⁴ Eine Powerwall von 3x3 42“ „High Definition⁵“-Displays wird über fünf Hochleistungsrechner angesteuert (s. Abb. 4.3) („Shared Space“).



Abbildung 4.3.: Powerwall des UbiComp-Projektes

Beamer Alternativ zum Einsatz der Powerwall besteht die Möglichkeit, Beamer einzusetzen.

¹RAID (Redundant Array of Independent Disks) – System zur Organisation mehrerer physikalisch vorhandener Festplatten zu einem logischen Laufwerk

²Blade-Server – spezielle Bauform von Rechnern, die übereinander in einem Baugruppenträger liegen

³IBM BladeCenter HS20 – <http://www-03.ibm.com/systems/bladecenter/hs20/index.html>

⁵„High Definition“-Displays sind hochauflösende Monitore mit 1920 x 1080 Pixeln

TableTop Zwei TableTop-Systeme ermöglichen Tisch-basierende Gruppen- oder Einzelanwendungen. Als Anzeigefläche kommt jeweils ein 48" „multi touch“-fähiges HD-Display zum Einsatz und wird in eine Tischplatte integriert (s. Abb. 4.4) („Shared Space“).



Abbildung 4.4.: Table-Top des UbiComp-Projektes

Klassisches Display Analog zu den horizontal installierten TableTop-Systemen gibt es Ausführungen der Panels mit Touchfunktionalität als freistehende, vertikal ausgerichtete Einheiten und als an der Wand befestigt.

PC-Stationen Zwei einzelne Desktop-Arbeitsplätze ermöglichen die Abgrenzung von persönlichen Bereichen („Personal Space“).

4.2. Logischer Entwurf

Aktuell gibt es sechs Arbeiten im Rahmen des CSCW-Projektes an der HAW Hamburg. Die von Philipp Roßberger erstellte Abbildung (4.5) zeigt eine serviceorientierte Architektur des Gesamtprojektes. Die Arbeiten lassen sich folgendermaßen in diese einordnen: Grundlegende serverseitige Dienste sind das „Verteilte Persistenz-System“ von [Kabalkin \(2007\)](#) und die Arbeit „Suchen und Finden im Collaborative Workspace“ von [Pein \(2007\)](#).

In unmittelbarem Zusammenhang stehen die Arbeiten „Physikbasierte Interaktion im Collaborative Workspace“ von [Roßberger \(2007\)](#) und die „Entwicklung einer Interaktionskomponente in Collaborative Workspaces“ vom Autor [Fischer \(2007\)](#). Hinzu kommt die Arbeit von

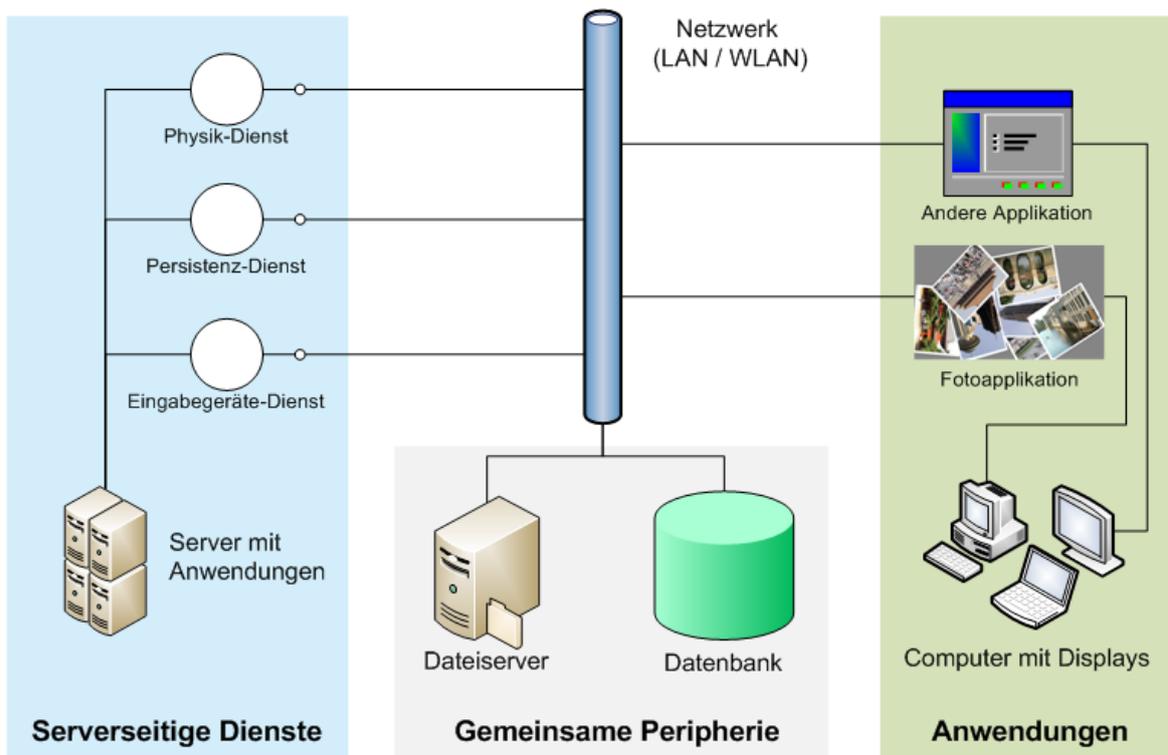


Abbildung 4.5.: Serviceorientierte Architektur des HAW Projektes CSCW

Köckritz (2007): „Virtueller Collaborative Workspace mit Remote Windows“. und der Bericht „Multihead-Display als interaktives Informationssystem“ von Napitupulu (2006). Das Zusammenspiel dieser Arbeiten, die Modularisierung und die damit verbundenen Abhängigkeiten können in der Model-View-Control-Architektur dargestellt werden.

Das ursprünglich nur für Softwaresysteme konzipierte Designmuster stammt originär aus Smalltalk-80 und wird in vielen Anwendungsbereichen verwendet, was die damit verbundene Definitionenvielfalt der einzelnen Bestandteile nach sich zieht. An dieser Stelle ist eine Konkretisierung der Module notwendig.

Nach Helman (1998) werden die Aufgaben der einzelnen Komponenten klar definiert, grundsätzlich erfolgt eine Modularisierung des Gesamtsystems in die Komponenten Eingabe, Bearbeitung/Steuerung und Ausgabe, abstrahiert sind diese Control, Model und View (vgl. Abbildung 4.6). Zwischen den Elementen gibt es präzise statische Verbindungen, diese Referenzierungen sind in Abbildung 4.7 zu sehen. Für definierte Kommunikationsschnittstellen bestehen zwischen allen Komponenten Basisverbindungen, weitere Beziehungen sind nur begrenzt vorhanden:

Model Das Modell verwaltet Daten bzw. Informationen und definiert darauf anwendbare

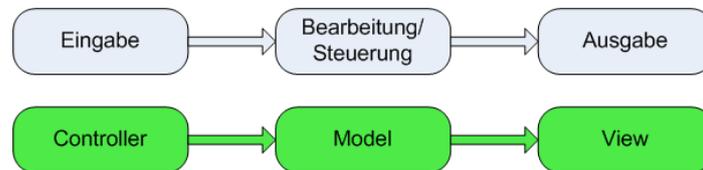


Abbildung 4.6.: Model-View-Control - Abstraktion

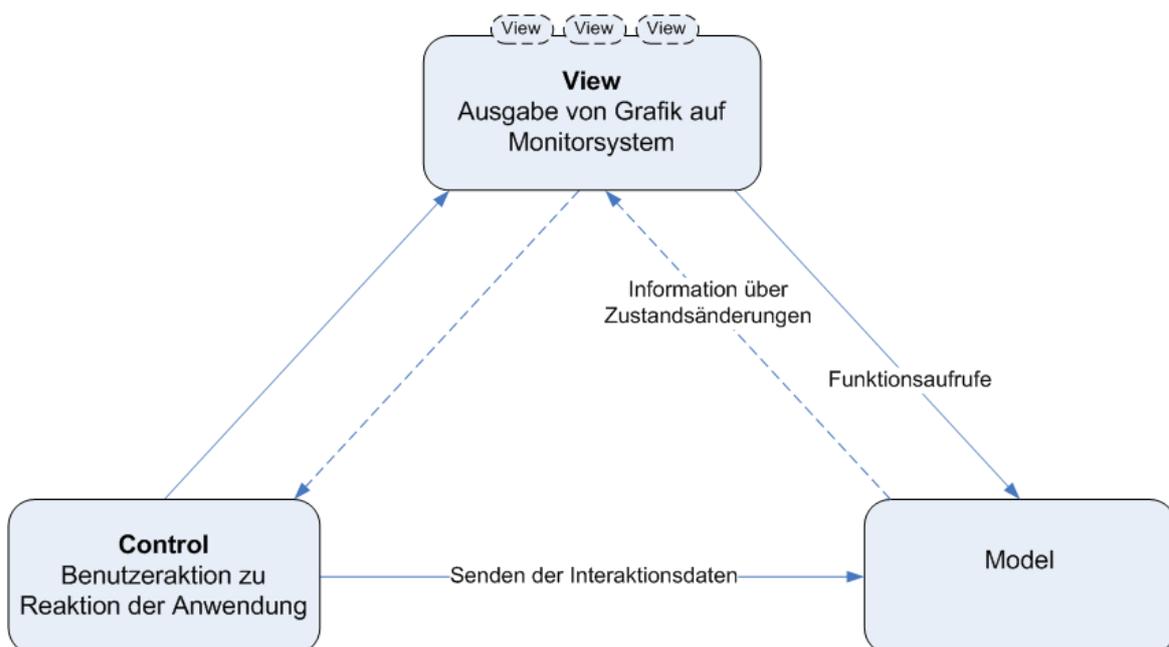


Abbildung 4.7.: Model-View-Control - Relationen

Funktionen. Es koordiniert die Systemzustände sowie Prozessabläufe, bei Änderungen wird die betroffene View darüber informiert. Das Modell hat keinen Einblick in die exakte Umsetzung/Implementierung der Viewports.

Eine beispielhafte Anwendung und die damit verbundene Steuerung der Systemzustände und Systemprozesse wird von Philipp Roßberger realisiert. Konkret entsteht eine mehrbenutzerfähige Arbeitsfläche, auf der Informationen in Form von Bildern abgelegt werden können. Hervorzuheben ist die Gestaltung des Arbeitsplatzes – die Manipulationseigenschaften der Objekte richten lassen bekannte Interaktionen zu (Natur, Desktopmetapher), insbesondere können Objekte über sechs Freiheitsgrade bewegt werden. Bewegungen von Objekten werden der natürlichen Physik entsprechend unterstützt, dies soll einer intuitiven Bedienung entgegenkommen. Abbildung 4.8 zeigt einen Screenshot der Anwendung.



Abbildung 4.8.: Modell - Anwendung von Philipp Rossberger

View Die Aufgabe ist das Rendern der Grafik des Modells für die Displayoberflächen. Die View kann mehrere Subviews instantiiieren, wenn beispielsweise mehrere Ausgabegeräte zum Einsatz kommen, um deren Inhalt individuell anzupassen. Der View ist bekannt, wie das sich bedienende Modell gestaltet ist, alle Funktionen sind aufrufbar. Wird es vom Modell über Zustands-, Prozess- oder Datenänderungen informiert, holt es sich die benötigten Daten, um betroffene Bereiche der Anzeige neu zu rendern. Die

Verbindung zum Controller bleibt auf Basisebene. Um die Möglichkeit eines Controller-Tausches zu gewährleisten, werden die Abhängigkeiten minimal gehalten.

Die Darstellungsebene und die damit verbundene verteilte Anzeige der Informationen aus dem Modell wird durch Oliver Köckritz und Jan Napitupulu behandelt. Abbildung 4.9 zeigt die Verteilung einer Ansicht auf einer Powerwall, Abbildung 4.10 zeigt die Verteilung einer Anwendung auf zwei Computersystemen.



Abbildung 4.9.: View - Verteilte Ansicht von Jan Napitupulu

Controller Der Controller bestimmt das Verhalten der Triade und muss daher beide Typen genau kennen. Aufgabe ist die Übersetzung der Benutzereingabe zum respektiven Verhalten der Anwendung. Abhängig vom aktuellen Zustand der Anwendung werden aus Eingabedaten Prozessdaten für das Modell generiert.

Durch den Controller wird die Interaktion mit dem Gesamtsystem umgesetzt. Dazu gehört die Einbindung der Interaktionsgeräte, die Datenaufbereitung und -auswertung in Abhängigkeit vom aktuellen Systemzustand. Die Funktion des Controllers ist Gegenstand dieser Arbeit und wird detailliert im folgenden Kapitel 4.2.1 behandelt.

4.2.1. Controller: Technologische Ebene

Die Aufteilung der Controller-Komponente in Ebenen ist notwendig, um eine geeignete Modularität des Eingabegerätes zu gewährleisten. Dadurch werden Skalierungs- und Erweite-



Abbildung 4.10.: View - Verteiltes Ping-Pong von Oliver Köckritz

rungsmöglichkeiten auf weitere Geräte oder Szenarien (bzw. Anwendungen) erreicht. Zweckmässig aufteilen lässt sich der Controller (im Sinne des MVC-Musters) in die Module: hardware-spezifische Datenerfassung, neutrale technische Datenaufbereitung und anwendungsspezifische Datenauswertung.

Erfassung der Gerätedaten

Auf dieser Ebene werden die Daten des bzw. der Eingabegerät(e) erfasst und dem System zur Verfügung gestellt. Dazu werden geeignete Datenpakete über eine definierte Kommunikationsschnittstelle zur Aufbereitung übergeben.

Aufbereitung der Gerätedaten

Diese Ebene ist das Bindeglied zwischen eingabespezifischer Datenerfassung und anwendungsspezifischer Datenauswertung. In beide Richtungen werden Schnittstellen mit eindeutigen Protokollen definiert, wodurch die Modularität des Gesamtsystems ermöglicht wird.

Die Datenstrukturen zur anwendungsspezifischen Auswertung beinhalten, wie in Kapitel 3.2 beschrieben, folgende Informationen: Zeigerpositionierung, Orientierung, Lokalisierung, Wechselwirkung, Bewegung und Haptik.

Die Schnittstellenbeschreibung zur Datenerfassungsebene ist direkt vom Interaktionskonzept abhängig. Erfassungs- und Aufbereitungsschritte werden so lange getrennt abgearbeitet, also auf der ersten Ebene angewandt, wie sie unterschiedlich sind. Kommen gleiche Algorithmen zu Einsatz, ist die Verbindung dieser auf Aufbereitungsebene sinnvoll.

Auswertung der Interaktionsdaten

Die anwendungsspezifische Ebene umfasst zwei Schritte:

Logische Aufbereitung der Eingabedaten: Abhängig von der Applikation werden nur die Daten gefiltert, die als Interaktionsdaten in Frage kommen. Als Kernfrage stellt sich hier: Welches Verhalten soll das System als eine vom Benutzer intendierte Interaktion analysieren? Zum Beispiel: Eine schnelle Bewegung nach unten kann auf drei Arten interpretiert werden. Als schnelle Bewegung nach unten, als keine Interaktion (der Arm wurde vom Benutzer hingelassen), oder als eine Aufforderung die Anwendung auszuschalten (Eingabegerät wurde abgelegt).

Die zweite Stufe ist die logische Interpretation der validierten Eingabedaten und bereitet die Modelldaten für die Übergabe an das Modell vor (vgl. 4.2). Situationsbedingt bzw. abhängig vom Zustand der Anwendung und der Ausgabe durch die View werden Eingabemöglichkeiten definiert. Darauf basierend können Eingabedaten verworfen werden oder bestimmte Bewegungsabläufe als eine Eingabe interpretiert werden. Wird beispielsweise eine Videosequenz abgespielt, ist eine Zeigehandlung nicht möglich – Zeigedaten werden ignoriert. Um diese Sequenz abzubrechen, kann eine kreisförmige Bewegung ausgeführt werden, diese Bewegung würde erkannt und akzeptiert werden.

4.3. Machbarkeitsstudie

In diesem Abschnitt wird untersucht, in wieweit sich die Entwicklung und damit das Konzept eines Interaktionsgerätes auf Basis einer Kamera als Eingabemöglichkeit gemäß den Anforderungen aus 3.3 prinzipiell realisieren lässt. Hierzu erfolgt die hard- und softwareseitige Entwicklung eines Prototypen. Ziel ist nicht die vollständige Realisierung des Gerätes – in erster Linie stellt sich die Frage nach der Umsetzbarkeit einer bestimmten Kernfunktionalität. Grundlage für die in Kapitel 3.2 erarbeiteten Eingabemetaphern und gleichzeitig primäre Interaktionsanforderung ist die Zeigegestik (z.B. Selektion und Manipulation von Objekten).

Die freie Interaktion im interaktiven Raum (vgl. Kapitel 4.1) setzt die nahtlose Übertragung zwischen mehreren Systemen und damit mehreren Displays voraus. Für die Validierung dieses Konzeptes wird der Einsatz vorerst auf die Installation eines Displays beschränkt.

Zum Ziel hat diese Implementierung das beliebige Bewegen eines Objektes auf einem Display, gemäß der Desktopmetapher entspricht dies der Umsetzung eines Mauszeigers. Hierbei soll der Benutzer in seiner räumlichen Positionierung nicht eingeschränkt werden, die Interaktion muss aus beliebigem Winkel (vor der Anzeigefläche) und beliebiger Entfernung (entsprechend eines Raumes) vollzogen werden können.

4.3.1. Vorgehensweise

Die Kamera wird auf den zu interagierenden Bereich gerichtet, der Mittelpunkt des Kamerabildes entspricht dem Zeigepunkt. Um dessen Koordinaten im Bereich der Anzeigefläche und daraus die absolute Mausposition bestimmen zu können, muss die Anzeigefläche im Bild identifiziert werden. Dies wird durch die markante Hervorhebung der Ränder erreicht. Um eine eindeutige Identifikation bereits auf dieser Ebene zu erreichen, werden in den Ecken angebrachte Infrarotdioden eingesetzt. Ein Infrarotsperfilter vor dem Objektiv der Kamera filtert das Licht nach Wellenlänge und verhindert das Eindringen des sichtbaren Anteils des Lichts. Mit Hilfe von Algorithmen der Bildmustererkennung lassen sich diese „Tags“ effektiv und eindeutig extrahieren. Aus den gewonnenen Informationen (vier Eckpunkte und Zeigepunkt) lässt sich die absolute Position des Zeigepunktes auf dem Bildschirm geometrisch ermitteln.

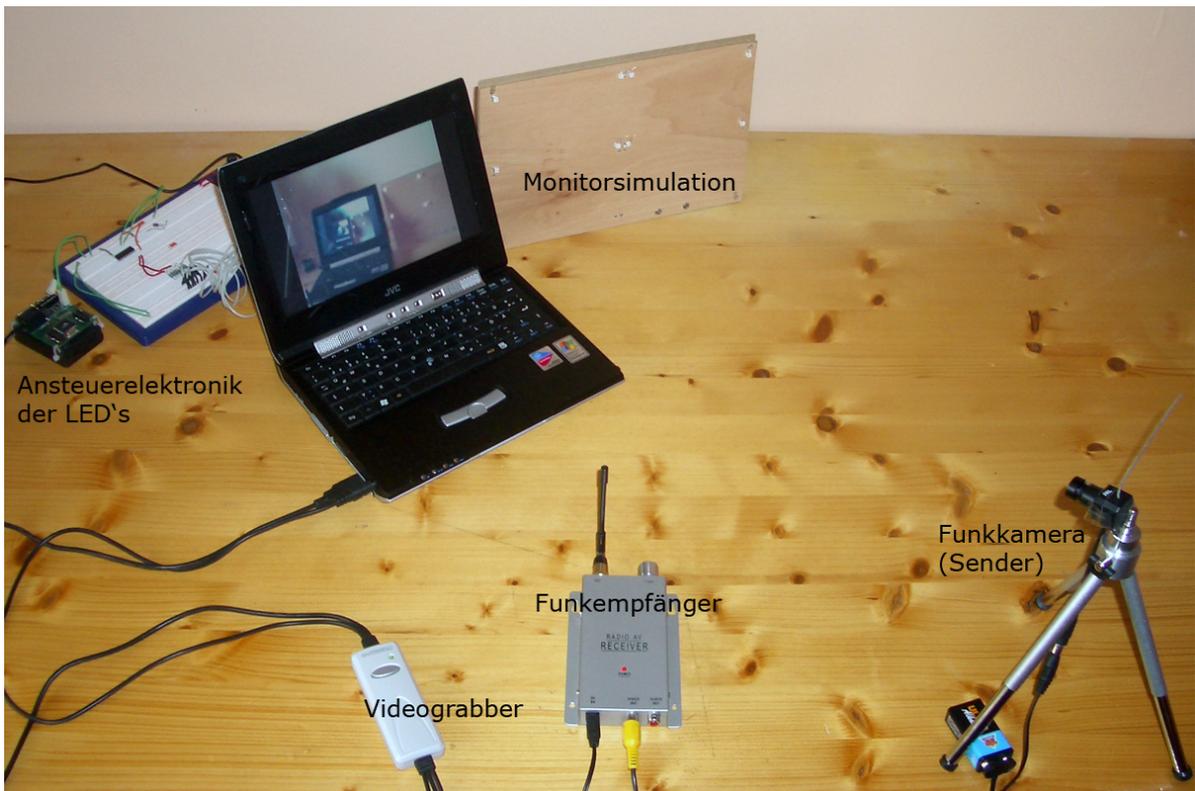
4.3.2. Hardware und Entwicklungsumgebung

Eingesetzt wird eine handelsübliche Miniaturkamera mit einer Auflösung von 380 TV Linien. Die Kommunikation erfolgt via Funk auf 2,4GHz, ein Empfänger stellt das Signal über Composite Video (FBAS) zur Verfügung und eine Video-Grabberkarte übersetzt dieses dem Computer über USB mit einer maximalen Abtastrate von 25 Bilder pro Sekunde in einen Videostream.

Die Tags am Bildschirmrand bilden Standard-Infrarot-Leds mit einer kurzwelligen Strahlung von 950 nm bei 100 mA und einer Sperrspannung von 5 V, die Filterung des IR-Lichts wird durch handelsübliche IR-Folie ermöglicht.

Die Auswertung und Berechnung erfolgt auf einem Notebook mit 1,8 GHz Centrino-Prozessor und 1 GB RAM. Als Betriebssystem wird Microsoft Windows XP SP2 eingesetzt.

Der gesamte Aufbau ist in Abbildung 4.11 zu sehen. Zu entwickeln sind zwei Softwareapplikationen: das Modul der Videoaufnahme und -auswertung sowie eine beispielhafte Anwendung, welche die ausgewerteten Daten grafisch darstellt (z.B. in Form eines Mauszeigers).



(a)



(b) mit Infrarotfilter

Abbildung 4.11.: Testaufbau

Programmiersprache Die Auswertung des Videostreams ist zeitkritisch, was die Verwendung von C++ im Bereich der Video-Zerlegung und -Aufarbeitung sowie der Dateninterpretation rechtfertigt. Die grafische Oberfläche wird mit Java erstellt.

Bibliotheken Das Zerlegen des Videostreams in Einzelbilder erfolgt durch „IC Imaging Control“ von Imaging Source⁶. Es handelt sich um eine Bibliothek in C++.

Aufgrund eigener Erfahrungen und des vorhandenen Know-Hows an der HAW Hamburg wird für die Bilderverarbeitung bzw. Bildmustererkennung die „LTI-Lib“ der Technischen Hochschule Aachen⁷ eingesetzt. Diese Bibliothek ist in C++ verfügbar und kann bequem mit IC Imaging Control verknüpft werden.

Zur Gestaltung der grafischen Benutzeroberfläche wird das Standard Widget Toolkit (SWT) von Eclipse⁸ eingesetzt.

Tools Mit C++ bietet sich unter Windows als Entwicklungsumgebung Microsoft Visual Studio 2003 an (VS2005 ist nicht interessant, da die LTI-Lib für VS2003 implementiert ist).

Die Java-Komponenten werden unter Eclipse⁹ entwickelt.

4.3.3. Systemaufbau

Die Entwicklung des Prototyps erstreckt sich auf drei Ebenen: die Datenerfassung, die Aufbereitung und die Auswertung. Der Fokus ist auf die Datenerfassungs- und die Aufbereitungsebene gerichtet, also die hardwaretechnische Umsetzung und Videoanalyse. Die grafische Oberfläche ist ein Fenster, in dem ein Mauszeiger bewegt und ein dreidimensionaler Würfel gedreht werden kann. Die Daten des Interaktionsgerätes werden via UDP¹⁰ an den GUI¹¹-Rechner gesendet. Die Auswertungsebene ist nicht Gegenstand des Prototypbaus und wird daher nicht näher betrachtet.

Datenerfassung

Die Datenerfassung ist ein Prozess von zwei Schritten, Abbildung 4.12 veranschaulicht dies:

⁶IC Imagine Control von Imaging Source: <http://www.imagingcontrol.com/>

⁷LTI-Lib der Technischen Hochschule Aachen: <http://www.techinfo.rwth-aachen.de/Software/Ltilib/>

⁸Standard Widget Toolkit: <http://www.eclipse.org/swt/>

⁹Eclipse (<http://www.eclipse.org>) - Software-Plattform und integrierte Entwicklungsumgebung

¹⁰UDP (User Datagram Protocol) – verbindungsloses Netzprotokoll in der Transportschicht

¹¹GUI (Graphical User Interface) – grafische Benutzeroberfläche

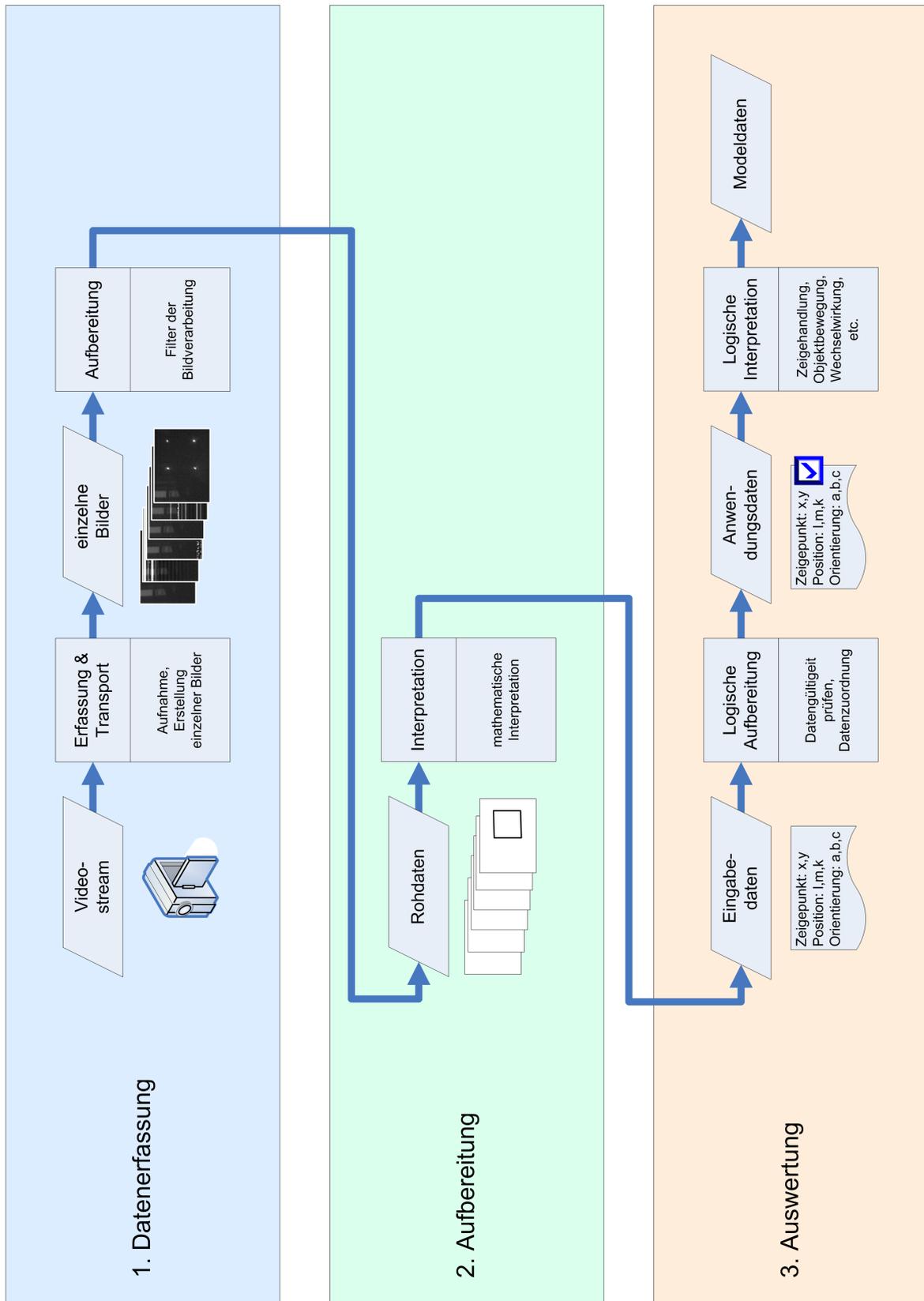
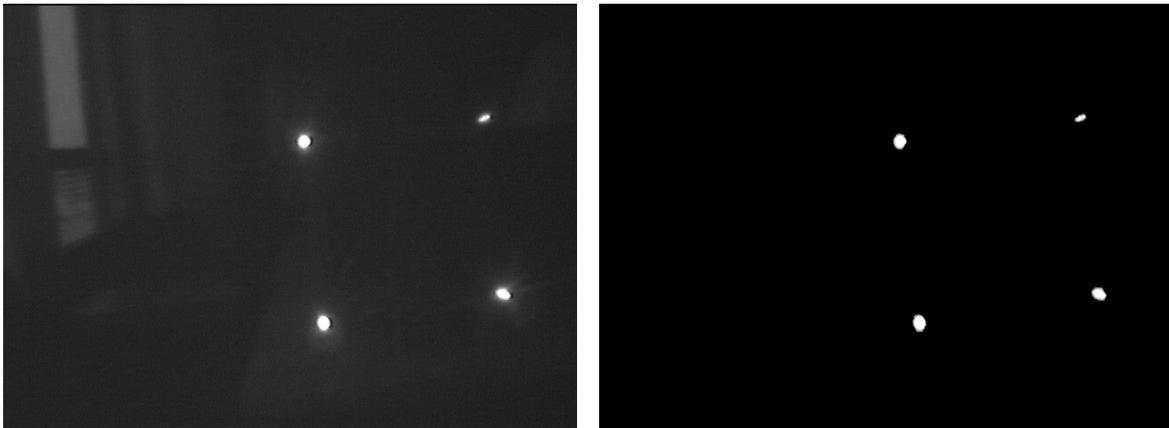


Abbildung 4.12.: Schichtmodell des Systems

1. Erfassung und Transport: Die Kamera nimmt den gezeigten Bereich als Video auf und transportiert ihn via Funk zum Empfänger. Der Grabber stellt dem PC-System den Stream zur Verfügung, zerlegt in einzelne Bilder können die Daten weiterverarbeitet werden.
2. Aufbereitung: Das einzelne Bild wird aufbereitet, um effektiver visuelle Informationen hergeben zu können:

Bildverarbeitung Mit Hilfe der „IC Imagin Control“-Bibliothek können einzelne Bilder des Videostreams extrahiert werden. Da sie nacheinander bearbeitet werden, hängt die mögliche Framerate von der Verarbeitungs- bzw. Berechnungsgeschwindigkeit des Gesamtsystems ab. Nach einer Grauwertabstufung auf 8 Bit (mit LTI-Lib) zeigt Abbildung 4.13(a) ein einzelnes Frame, wie es direkt von der Kamera mit montiertem Infrarotsperfilter erfasst wird.



(a) Kamera-Rohbild

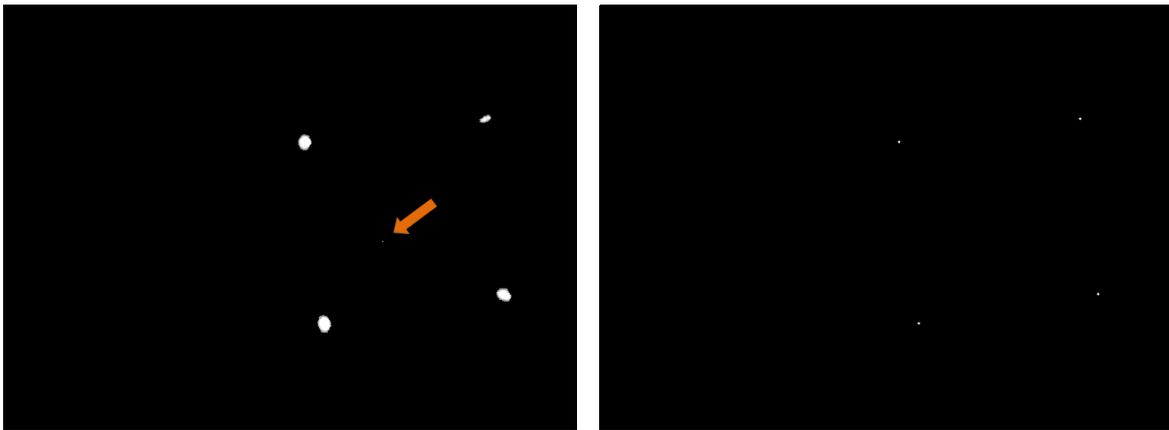
(b) nach Anwendung des „optimalen Schwellwertes“

Abbildung 4.13.: Bildverarbeitung: Schwellwerte

Neben den im Bild deutlich zu erkennenden vier Infrarot-LEDs fallen weniger helle Lichtschleiern und andere Lichtquellen, wie in diesem Beispiel Tageslicht, auf. Um die Datenmenge auf die wesentlichen Informationen zu reduzieren und somit die vier Punkte eindeutig hervorzuheben, ist die Anwendung eines Filters nützlich. Durch die Validierung gegen einen geeigneten Schwellwert bleiben die wesentlichen Informationen im Bild bestehen, der Rest wird schwarz. Nach [Sonka u. a. \(2007\)](#) eignet sich dafür die Berechnung und Anwendung des „optimalen Schwellwertes“. Basierend auf dem Grauwert-Histogramm wird der Schwellwert festgelegt, das Ergebnis ist in Abbildung 4.13(b) zu sehen.

Extraktion Mit Extraktion ist hier nicht das bekannte Verfahren für Bildmustererkennung gemeint, sondern die Berechnung der konkreten vier einzelnen Mittelpunkte

te des Bildes. Angewandt wird die geometrische Momentenberechnung, welche die einzelnen Schwerpunkte liefert. Es wird wie folgt vorgegangen: Durch eine Schwerpunktberechnung (basierend auf den Farbwerten) auf das gesamte Bild erhält man den „Mittelpunkt“ zwischen den vier Eckpunkten, zu sehen in Abbildung 4.14(a).



(a) Einzelnes Bild mit Gesamtschwerpunkt

(b) Einzelnes Bild mit vier Schwerpunkten

Abbildung 4.14.: Bildverarbeitung: Schwerpunkte

Mit einer gedachten horizontalen und einer vertikalen Linie durch den Mittelpunkt kann die Fläche in vier Sektoren aufgeteilt werden. Die erneute Anwendung der Schwerpunktberechnung auf jeden einzelnen Quadranten ergibt die einzelnen Schwerpunkte der Eckpunkte, siehe Abbildung 4.14(b).

Das Ergebnis sind die Koordinatenpaare der LED-Spots im Koordinatensystem der Anzeigefläche, was die Anwendung von mathematischen/geometrischen Formeln ermöglicht.

Datenauswertung

Die Datenaufbereitung (Stufe 2 nach Abbildung 4.12) erfolgt in einem Schritt, bezeichnet als mathematische Interpretation. Aus den Informationen des Datensatzes (Bildinformationen) lassen sich Eigenschaften bezüglich der Haltung (Position, Orientierung) des Eingabegerätes gewinnen:

Im Verbund bilden die Koordinaten der Eckpunkte ein verzerrtes Viereck, welches das Display repräsentiert. Hieraus lassen sich Informationen über die Verwendung des Eingabegerätes errechnen:

Zeigerpositionierung Wie bereits oben erwähnt, bildet der Bildmittelpunkt des Kamerabildes den Punkt, auf den gezeigt werden soll. Das durch die vier Eckpunkte aufgespannte Viereck ist perspektivisch verzerrt, da der Benutzer das Eingabegerät von beliebiger Position halten kann. Das Viereck ist in einem geradlinigen Koordinatensystem aufgespannt, die Koordinatenachsen sind jedoch nicht orthogonal zueinander, die Abbildung des Vierecks erfolgt im Koordinatensystem der Bildebene, siehe 4.15. Punkt P markiert den Mittelpunkt des Kamerabildes.

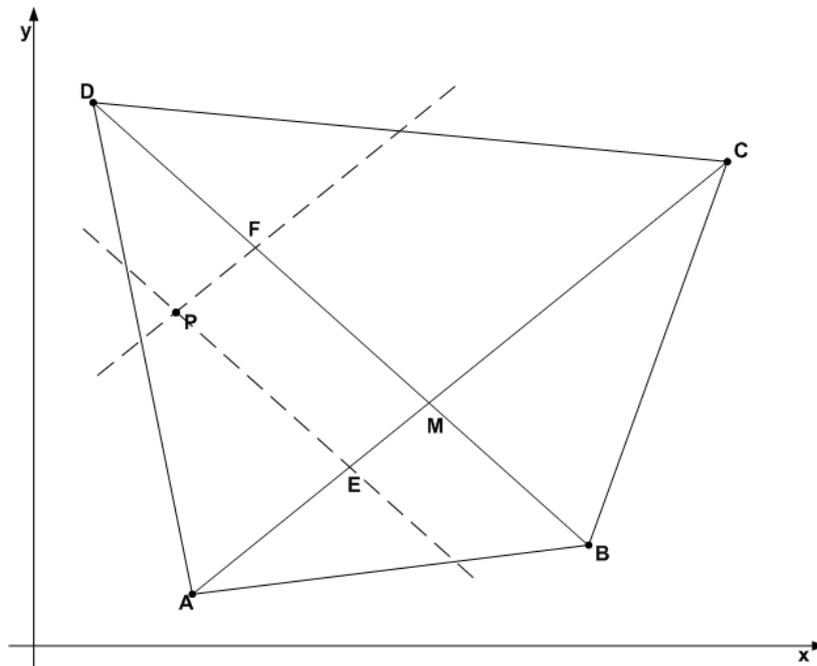


Abbildung 4.15.: Perspektivisch verzerrter Bildschirm

Der erste Schritt ist eine Transformation der Bildpunkte in ein kartesisches Koordinatensystem. Das Zielobjekt ist in Abbildung 4.16 zu sehen. **Hinreichende Annahme:** Für die Implementierung eines Prototyps reicht eine näherungsweise geometrische Berechnung aus, daher können die Verhältnisse der Eckpunkte zum Mittelpunkt in beiden Vierecken als gleich betrachtet werden, folgende Gleichung soll zur Berechnung der Koordinaten des Punktes P gelten:

$$P'.x = \frac{\overline{A'M'}}{\overline{A'E'}} * \overline{A'B'}$$

$$P'.y = \frac{\overline{B'M'}}{\overline{B'F'}} * \overline{B'C'}$$

mit

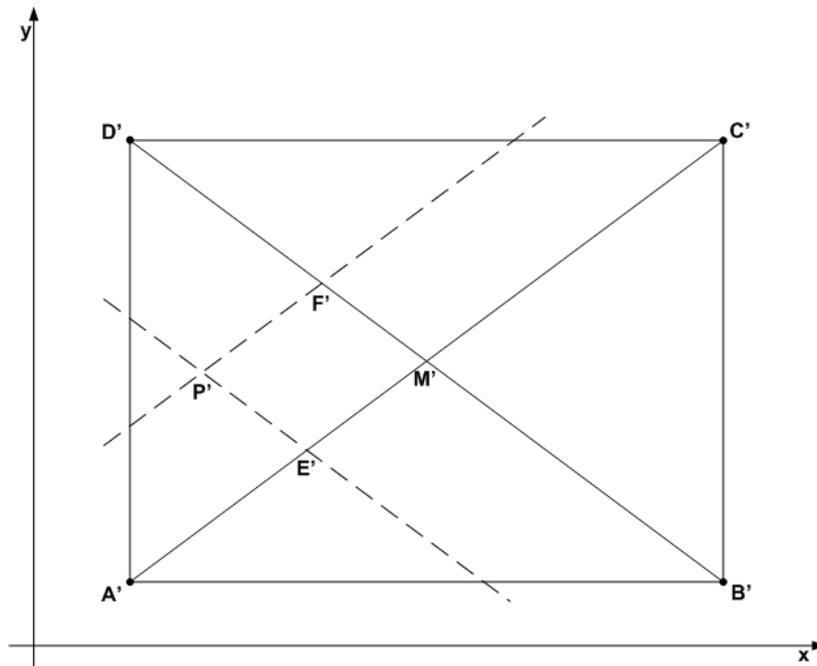


Abbildung 4.16.: Entzerrter Bildschirm

$$\overline{A'M'} = \sqrt{\overline{A'B'}^2 * \overline{B'C'}^2}$$

$$\overline{A'E'} = \frac{\overline{AE}}{\overline{AM}} * \overline{A'M'}$$

$$\overline{A'B'} = \text{Displayauflösung.x}$$

$$\overline{B'M'} = \overline{A'M'}$$

$$\overline{B'F'} = \frac{\overline{BF}}{\overline{BM}} * \overline{B'M'}$$

$$\overline{B'C'} = \text{Displayauflösung.y}$$

Die Displayauflösung wird auf 1.0 mit vier Nachkommastellen normiert.

Orientierung Die Orientierung des Eingabegerätes wird durch die Winkel zu der horizontalen und der vertikalen Ebene angegeben. Dies ermöglicht eine Auswertung der Orientierung im dreidimensionalen Raum. Wie bei der Zeigerpositionierung ist die folgende Berechnung lediglich eine grobe Annäherung. In [Abbildung 4.17](#) ist das rotierte Viereck zu sehen, stellt man sich eine Ebene vor und lässt diese um die x-Achse (α) und um die y-Achse (β) rotieren, gelten aufgrund der „hinreichenden Linearität“ folgende Gleichungen:

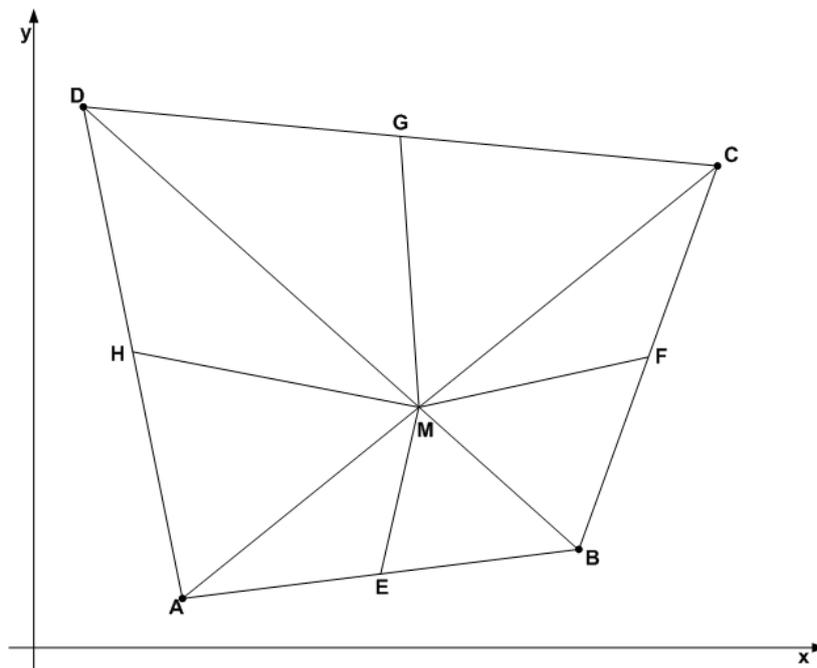


Abbildung 4.17.: Rotiertes Rechteck

$$\alpha = 180.0 * \frac{\overline{ME}}{\overline{ME} + \overline{MG}}$$

$$\beta = 180.0 * \frac{\overline{MF}}{\overline{MF} + \overline{MH}}$$

4.3.4. Auswertung

Trotz der Ungenauigkeiten bei den geometrischen Berechnungen zeigt die praktische Validierung, dass sich der theoretische Ansatz mit der Praxis annähernd deckt und das Konzept grundsätzlich funktioniert.

Bei Betrachtung des Echtzeitverhaltens der Zeigerpositionierung fällt die geringe Erneuerungsrate von etwa 60 ms pro Datensatz auf. Die Anwendbarkeit ist gegeben, schnelle Bewegungen führen jedoch zum leichten „Springen des Zeigers“. Gesteigert werden kann die Performanz durch die Verwendung effizienterer Bilderkennungsalgorithmen, besonders bei der Verarbeitung und Extraktion.

Eine Erhöhung der Zuverlässigkeit und Genauigkeit kann durch den Einsatz stärkerer IR-

LEDs (auch mit größerem Abstrahlwinkel) und einem darauf abgestimmten Infrarotsperfilter erfolgen.

Schnelle Bewegungen der Kamera führen aufgrund der relativ langen Belichtungszeit der verwendeten Kamera zu Bewegungsunschärfe (Motion Blur, Verwackeln), wie in Abbildung 4.18 zu erkennen. Das Resultat ist eine fehlerhaftete Berechnung der LED-Koordinaten. Da bei der einfachen Kamera keine Möglichkeit besteht, den Shutter einzustellen (Verkürzung der Belichtungszeit), bietet sich folgender Versuch an: Periodisches Triggern (Ansteuern) der LEDs mit einem Vielfachen der Framerate der Kamera könnte die Bewegungsunschärfe um die Dioden herum unterdrücken. Hier ist zu prüfen, inwiefern die Lichtquellen noch stark genug als solche interpretierbar sind, da dies die Lichtintensität reduziert.



Abbildung 4.18.: Bewegungsunschärfe im Rohbild

Die Erweiterung um Softwaremodule zum Filtern und Regeln der Zeigerinformationen steigert das ergonomische Bedienungsempfinden, da ein „Zittern“ unterdrückt werden kann.

Um die Rohdaten mathematisch aufarbeiten zu können, müssen alle vier Infrarot-Spots im Bild erscheinen. Aktuell ist dafür die Einhaltung eines relativ hohen und unpraktikablen Abstands notwendig. In Abbildung 4.19(a) ist dargestellt, wie sich das Längenverhältnis in Abhängigkeit vom Blickwinkel der Kamera errechnet: Das Rechteck \overline{ABCD} stellt die Anzeigefläche mit der Breite b dar, der Punkt P repräsentiert den Zeigepunkt, der in dieser Situation den äußersten Rand fokussiert. Es wird davon ausgegangen, dass der Benutzer und damit die Kamera zentral mit der Distanz d vor dem Bildschirm steht. Das Verhältnis errechnet sich

nach:

$$\frac{d}{b} = \frac{1}{2 * \tan \frac{\alpha}{4}}$$

Das verwendete Standard-Objektiv ermöglicht nur einen Bildwinkel von etwa 45°, es muss das 2,5-fache der Bildschirmbreite an Mindestabstand gehalten werden, um den gesamten Bereich nutzen zu können.

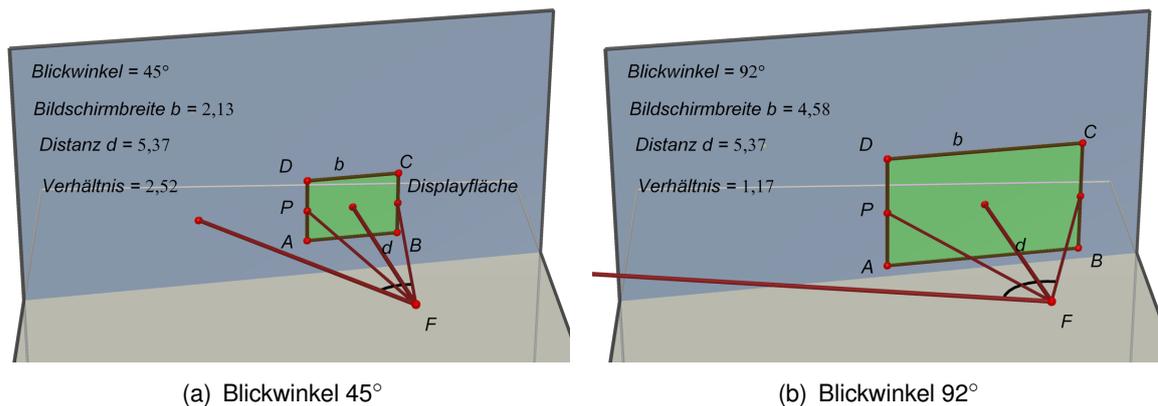


Abbildung 4.19.: Längenverhältnisse - Distanz zu Bildschirmbreite

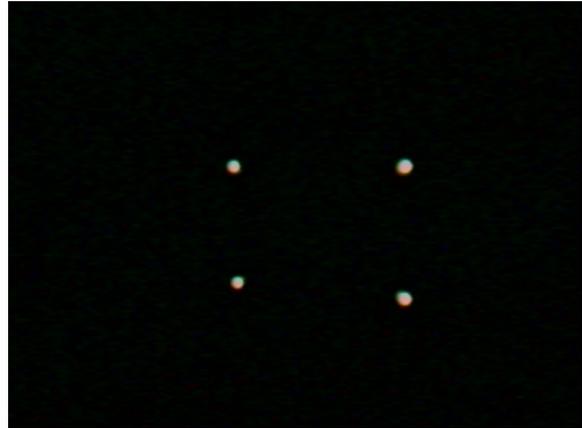
Eine Optimierung mit 3,6 mm-Objektiv erhöht den Blickwinkel auf 92°, der Mindestabstand reduziert sich dadurch auf das 1,2-fache der Bildschirmbreite und ist somit als Interaktionsdistanz absolut ausreichend (vgl. Abb. 4.19(b)). Im Versuch ist das Bild jedoch nicht verwendbar. Diese Linse besteht aus Glas und ist daher nicht für die Bündelung von Licht im mittleren und thermischen Infrarotbereich geeignet, sie ist für dieses Spektrum undurchsichtig. Verwendbar sind Linsen aus Sondermaterialien wie Kristallen (Calciumfluorid z.B.) und Halbleitermaterialien wie Silizium und Germanium.

Das geringe eintretende Licht (sichtbarer Anteil der Infrarotdioden) wurde durch den automatischen Weißabgleich der einfachen Miniaturkamera verstärkt, insgesamt war das Bild viel zu hell und es konnten keine eindeutigen Lichtquellen vom restlichen Bild unterschieden werden.

Der hohe Bedarf an Rechenleistung bei der Umsetzung als Prototyp, besonders bei der Bildverarbeitung, schlägt sich kritisch auf die Mehrbenutzerfähigkeit nieder. Um mehrere Geräte in das System zu integrieren, ist sehr viel Rechenpower erforderlich.



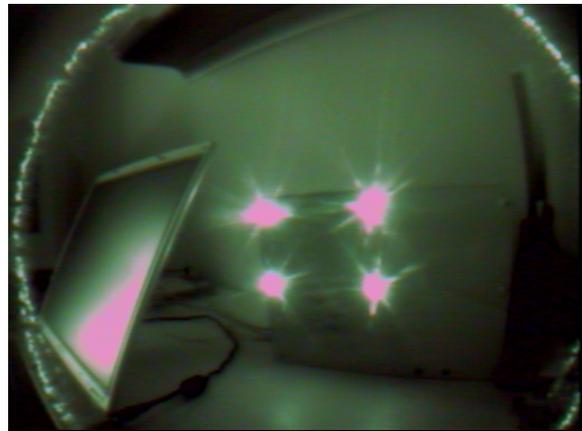
(a) Standard-Objektiv, ohne IR-Filter



(b) Standard-Objektiv, mit IR-Filter



(c) 3,6 mm-Objektiv, ohne IR-Filter



(d) 3,6 mm-Objektiv, mit IR-Filter

Abbildung 4.20.: Vergleich - Objektive mit und ohne IR-Filter

4.4. Übertragbarkeit auf Standardgeräte

Das vorherige Kapitel zeigt, dass das theoretische Konzept in der Praxis funktioniert, die technische Umsetzung für eine praktikable Anwendung jedoch stark optimiert werden muss. Bevor selbständig versucht wird, an dieser Stelle eine Lösung zu entwickeln, ist die Recherche auf dem kommerziellen Markt sinnvoll. Besonders in der Computerspieleindustrie sind viele Innovationen im Rahmen der Eingabegeräte zu verzeichnen, was in der Natur der Anwendung liegt.

4.4.1. Beschreibung

Die Firma PixArt hat in Kooperation mit Nintendo einen „Multi-Object Tracking“(MOT)-Sensor entwickelt, der den technischen Funktionsumfang der Datenerfassungsebene des Prototypen abdeckt (vgl. Abb. 4.12). Damit werden die beiden Funktionsschritte Aufnahme, Erstellung einzelner Bilder und Bildverarbeitung von einem Gerät bewältigt.

Die Kombination von einem in CMOS-Technologie gefertigten aktiven Pixelsensor mit einem digitalen Signalprozessor unter Verwendung eines Infrarotsperrefilters liefert eine Sensoreinheit, die mehrere Infrarot-Spots auf einem aufgenommenen Bild erkennt und deren Koordinaten bestimmt. Durch die Implementierung auf Hardwareebene ergibt sich eine erreichbare Mindestgeschwindigkeit der Auswertung von mindestens 20 ms pro Bild ¹².

Da es sich um ein Gerät für den Endverbraucher handelt, ist es relativ kostengünstig zu beziehen. Zudem hat sich bereits eine stetig wachsende Gemeinde gebildet (z.B. Gemeinschaft „wiili.org“¹³), die bemüht ist, technisches Know-How und damit die Möglichkeit der Verwendung für eigene Interessen zu entwickeln.

Verbaut wird diese Komponente in dem Nintendo Wiimote Controller der Nintendo Wii. Die Eingabemöglichkeiten des Controllers bezüglich der Wii-Konsole sind gegenüber herkömmlichen Systemen umfangreich, Nintendo wirbt mit intuitiver Steuerung durch (s. Abb. 4.21 und 4.22):

Physikalische Bewegung Ein integrierter Beschleunigungssensor macht die Analyse von Bewegungen möglich, ähnlich wie es beim Projekt von [Rekimoto \(1996\)](#) vollzogen wird.

Einsatz eines Zeigers Durch den oben vorgestellten MOT-Sensor ist die Umsetzung von Zeigegestiken möglich.

Möglichkeiten der Wechselwirkung über Buttons

Kabellose Kommunikation Kommunikation via Bluetooth, wodurch die Mehrbenutzerfähigkeit (bis zu vier Spieler) an der Konsole gewährleistet ist.

Feedback Für akustische, optische und haptische Rückmeldungen sind Lautsprecher, LED's und ein Vibrator installiert.

Von besonderem Interesse ist im Hinblick auf den Prototypenbau die Umsetzung der Zeigerpositionierung. Nintendo markiert mit Hilfe einer IR-Diodenleiste (Sensorbar) die Ober- oder Unterkante des Ausgabemonitors, die zwei Infrarotspots repräsentiert (siehe Abb. 4.23). Der

¹²„Information basiert auf Auswertegeschwindigkeit des Wiimote-Controllers, daher nur als Mindestangabe“

¹³wiili.org: Internet-Wiki zum Thema der Portierung einer GNU/Linuxumgebung auf die Wii-Konsole

Eingabe

- Infrarot-Modul:
PixArt Multi-Object Tracking™ engine
(MOT sensor™)
- 3-achsiger Beschleunigungssensor:
Analog Devices ADXL330
- 12 Tasten

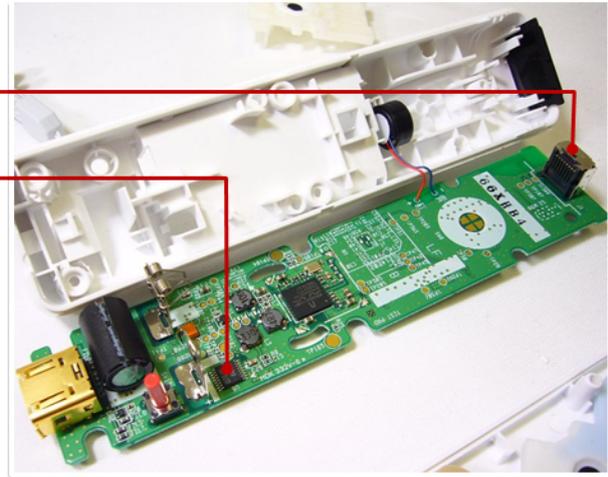


Abbildung 4.21.: Nintendo Wiimote: Kameramodul und Beschleunigungssensor

MOT-Sensor liefert die absolute Position dieser Punkte im Bild, somit kann die Konsole die relative Position zum Bildschirm errechnen.

4.4.2. Aufgabenstellung und Umsetzung

Grundlegend müssen zwei Fragestellungen beantwortet werden: Inwiefern ist die Integration in ein Wii-fremdes System möglich und falls ja, kann die Komponente im Sinne der Anforderungen als Interaktionsgerät für den interaktiven Raum verwendet werden?

Rückblickend werden die Anforderungen an eine Interaktionskomponente weitestgehend erfüllt, unklar ist die mögliche Genauigkeit des Systems. Zudem erfolgt zwar eine Zeigerpositionierung, aber die Interpretierung der Position aus der Bildebene in den Objektraum ist nicht möglich, da nach dem Wii-Konzept nur zwei Referenzpunkte zur Verfügung stehen. Beispiel: Wird der Abstand der IR-Spots geringer, ist nicht differenzierbar, ob sich der Benutzer vom Bildschirm entfernt oder seitlich um den Monitor bewegt und sich dadurch der Winkel verändert. Notwendig ist hier die Etablierung zweier weiterer Punkte (siehe Prototyp), um gerade diese Informationen gewinnen zu können.

Es gibt bereits diverse Treiber und Analysetools für Windows (GlovePIE, WiinRemote, RMX, wiimote), Linux (WMD, CWiid, Wiiewer, Perlwiimote) und Mac (DarwiinRemote, Remote Buddy), um mit den Daten der Wiimote zu arbeiten. Eine Anwendung mit dem Ziel dieser Arbeit, also der mehrbenutzerfähigen Integration in eine computergestützte Umgebung, existiert nicht, daher kann auf diese Tools nicht unmittelbar zurückgegriffen werden.

Ausgabe

- Vibrator
- Lautsprecher

Kommunikation

- Bluetooth HID:
Broadcom BCM2042

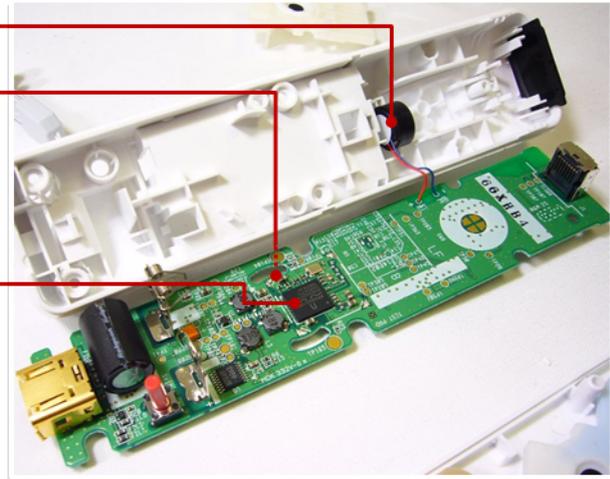


Abbildung 4.22.: Nintendo Wiimote: Rückmeldungen



Abbildung 4.23.: Zerlegte Nintendo SensorBar, Bild von „informit.com“

Zusammenfassend ergeben sich folgende Portierungsaufgaben, die in einer Testimplementierung erfolgreich bewältigt wurden:

Integration der Wiimote in ein Wii-fremdes System Da die Kommunikation via Bluetooth stattfindet und nicht verschlüsselt wird, ist die Einbindung in ein anderes System ohne weiteres möglich. Zum Beispiel unter Windows: Die Verbindung zur Fernbedienung wird mit einem handelsüblichen Bluetooth-Adapter hergestellt. Das Gerät offeriert einen Service für die Verwendung als HID-Eingabegerät. Nach Etablierung der Verbindung wird eine socket-Verbindung hergestellt, die das Senden von Kommandos und das Lesen der zurückgegebenen Datenströme ermöglicht. Das System unterstützt verschiedene Auswertungskonfigurationen, basierend auf den integrierten Sensoren und Schnittstellen können folgende Daten abgefragt werden: der Status der Tasten, die Werte des dreidimensionalen Beschleunigungssensors, die Daten des Infrarotsensors und empfangene Informationen der Erweiterungsschnittstelle, über die das Nintendo Nunchuck angeschlossen werden kann (ein Gerät für die zweite Hand, in dem Tasten

und ein Beschleunigungssensor verbaut sind – wird hier nicht genutzt). Über ein Konfigurationsregister, das via Datenstrom geschrieben wird, kann das zu empfangende Datenpaket eingestellt werden:

Im Rahmen dieses Prototypbaus sind die Daten der Tasten, der Infraroteinheit und sekundär der Beschleunigungssensor interessant. Abbildung 4.24 zeigt den Aufbau des Datenstroms, der über die socket-Verbindung von der Wiimote ausgelesen wird.

Byte	Funktion		Bit							
			7	6	5	4	3	2	1	0
0	Kanal		Daten-Auwertungs-Modus (7:0)							
1	Tasten					+	Up	Down	Right	Left
2			Home			-	A	B	1	2
3	Bewegung		x(7:0)							
4			y(7:0)							
5			z(7:0)							
6	Infrarot	Punkt 1	x(7:0)							
7			y(7:0)							
8			y (9:8)	x (9:8)	Size (3:0)					
9		Punkt 2	x(7:0)							
10			y(7:0)							
11			y (9:8)	x (9:8)	Size (3:0)					
12		Punkt 3	x(7:0)							
13			y(7:0)							
14			y (9:8)	x (9:8)	Size (3:0)					
15		Punkt 4	x(7:0)							
16			y(7:0)							
17			y (9:8)	x (9:8)	Size (3:0)					

Abbildung 4.24.: Datensatz der Wiimote

Mehrbenutzerfähigkeit Die Art der Kommunikation bestimmt gleichzeitig die Mehrbenutzerfähigkeit. Durch den Einsatz von Bluetooth als Kommunikationsmittel wird die Integration von sieben Geräten mit einem Empfänger auf einem System ermöglicht.

Erweiterung des Zeigesystems auf vier Referenzpunkte Mitglieder der Wiki-Webseite

„wiili.org“¹⁴ haben herausgefunden, wie Kommandos an die Wiimote gesendet werden können und wie die jeweils zurückkommenden Streams zu interpretieren sind. Das Ergebnis ist bereits im vorherigen Absatz beschrieben. Mit dieser Einstellung werden vier IR-Punkte übertragen, was die gewünschte Auswertung ermöglicht.

4.4.3. Auswertung

Die Portierung der Wiimote ist erfolgreich gelungen, im Rückblick auf den Bau des Prototyps hat der Autor eine Hardware mit wesentlich größerem Potential erhalten. Technisch wird das System besonders im Hinblick auf die Geschwindigkeit und Mehrbenutzerfähigkeit allen Anforderungen gerecht. Zusätzlich bietet das Gerät Möglichkeiten, nichtfunktionale Anforderungen, die nicht vom Prototyp abgedeckt werden sollten, zu erfüllen.

Die Wiimote kann mit geringem Aufwand in das Gesamtsystem in die erste (Datenerfassung) der drei Ebenen eingebunden werden. Die Aufbereitung und Auswertung bleiben unangetastet. Die Datenerfassungsebene wird jedoch erheblich gestrafft: die Videoaufnahme, Bildverarbeitung und Extraktion übernimmt der Controller. Abbildung 4.25 zeigt die Ebenenstruktur mit dem integrierten Eingabegerät.

4.5. Optimierung

Nachdem aus technischer Sicht die Portierung der Nintendo Wiimote in das Szenario möglich ist und die Verwendbarkeit mit einer Auswertungsgeschwindigkeit von 50 Bildern pro Sekunde vielversprechend klingt, besteht hier keine Notwendigkeit der Optimierung, so dass der Fokus auf die Interpretation der Daten verlegt werden kann. Bei der ersten Implementierung (vgl. 4.3) sind bewusst keine Anforderungen an eine hohe Genauigkeit der Berechnungen erhoben worden. Nach erfolgreicher Validierung des Konzeptes und gelungener Portierung der Hardware wird die mathematische Auswertung der Rohdaten optimiert. Die Interpretation von Daten aus Bildern und der Rückschluss auf bestimmte Objekte im Raum fällt in den Forschungsbereich Photogrammetrie (vgl. A.1 im Anhang).

4.5.1. Bestimmung der Zeigeposition

Wie bereits bei der Entwicklung des Prototyps in Kapitel 4.3.1 erläutert, ist als Zeigepunkt der Mittelpunkt des Kamerabildes zu verstehen. Die Aufgabe liegt in der Transformation des

¹⁴wiili.org: Internet-Wiki zum Thema der Portierung einer GNU/Linuxumgebung auf die Wii-Konsole

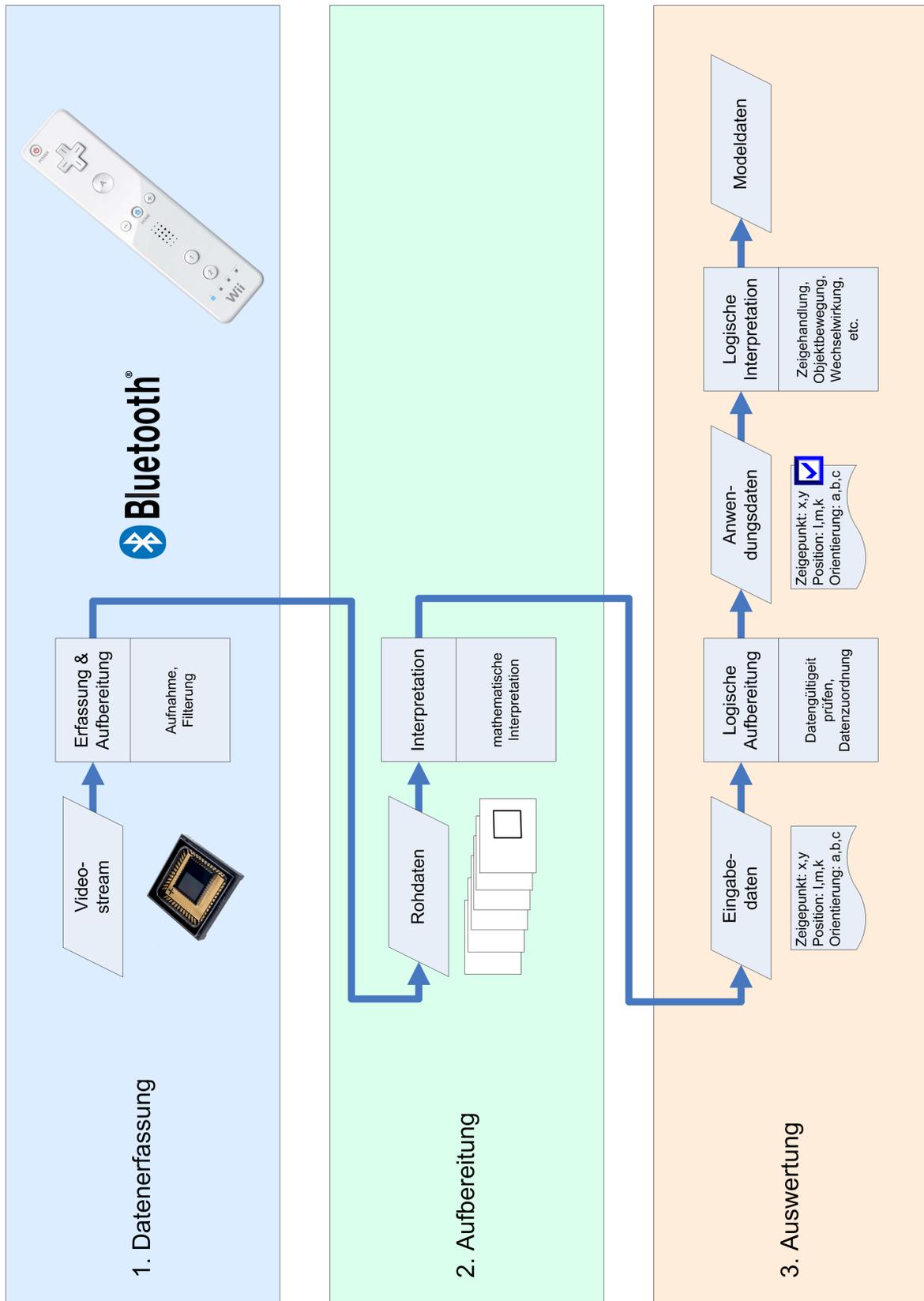


Abbildung 4.25.: Schichtmodell des Systems mit der Wiimote

Punktes bezogen auf das perspektivisch verzerrte Rechteck in das durch den Monitor aufgespannte Rechteck:

Da die Bildpunkte der vier Eckpunkte, welche im Objektraum ein Rechteck aufspannen, bekannt sind, lassen sich unter Berücksichtigung folgender Eigenschaften der Zentralprojektion weitere Punktepaare geometrisch berechnen:

1. Punkte die vor der Projektion auf einer Geraden liegen, liegen auch nach der Projektion auf einer Geraden.
2. Parallele Geraden haben denselben Fluchtpunkt.
3. Das Doppelverhältnis ist gegenüber perspektivischen Transformationen invariant.

Nach Regel 1 lässt sich ein fünftes Punktepaar hinzufügen: Der Schnittpunkt der Diagonalen des Rechtecks (Mittelpunkt M) in der Objektebene muss, da sich dieser Punkt auf jeder einzelnen Diagonale befindet, wieder in deren gemeinsamen Schnittpunkt (M') der Diagonalen im Viereck der Bildebene sein (zu sehen in Abbildung 4.26).

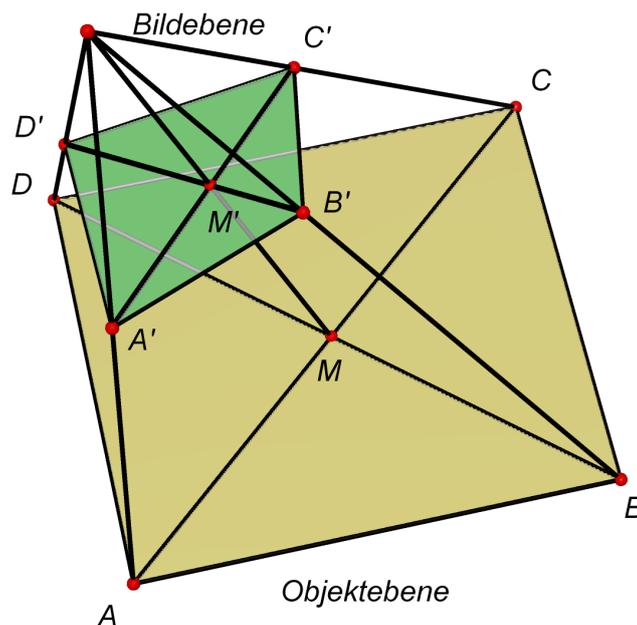


Abbildung 4.26.: Beide Ebenen

Da das Mittelpunktpaar bekannt ist, lassen sich die Punktepaare der Seitenmitten berechnen. In einem Rechteck sind die sich gegenüberliegenden Seiten (z.B. \overline{AB} , \overline{DC}) und die Gerade durch die Seitenmittelpunkte und damit durch den Mittelpunkt (\overline{HMF}) parallel zueinander. Parallele Geraden schneiden sich gemäß Regel 2 in einem Fluchtpunkt FL'_y). Die

Schnittpunkte des Strahls $\overline{FL'_yM'}$ mit den Seiten $\overline{B'C'}$, $\overline{A'D'}$ bilden folglich deren Seitenmitten (F' und H'), vgl. Abbildung 4.27. Analog hierzu erfolgt die Berechnung der Seitenmitten von $\overline{A'B'}$ und $\overline{C'D'}$.

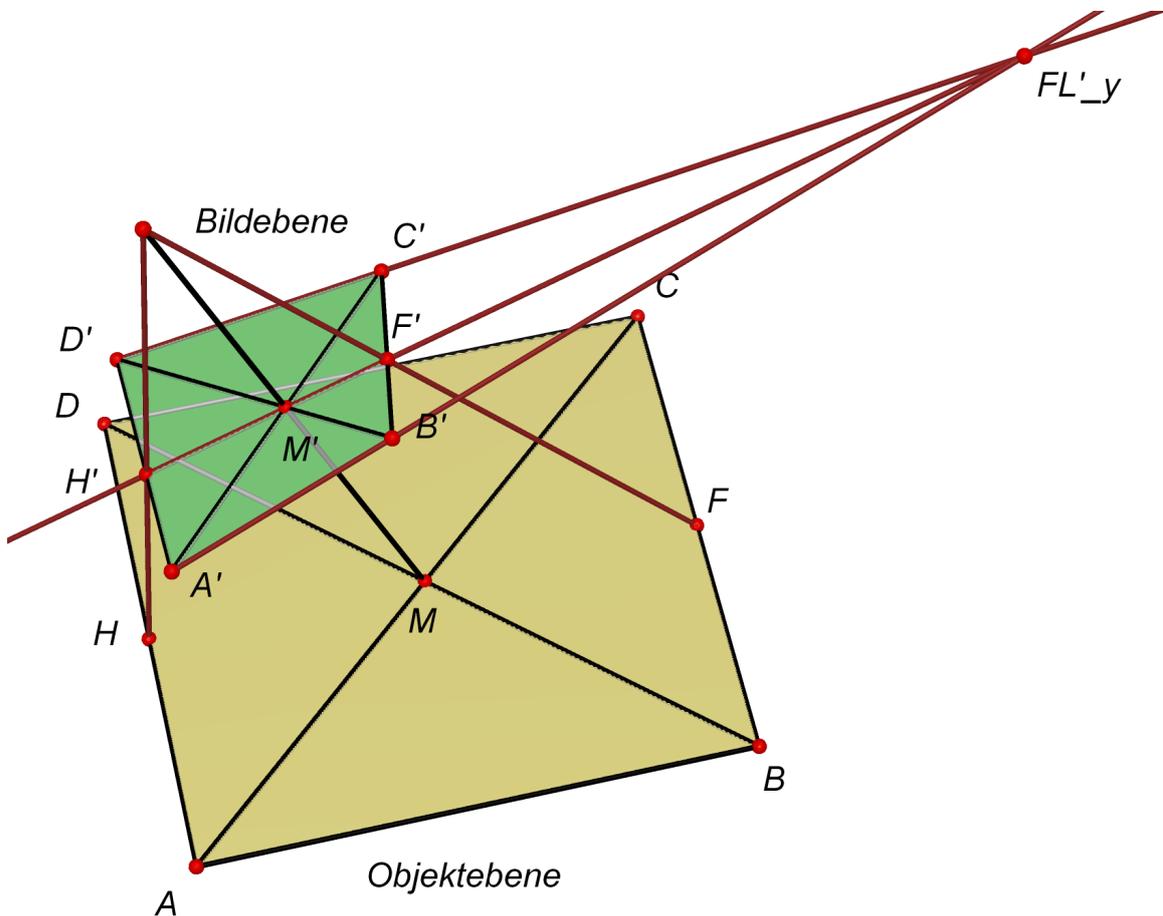


Abbildung 4.27.: Errechnung der Seitenmitten über den horizontalen Fluchtpunkt

Basierend auf diese Vorarbeit lässt sich zu jedem beliebigen Punkt in einer Ebene der entsprechende Punkt in der anderen Ebene ermitteln. Folgendes Beispiel verdeutlicht die Berechnung des Zeigepunktes:

Gegeben ist der Zeigepunkt P' in der Bildebene. Die beiden Koordinaten des Punktes werden nacheinander errechnet, die y-Position in der Objektebene:

Ein Strahl vom Fluchtpunkt FL'_y durch P' erzeugt im Schnitt mit der Seite $\overline{A'D'}$ den Punkt P'_y und entspricht in der Objektebene der Parallelen zu \overline{AB} durch P (s. Abb. 4.28). Die Länge der Strecke $\overline{AP_y}$ gibt die y-Koordinate des Punktes P wieder. Deren Berechnung erfolgt

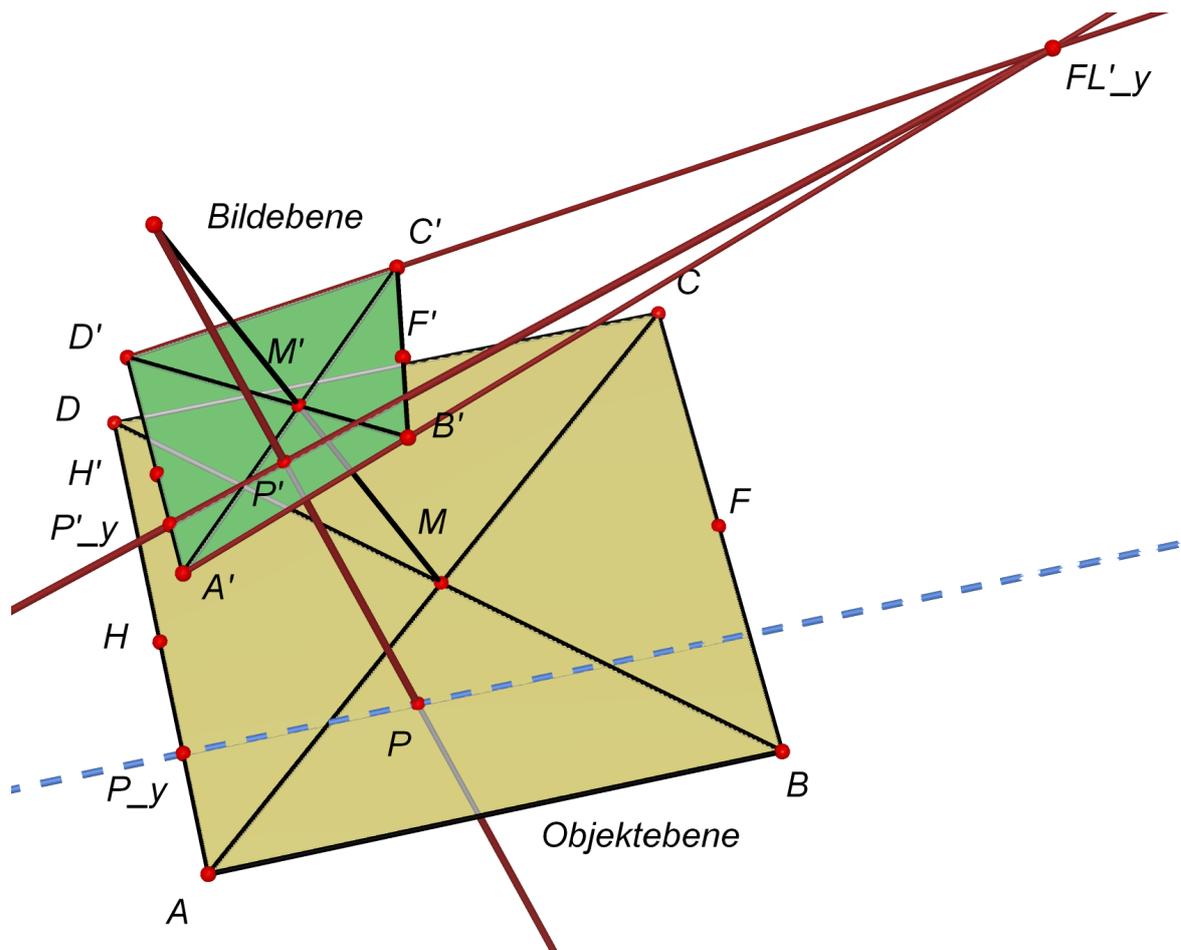


Abbildung 4.28.: Errechnung eines einzelnen Punktes über den horizontalen Fluchtpunkt

mit Hilfe des Doppelverhältnis, welches gemäß Regel 3 bei der Zentralprojektion gültig ist - veranschaulicht in Abbildung 4.29 gilt:

$$\frac{\overline{AH} * \overline{P_y D}}{\overline{AD} * \overline{P_y H}} = \frac{\overline{A'H'} * \overline{P'_y D'}}{\overline{A'D'} * \overline{P'_y H'}}$$

mit

$$\overline{PH} = \overline{AH} - \overline{AP} \text{ und } \overline{PD} = \overline{AD} - \overline{AP}$$

ergibt sich:

$$R = \frac{\overline{A'H'} * \overline{P'_y D'}}{\overline{A'D'} * \overline{P'_y H'}} * \frac{\overline{AD}}{\overline{AH}}$$

$$\overline{AP_y} = \frac{\overline{AD} - R * \overline{AH}}{1 - R}$$

Analog hierzu verhält sich die Berechnung für die x-Koordinate, es gilt:

$$R = \frac{\overline{A'E'} * \overline{P'_x B'}}{\overline{A'B'} * \overline{P'_x E'}} * \frac{\overline{AB}}{\overline{AE}}$$

$$\overline{AP_x} = \frac{\overline{AB} - R * \overline{AE}}{1 - R}$$

4.5.2. Lokalisierung

Die Interpretation der Position und Orientierung der Kamera im Raum erfordert die Auflösung der Abbildungsgleichungen aus A.1.

$$R = \begin{pmatrix} \cos \beta \cos \gamma & \cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma & \sin \alpha \sin \gamma - \cos \alpha \sin \beta \cos \gamma \\ -\cos \beta \sin \gamma & \cos \alpha \cos \gamma - \sin \alpha \sin \beta \sin \gamma & \sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma \\ \sin \beta & -\sin \alpha \cos \beta & \cos \alpha \cos \beta \end{pmatrix}$$

$$u = \Delta u_H - c * \frac{r_{11} * (x - t_x) + r_{12} * (y - t_y) + r_{13} * (z - t_z)}{r_{31} * (x - t_x) + r_{32} * (y - t_y) + r_{33} * (z - t_z)} \quad (4.1)$$

$$v = \Delta v_H + c * \frac{r_{21} * (x - t_x) + r_{22} * (y - t_y) + r_{23} * (z - t_z)}{r_{31} * (x - t_x) + r_{32} * (y - t_y) + r_{33} * (z - t_z)} \quad (4.2)$$

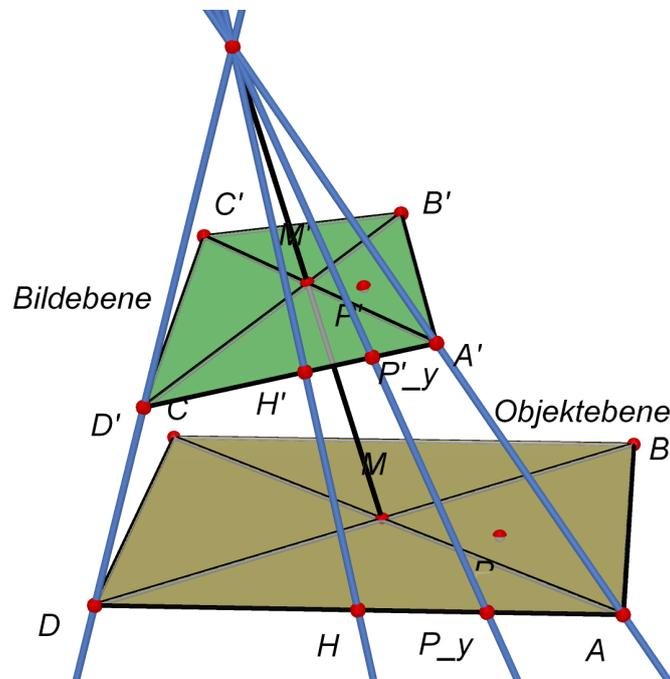


Abbildung 4.29.: Anwendung des Doppelverhältnisses

mit

- u: Bildkoordinate x
- v: Bildkoordinate y
- Δu_H : Bildhauptpunktverschiebung x
- Δv_H : Bildhauptpunktverschiebung y
- r_{11} - r_{33} : Elemente der Rotationsmatrix

Ausgehend von der bereits kalibrierten Kamera und den somit bestimmten inneren Parametern, ergeben sich zwei Gleichungen mit jeweils sechs Unbekannten: den drei Dimensionen des Translationsvektors \vec{t} und die Rotation um die drei Achsen mit den Winkeln α , β und γ . Hierzu ist die Aufstellung eines Gleichungssystems mit mindestens sechs Gleichungen erforderlich – dies ist möglich, da mindestens sechs Punktpaare bekannt sind (vier Eckpunkte, Mittelpunkt und Zeigepunkt).

Aufgrund der eingesetzten Technologie (digitale Kamera) sind die Messergebnisse der Bildpunkte fehlerbehaftet, was zur Folge hat, dass es keine eindeutige Lösung geben kann. Der Einsatz von Fehlergleichungen ermöglicht die Entwicklung einer verbesserten Lösung, bezeichnet als Ausgleichsrechnung:

$$L_i + \varepsilon_i = f_i(\xi_1, \xi_2, \xi_3, \dots, \xi_m) \text{ mit } i = 1, 2, 3, \dots, n$$

Gefragt wird nach einer Lösung, bei der die Parameter $(\xi_1, \xi_2, \xi_3, \dots, \xi_m)$ so bestimmt werden, dass die Differenzen ε_i zu den Beobachtungen L_i minimal werden. Gängig hierfür ist der Ausdruck nach Gauss:

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [f_i(\xi_1, \xi_2, \xi_3, \dots, \xi_m) - L_i]^2 \rightarrow \text{Minimum}$$

Da die unbekannt Parameter $(\xi_1, \xi_2, \xi_3, \dots, \xi_m)$ in den Abbildungsgleichungen nichtlinear miteinander verknüpft sind, ist eine direkte Auflösung nach diesen nicht möglich - es muss eine lineare Ersatzfunktion gesucht werden. Wie in [Meisel \(2005\)](#) beschrieben, wird zu einem nichtlinearen Gleichungssystem ein linearisiertes Fehlergleichungssystem erstellt:

$$L_i + \varepsilon_i = f_i(\hat{\xi}_1, \hat{\xi}_2, \hat{\xi}_3, \dots, \hat{\xi}_m) + \frac{\delta f_i(\xi)}{\delta \xi_1} \Big|_{\xi} \cdot (\bar{\xi}_1 - \hat{\xi}_1) + \frac{\delta f_i(\xi)}{\delta \xi_2} \Big|_{\xi} \cdot (\bar{\xi}_2 - \hat{\xi}_2) + \dots$$

Um die nichtlineare Ausgleichsrechnung durchführen zu können, müssen die Abbildungsgleichungen im nächsten Schritt in das Format der Fehlergleichungen konvertiert werden, dies geht jedoch über den Rahmen dieser Arbeit hinaus und bleibt somit nur in der Theorie angedacht.

4.5.3. Portierung in ein Multi-Monitor-System

Bisher wurde die Umsetzung auf die Interaktion in einer Ein-Monitor-Umgebung realisiert. Eine Skalierung auf mehrere Ausgabesysteme wurde in Hinblick auf die zur Verfügung stehende Zeit nicht umgesetzt. Zwei Fragestellungen ergeben sich jedoch vorab:

Große Anzeigefläche

Wie bereits aus der Prototyprealisierung bekannt, ist der notwendige Abstand zur Interaktionsfläche nicht praktikabel und muss optimiert werden. Die Wiimote scheint optische Komponenten integriert zu haben, wodurch der Aufnahmewinkel zur Kameralösung vergrößert ist. Bei sehr grossen Anwendungen ist selbst diese Verbesserung nicht ausreichend, daher muss grundsätzlich ein Lösungsweg gefunden werden, der eine beliebige Skalierung der Anzeigeebene zulässt.

Lösungsansatz:

- **Verwendung mehrerer LED-Bereiche:** Eine Aufteilung der Fläche in mehrere kleine Rechtecke (siehe Abb. 4.30(a)) ist nur bedingt einsetzbar. Die Wiimote kann nur vier Infrarotspots gleichzeitig liefern. Es muss dafür gesorgt werden, dass entweder der Abstand der LED's voneinander gross genug ist oder dass die LED's an- und ausgeschaltet werden können. In beiden Fällen muss der aktuelle Interaktionsbereich identifiziert werden können, möglich wäre dies durch Ein- und Ausblenden der LED's.

Problematisch ist diese Lösung, sobald mehrere Geräte verwendet werden. Eine Koordination der Ein- und Ausschaltzyklen ist notwendig, evtl. gibt es an dieser Stelle zeitliche Beeinträchtigungen bei der Interaktion.

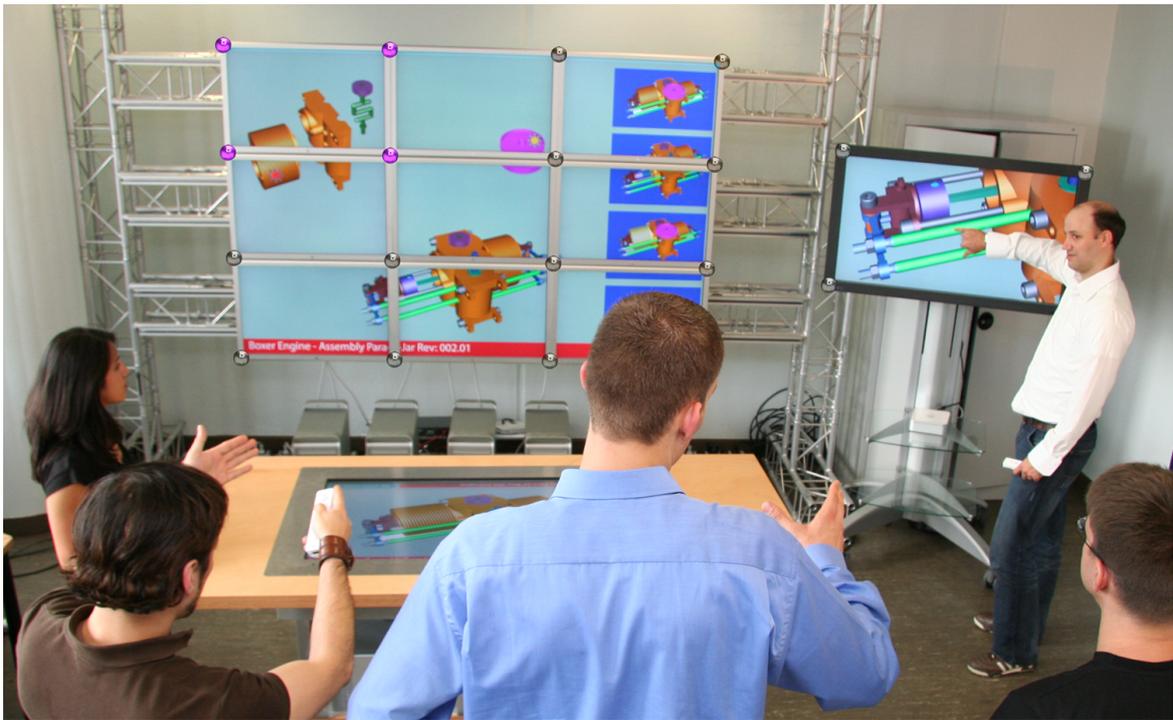
- **Verwendung eines einzelnen LED-Referenzbereiches:** Die geometrische Position ist über die Ränder des Rechtecks hinweg berechenbar, quasi unendlich skalierbar. Zu jedem Punkt in der Bildebene kann der korrespondierende Punkt in der Objektebene berechnet werden, Basis hierfür sind die vier Referenzpunkte, siehe Abbildung 4.30(b).

Voneinander entfernte Anzeigeflächen

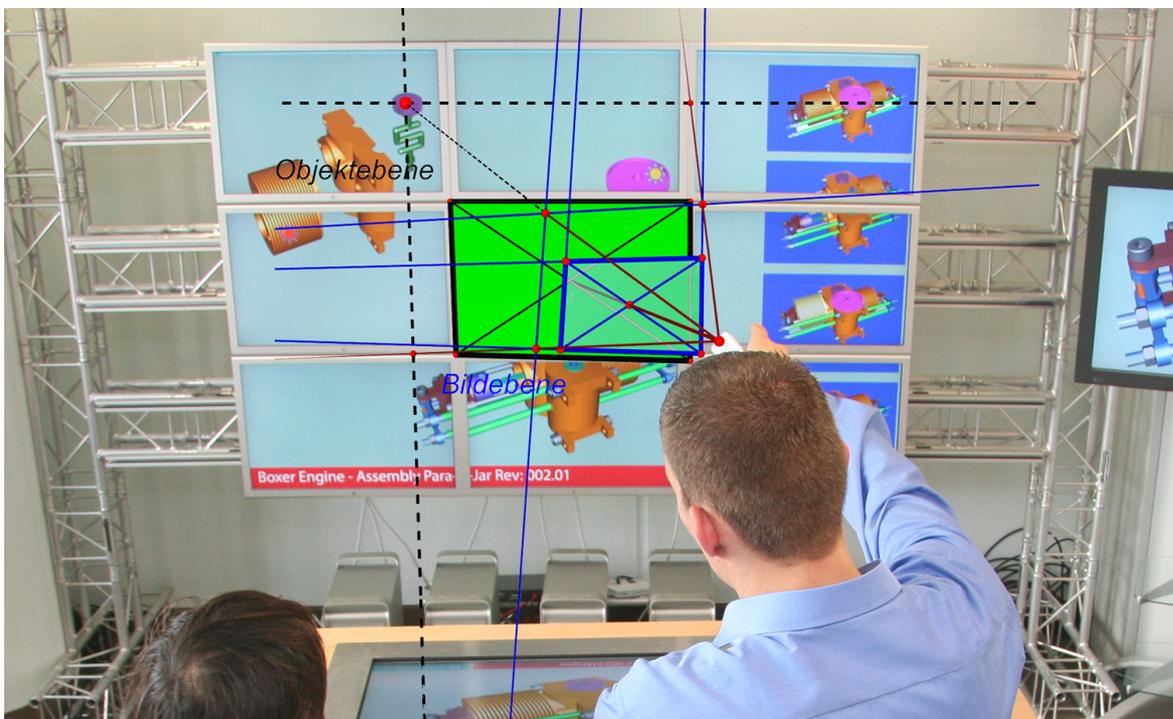
Die Vorgehensweise entspricht jener, die für mehrere LED's verwendet wird: Ausgangspunkt ist, dass jede Anzeigefläche für sich mit LED's zur Markierung versehen ist. Hier stellt sich hauptsächlich die Frage der Unterscheidung.

4.6. Auswertung

Es ist gelungen einen Prototypen umzusetzen, der den Anforderungen in Hinblick auf die Zeigegestik in allen Punkten gerecht wird. Als Bonus kommt hinzu, dass diese Hardwarebasis ein gutes Potential an Erweiterbarkeit hinsichtlich Multimodalität mit sich bringt. Die in Kapitel 4.4 beschriebene Ausstattung in Form von Peripherie ermöglicht die Analyse von



(a) Installation mit grossflächiger LED-Bestückung



(b) Pointing ausserhalb des Rechtecks mit einem Referenzviereck

Abbildung 4.30.: Ansätze der Multi-Monitor Umsetzung

Bewegungen und bietet die Möglichkeit, dem Nutzer akustisches und haptisches (Vibration) Feedback wiederzugeben.

Die Ergänzung des Eingabemodus um Sprache wurde im Analysekapitel [3.3](#) festgelegt, jedoch nicht umgesetzt, denn die Forschung der Spracherkennung durch Computer ist so weit fortgeschritten, dass eine Anwendung in diesem Rahmen (nur sehr geringes Vokabular notwendig) als praktikabel angesehen werden kann.

5. Schluss

5.1. Zusammenfassung

Das Projekt „UbiComp“ der HAW Hamburg hat als Ziel die Steigerung der Effizienz und Ergonomie der Zusammenarbeit einer Personengruppe durch die Integration von technischen Systemen in die Arbeitsumgebung.

Dieses sollte so geschehen, dass der Benutzer es nicht aktiv wahr nimmt, der Leitsatz ist hier die intuitive Handhabung des Systems: was nicht erlernt werden muss, wird einfach benutzt. Diese Arbeit beschäftigte sich mit der Komponente Interaktion in einer solchen Umgebung.

Zum Verständnis wurden im Grundlagenkapitel 2 die von diesem Thema betroffenen Forschungsgebiete erläutert: Mensch-Maschine-Interaktion, multimodale Interaktion und Computergestützte Gruppenarbeit.

Im Analysekapitel 3 wurde das Forschungsprojekt RESCUE erläutert und die Komponente Leitstand als ein zweckmässiges Szenario (Kapitel 3.1) für eine kollaborative Arbeitsumgebung bestimmt. Aus den dort definierten Anwendungsfällen konnten konkrete Anforderungen an ein geeignetes Interaktionskonzept erhoben werden (Kapitel 3.2). Vordergründig wurde die Hand als das Interaktionsmedium festgelegt, unterstützende sollte die Sprache eingesetzt werden (Kapitel 3.3). Die Recherche nach aktuellen Projekten in Hinblick auf Handbasierende Interaktion lieferte verschiedene Lösungsansätze (Kapitel 3.4), wobei die Verwendung einer Kamera als direktes Eingabegerät das grösste Potential bot.

Im Rahmen dieser Vorarbeit wurde die Idee entwickelt eine Kamera als Eingabegerät zu verwenden und die Interaktionsfläche mit Infrarotlicht zu markieren. In einer Machbarkeitstudie (Kapitel 4.3) konnte mit einem handelsüblichen, günstigen Kamerasystem gezeigt werden, dass der theoretische Ansatz umsetzbar ist und vielversprechend funktioniert. Die Anforderungen an den Rechenaufwand der Datenverarbeitung waren jedoch zu hoch und so nur schwierig mehrfach einsetzbar (Mehrbenutzerfähigkeit) in einer Arbeitsumgebung.

Die Suche nach geeigneter Hardware konnte erfolgreich abgeschlossen werden, als der Nintendo Wii Controller Wiimote auf den Markt kam (Kapitel 4.4). Die Zeigegestik funktionierte nach einem annähernd gleichen Prinzip, die notwendigen Daten wurden jedoch direkt auf

dem Controller ausgewertet, was enorme Leistungseinsparungen auf Seite des Rechners mit sich brachte.

Es wurde eine Anwendung programmiert, mit der die Wiimote in das Entwicklersystem eingebunden werden konnte und somit für die Interaktion in der Arbeitsumgebung zur Verfügung stand. Die Algorithmen aus der Machbarkeitsstudie waren direkt übertragbar und das Gerät konnte als Zeigestock eingesetzt werden.

Optimierungen der mathematischen Algorithmen (Kapitel 4.5.1) erlauben eine wesentlich genauere Positionsbestimmung der Zeigegestik, was sich bei der Skalierung der Anzeigefläche besonders stark auswirkt (Kapitel 4.5.3) bzw. diese erst möglich macht.

Es ist gelungen einen Prototypen umzusetzen, der auf dem oben erwähnten Leitgedanken, der Intuitivität, basiert und somit ein geeignetes Interaktionsgerät für eine Umgebung im Sinn des ubiquitären Computing ist. Das Gerät kann wie ein Stock betrachtet werden, der in die Hand genommen wird, um anderen etwas zu zeigen. Sollen Objekte, auf die man zeigt, bewegt werden, wird „zugedrückt“, also eine im „Stock“ integrierte Taste gewählt und die Wechselwirkung mit dem Objekt findet statt.

Ursprünglich sollte diese Funktionalität durch Sprache unterstützt werden, hierzu in Kapitel 5.2.2 mehr.

Wie im Entwurf (Kapitel 4.2) beschrieben, wurde die Datenauswertungsebene nur oberflächlich umgesetzt, um die Ergebnisse des Eingabegerätes zu visualisieren. Da diese Ergebnisse sehr vielversprechend sind, bietet diese Arbeit eine gute Ausgangslage für die Umsetzung eines vollständigen multimodalen Interaktionssystems auf Anwendungsebene (Kapitel 5.2.3). Einen Ansatz der Synergie von Modalitäten ist in Kapitel 5.2.2 beschrieben.

5.2. Ausblick

5.2.1. Weitere Einsatzmöglichkeiten des Zeigestocks

Exemplarisch für ein CSCW-System wurde in dieser Arbeit ein Leitstand im Rahmen von Disaster-Szenarien verwendet. Einsetzbar ist das Zeigegerät in denkbar vielen Umgebungen, welche die Kollaboration mehrerer Personen unterstützen sollen. Beispielhaft sind neben weiteren Arbeitsumgebungen auch Bereiche des Home-Entertainment, von der Steuerung von Multimediageräten bis hin zu Home-Control-Lösungen.

Entfernt man sich von der CSCW-Metapher, ohne dabei das UbiComp-Thema ausser Acht zu lassen, sind weitere Einsatzmöglichkeiten denkbar: Ein vollständig anderer Ansatz ist

die Verwendung des Zeigegerätes in industrieller Umgebung. Hier wird die Interaktionskomponente als solches mit der Charakteristik der Zeigefunktion fokussiert. Vorstellbar ist ein Hochregallager, in dem automatisierte Flurförderzeuge eingesetzt werden (s. Abb. 5.1). Aktuell werden Lagerplätze über ihre Koordinaten identifiziert, der automatisierte Stapler fährt die Aufträge ab. Das System gewinnt an Interaktivität, wenn es dem Lagerist oder Logistiker möglich wird, Lagerplätze durch direkte Zeigegestik auszuwählen.



Abbildung 5.1.: Automatisierter Stapler im Hochregallager (Bild von Still GmbH)

Theoretisch könnte das durch die Ausstattung des Lagerbereiches mit Infrarotdioden ermöglicht werden, wobei dieses Konzept sicherlich vom Standpunkt der Steuerung und Interpretation der Daten (aufgrund der Grösse der Kulisse) sehr komplex ist. Es zeigt aber das Potential eines solchen Eingabegerätes.

5.2.2. Verbesserung der Zeigehandlung

Diese Arbeit hat gezeigt, dass eine Zeigehandlung auf Kamerabasis realisierbar ist. Um schlussendlich ein multimodales Eingabesystem zu erhalten, muss untersucht werden, wie sich diese Zeigegestik durch weitere Modalitäten unterstützen lässt. Konzeptionell wurde diese Fragestellung in dieser Arbeit behandelt, konkret im Grundlagenkapitel 2.2 und mit der Betrachtung vorhandener Interaktionslösungen im Analysekapitel 3.4.

Die konkrete Umsetzung steht noch aus, an dieser Stelle sollte für Spracherkennung ein geeigneter Wortschatz entwickelt werden, Bibliotheken sind bereits frei verfügbar und funktionieren mit definiertem Wortschatz einwandfrei.

Auch die Interpretation von Bewegung kann als unterstützendes Element der Zeigegestik verwendet werden. Richtet die Person das Zeigergerät auf einen speziellen Punkt, ist das gezielte Abstoppen der Hand- bzw. Armbewegung über den Beschleunigungssensor der Wiimote zu erkennen. Dies kann als eindeutiger Abschluss einer Zeigehandlung analysiert werden, für eine Wechselwirkung mit einem fokussierten Objekt ist es also nicht nötig, einen Knopf oder ein anderes Kommando z.B. in Form von Sprache zu erteilen.

Darüber hinaus gibt es weitere Gestiken, wie z.B. der Blickrichtung, die durch Eye-Tracking ermittelt werden kann. Erst das gleichzeitige Fokussieren eines Objektes mit der Wiimote und den Augen ruft eine Selektion hervor.

Die Verwendung mehrerer Modalitäten schafft die Möglichkeit, die einzelnen Interpretationen gegenseitig auf Plausibilität zu prüfen um so das bestmögliche Wissen über Eingabeereignisse zu gewinnen.

Besonders hilfreich ist dieses Vorgehen bei fehlerhaften oder fehlenden Eingabedaten, wie z.B. wenn bei der Spracherkennung laute Hintergrundgeräusche vorhanden sind und Daten dadurch nicht mehr verwertbar sind, oder wenn beim Einsatz der Wiimote die Infrarotdioden verdeckt sind und dadurch keine Zeigerposition bestimmbar ist. An dieser Stelle könnten die Bewegungsdaten innerhalb eines beschränkten zeitlichen Rahmens als Ersatz fungieren.

Eine typischer Anwendungsfall für eine notwendige Informationsfusion um die Genauigkeit der Daten zu sichern, ist in Abbildung 5.2 dargestellt. Die Aufgabe ist das Markieren mehrerer Objekte durch eine kreisförmige Bewegung auf der Anzeigefläche und der gleichzeitigen Verdeutlichung durch Sprache. Die Bewegung ist dem natürlichen Ablauf nachempfunden (zusätzlich analysierte Modalitäten: Sprache und Bewegung):

1. In dem Moment, als der Benutzer sich entscheidet, einen Bereich zu markieren, befindet sich der Cursor auf einem beliebigen Punkt, dem Startpunkt.
2. Der erste markante Punkt steht für die Koordinaten an dem der Benutzer beginnt den Bereich zu markieren. Zusätzlich beginnt er mit der Aussprache des Wortes „Das!“.

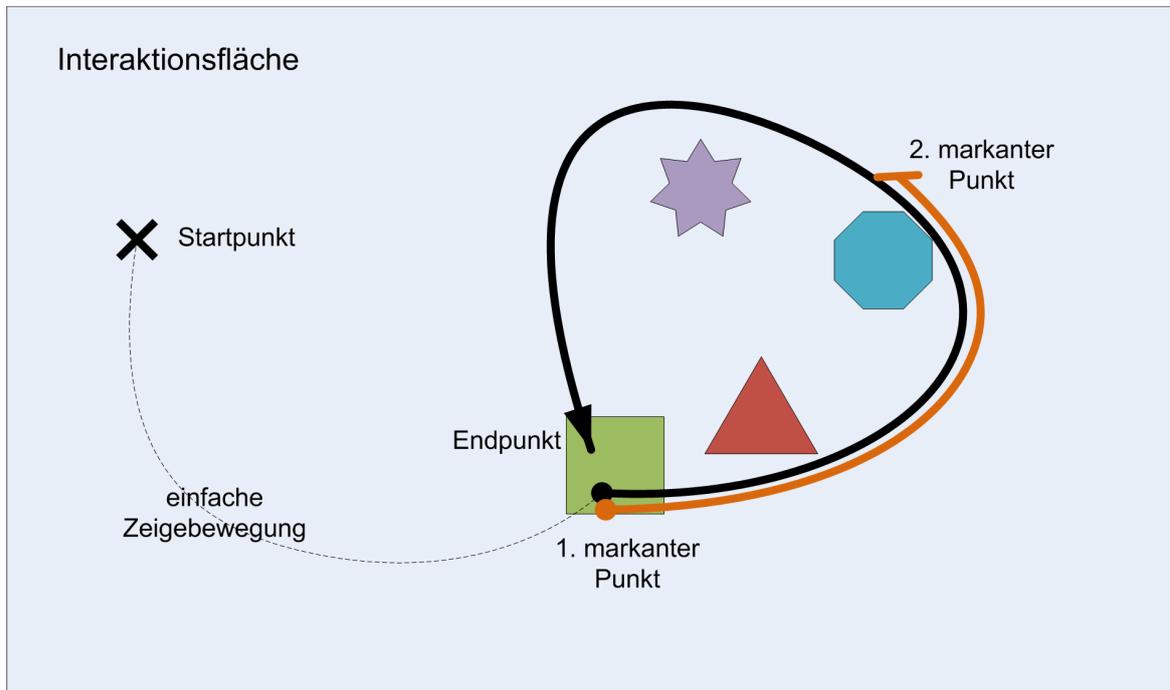


Abbildung 5.2.: Informationsfusion am Beispiel einer Bereichsselektion

3. Der zweite markante Punkt spiegelt zeitlich den Moment wieder, als das Wort „Das!“ ausgesprochen wurde.
4. Der Endpunkt schliesst den zu selektierenden Bereich.

Alle vier Symbole sollen markiert werden, das Beispiel zeigt, aus welchem Grund die Verschmelzung der Daten sehr hilfreich sein kann. Die Zeigengestik dient lediglich der absoluten Referenzierung der Objekte. Eine Bewegung könnte aus den Zeigendaten interpretiert werden, dafür wird jedoch der Beschleunigungssensor eingesetzt:

Bewegung Am Startpunkt beginnend wird die Bewegung als frei empfunden. Mit einer charakteristischen Verzögerung erreicht der Benutzer den ersten Zwischenpunkt, diese Gestik wird bereits als Selektion interpretiert und es findet die Auswahl des grünen Quadrates statt. Die identische Situation besteht ab dem 1. markanten Punkt, erst die Verzögerung am Endpunkt würde als Selektion gewertet werden.

Sprache Mit der alleinigen Aussprache von „Das!“ kann nicht konkret entschieden werden, welches Objekt gemeint ist, da sich der Zeigepunkt währenddessen verschoben hat. Denkbar wäre die Auswahl der Objekt unter dem Zeigegerät: am Anfang (Viereck), am Ende (Achteck) oder über den vollständig zeitlichen Bereich (Viereck, Dreieck, Achteck) des Sprechens.

Fusion Erst die Synergie der beiden Modalitäten schafft ein eindeutige Auswahl: Bis zum ersten markanten Punkt wurde sprachlich nichts geäußert. Der Ausruf „Das!“ bestimmt den Beginn der Selektion, da sich der Cursor bewegt, während ausgesprochen wird, wird davon ausgegangen, dass es sich um keine punktuelle Selektion handelt. Nach Bendeigung des Wortes ist der Zeiger noch immer in Bewegung, mit der starken Verzögerung zum Endpunkt kann der gemeinte Bereich eindeutig identifiziert werden: alles vier Objekte.

Dies ist nur ein Beispiel für die Möglichkeiten der logischen Verknüpfungen der Informationen.

5.2.3. Weitere Ideen

Anwendungsebene Bisher wurde das Eingabegerät und dadurch die Funktionalität Zeigegestik in den Vordergrund gestellt. Interessant werden jetzt die Applikationen, in denen es unterstützend, als Teil einer komplexeren Interaktion, eingebettet wird.

Die Perspektive, aus der das System betrachtet wird, verlagert sich, es geht weniger um die Interaktion als um den Anwendungshintergrund. Eine exemplarische Anwendung ist die Arbeit von [Roßberger \(2007\)](#), sie geht von der Betrachtungsweise aus genau in diese Richtung.

Interaktionsmodi Im vorherigen Kapitel [5.2.2](#) wurde kurz eine Selektionsgeste angesprochen. In diversen Anwendungsfällen sind bereits Gesten implementiert, die Maussteuerung des Webbrowsers Opera¹ zum Beispiel oder die Steuerung von kleinen tragbarer Computer, wie einem PDA.

Im Rahmen der entfernten dreidimensionalen Interaktionsmöglichkeit durch das, in dieser Arbeit entwickelte Gerät, sind bisherige Konzepte zu überdenken. Sind diese Gesten portierbar oder muss das bestehende Gesten-Repertoire ganz neu überdacht werden?

¹Opera ist ein Webbrowser mit E-Mail-Client und weiteren Tools der Opera Software ASA

Literaturverzeichnis

- [Ackoff 1989] ACKOFF, R.L.: From Data to Wisdom. In: *Journal of Applied Systems Analysis* 16 (1989), S. 3–9
- [Ballagas u. a. 2006] BALLAGAS, Rafael ; BORCHERS, Jan ; ROHS, Michael ; SHERIDAN, Jennifer G.: The Smart Phone: A Ubiquitous Input Device. In: *IEEE Pervasive Computing* 5 (2006), Nr. 1, S. 70. – ISSN 1536-1268
- [Ballagas u. a. 2005] BALLAGAS, Rafael ; ROHS, Michael ; SHERIDAN, Jennifer G.: Sweep and point and shoot: phonecam-based interactions for large public displays. In: *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*. New York, NY, USA : ACM Press, 2005, S. 1200–1203. – ISBN 1-59593-002-7
- [Böhm 2007] BÖHM, Dr.-Ing. J.: *Kamera-Navigation in Punkt-Wolken*. January 2007. – URL <http://www.fh-oow.de/3dtage/folien/boehm.pdf>
- [Bi u. a. 2005] BI, Xiaojun ; SHI, Yuanchun ; CHEN, Xiaojie ; XIANG, Peifeng: Facilitating interaction with large displays in smart spaces. In: *sOc-EUSAI '05: Proceedings of the 2005 joint conference on Smart objects and ambient intelligence*. New York, NY, USA : ACM Press, 2005, S. 105–110. – ISBN 1-59593-304-2
- [Burger 1997] BURGER, Cora: *Groupware : Kooperationsunterstützung für verteilte Anwendungen*. Heidelberg : dpunkt, 1997. – ISBN 3-920993-60-8
- [Buxton und Moran 1990] BUXTON, Bill ; MORAN, Tom: EuroPARC's integrated interactive intermedia facility (IIIF): early experiences. In: *Proceedings of the IFIP WG 8.4 conference on Multi-user interfaces and applications*. Amsterdam, The Netherlands, The Netherlands : Elsevier North-Holland, Inc., 1990, S. 11–34. – URL <http://www.billbuxton.com/iiif.html>. – ISBN 0-444-88750-1
- [Cao und Balakrishnan 2003] CAO, Xiang ; BALAKRISHNAN, Ravin: VisionWand: interaction techniques for large displays using a passive wand tracked in 3D. In: *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 2003, S. 173–182. – ISBN 1-58113-636-6
- [van Dam 1997] DAM, Andries van: Post-WIMP user interfaces. In: *Commun. ACM* 40 (1997), Nr. 2, S. 63–67. – ISSN 0001-0782

- [Dan R. Olsen und Nielsen 2001] DAN R. OLSEN, Jr. ; NIELSEN, Travis: Laser pointer interaction. In: *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA : ACM Press, 2001, S. 17–22. – ISBN 1-58113-327-8
- [Davids 2007] DAVIDS, Arno: Ad-hoc Sensornetzwerk zur Gebäudeüberwachung und Navigationsunterstützung / HAW Hamburg. Februar 2007. – Seminarbericht
- [Ellis u. a. 1991] ELLIS, Clarence A. ; GIBBS, Simon J. ; REIN, Gail: Groupware: some issues and experiences. In: *Commun. ACM* 34 (1991), January, Nr. 1, S. 39–58. – URL <http://portal.acm.org/citation.cfm?id=99987>. – ISSN 0001-0782
- [Fischer 2006] FISCHER, Christian: Multimodale Interaktionen in Collaborative Workspaces / HAW Hamburg. URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2006/fischer/abstract.pdf>, Juli 2006. – Anwendung I
- [Fischer 2007] FISCHER, Christian: Interaktionskomponenten für Collaborative Workspaces / HAW Hamburg. URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master06-07-aw/fischer/report.pdf>, 02 2007. – Anwendung 2
- [Forlines u. a. 2005] FORLINES, Clifton ; BALAKRISHNAN, Ravin ; BEARDSLEY, Paul ; BAAR, Jeroen van ; RASKAR, Ramesh: Zoom-and-pick: facilitating visual zooming and precision pointing with interactive handheld projectors. In: *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 2005, S. 73–82. – ISBN 1-59593-271-2
- [Hachet u. a. 2005] HACHET, Martin ; POUDEIROUX, Joachim ; GUITTON, Pascal: A camera-based interface for interaction with mobile handheld computers. In: *SI3D '05: Proceedings of the 2005 symposium on Interactive 3D graphics and games*. New York, NY, USA : ACM Press, 2005, S. 65–72. – ISBN 1-59593-013-2
- [Helman 1998] HELMAN, Dean: Model-View-Controller. In: *Objective Toolkit Pro* (1998), May. – URL <http://ootips.org/mvc-pattern.html>
- [Hinck 2007] HINCK, Steffen: Einsatz von Wearable Computing in Desasterszenarien / HAW Hamburg. Februar 2007. – Seminarbericht
- [Ishii u. a. 1994] ISHII, Hiroshi ; KOBAYASHI, Minoru ; ARITA, Kazuho: Iterative design of seamless collaboration media. New York, NY, USA : ACM Press, 1994, S. 83–97. – ISSN 0001-0782
- [Johansen 1988] JOHANSEN, Robert: *GroupWare: Computer Support for Business Teams*. New York, NY, USA : The Free Press, 1988. – ISBN 0029164915

- [Johanson 2003] JOHANSON, Bradley E.: *Application coordination infrastructure for ubiquitous computing rooms*. Stanford, CA, USA, Dissertation, 2003. – Adviser-Armando Fox
- [Johnson-Lenz 1981] JOHNSON-LENZ, P. & T.: Consider the Groupware: Design and Group Process Impacts on Communication in the Electronic Medium. / Hiltz & Kerr, Studies of Computer-Mediated Communications Systems: A Synthesis of the Findings. Computerized Conferencing and Communications Center, New Jersey Institute of Technology, 1981 (16). – Research Report
- [Kabalkin 2007] KABALKIN, Mykhaylo: Verteiltes Persistenz-System / HAW Hamburg. Februar 2007. – Seminarbericht
- [Köckritz 2007] KÖCKRITZ, Oliver: Virtueller Collaborative Workspace mit Remote Windows / HAW Hamburg. Februar 2007. – Seminararbeit
- [Kienle 2005] KIENLE, Dr. A.: *Vorlesung: Informatik & Gesellschaft*. 2005
- [Lamont u. a. 2004] LAMONT, Judith ; GOTTA, Mike ; MCCORT, Kevin ; GARCIA, Jose ; FORQUER, Bill: *Collaboration: CARE Canada's use of collaboration and the latest trends in that industry*. 04 2004. – URL <http://www.kmworld.com/Articles/ReadArticle.aspx?ArticleID=9521>
- [Licklider 1960] LICKLIDER, J. C. R.: Man-Computer Symbiosis. In: *IRE Transactions on Human Factors in Electronics* HFE-1 (1960), March, S. 4–11. – URL <http://groups.csail.mit.edu/medg/people/psz/Licklider.html>. – ISSN 1058-6180
- [Lin u. a. 2007] LIN, Edward C. ; YU, Kai ; RUTENBAR, Rob A. ; CHEN, Tsuhan: A 1000-word vocabulary, speaker-independent, continuous live-mode speech recognizer implemented in a single FPGA. In: *FPGA '07: Proceedings of the 2007 ACM/SIGDA 15th international symposium on Field programmable gate arrays*. New York, NY, USA : ACM, 2007, S. 60–68. – ISBN 978-1-59593-600-4
- [Matveyev und Göbel 2003] MATVEYEV, Sergey V. ; GÖBEL, Martin: The optical tweezers: multiple-point interaction technique. In: *VRST '03: Proceedings of the ACM symposium on Virtual reality software and technology*. New York, NY, USA : ACM Press, 2003, S. 184–187. – ISBN 1-58113-569-6
- [Meisel 2005] MEISEL, Prof. Dr. A.: *Angewandte lineare Algebra und Ausgleichsrechnung*. March 2005. – URL https://www.informatik.haw-hamburg.de/cms/uploads/media/RV_LinAlgeb.pdf
- [Meisel 2006] MEISEL, Prof. Dr.-Ing. A.: *Das 3D-Kameramodell der Technischen Informatik der HAW*. September 2006

- [Miyaoaku u. a. 2004] MIYAOKU, Kento ; HIGASHINO, Suguru ; TONOMURA, Yoshinobu: C-blink: a hue-difference-based light signal marker for large screen interaction via any mobile terminal. In: *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 2004, S. 147–156. – ISBN 1-58113-957-8
- [Möller-Freuck 2007] MÖLLER-FREUCK, Niels-Christian: Polizei Desktop / HAW Hamburg. Februar 2007. – Seminarbericht
- [Möslein 1993] MÖSLEIN, Kathrin: CSCW als Arbeitssystem-Gestaltung. Institut für Informatik, Technische Universität München, 1993
- [Napitupulu 2006] NAPITUPULU, Jan: Multihead-Display als interaktives Informationssystem / HAW Hamburg. 2006. – Seminarbericht
- [O'Hara-Devereaux und Johansen 1994] O'HARA-DEVEREAUX, Mary ; JOHANSEN, Robert: *GlobalWork: Bridging Distance, Culture and Time*. San Francisco : Jossey-Bass, 1994. – 439 S
- [Paley u. a. 1998] PALEY, W. B. ; BUXTON, Bill ; FEINER, Steve ; ISHII, Hiroshi ; MOUNTFORD, S. J. ; SHAFER, Steve ; WEISER, Mark: The sorcerer's apprentice (panel): ubiquitous computing and graphics. In: *SIGGRAPH '98: ACM SIGGRAPH 98 Conference abstracts and applications*. New York, NY, USA : ACM, 1998, S. 206–208. – ISBN 1-58113-046-5
- [Pein 2007] PEIN, Pascal: Suchen und Finden im Collaborative Workspace / HAW Hamburg. Februar 2007. – Seminarbericht
- [Piening 2007] PIENING, Andreas: RESCUE: Leitstand für Disaster-Szenarien / HAW Hamburg. URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master06-07-proj/piening/report.pdf>, Februar 2007. – Seminarbericht
- [Rekimoto 1996] REKIMOTO, Jun: Tilting operations for small screen interfaces. In: *UIST '96: Proceedings of the 9th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 1996, S. 167–168. – ISBN 0-89791-798-7
- [Roßberger 2007] ROSSBERGER, Philipp: Physikbasierte Interaktion im Collaborative Workspace / HAW Hamburg. Februar 2007. – Seminararbeit
- [Rohs u. a. 2006] ROHS, Michael ; ESSL, Georg ; ROTH, Martin: CaMus: live music performance using camera phones and visual grid tracking. In: *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression*. Paris, France, France : IRCAM & Centre Pompidou, 2006, S. 31–36. – ISBN 2-84426-314-3

- [Schlichter u. a. 2001] SCHLICHTER, Johann ; REICHWALD, Ralf ; KOCH, Michael ; MÖSLEIN, Kathrin: Rechnergestützte Gruppenarbeit (CSCW). In: *i-com Zeitschrift für interaktive und kooperative Medien* 0 (2001), Nr. 1, S. 5–11
- [Scott 2005] SCOTT, Stacey D.: *Territoriality in collaborative tabletop workspaces*. Calgary, Alta., Canada, Canada, Dissertation, 2005. – URL <http://portal.acm.org/citation.cfm?id=1123527>
- [Shizuki u. a. 2006] SHIZUKI, Buntarou ; HISAMATSU, Takaomi ; TAKAHASHI, Shin ; TANAKA, Jiro: Laser pointer interaction techniques using peripheral areas of screens. In: *AVI '06: Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA : ACM Press, 2006, S. 95–98. – ISBN 1-59593-353-0
- [Slay und Thomas 2006] SLAY, Hannah ; THOMAS, Bruce: Interaction and visualisation across multiple displays in ubiquitous computing environments. In: *Afrigraph '06: Proceedings of the 4th international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*. New York, NY, USA : ACM, 2006, S. 75–84. – ISBN 1-59593-288-7
- [Sonka u. a. 2007] SONKA, Milan ; HLAVAC, Vaclav ; BOYLE, Roger: *Image Processing, Analysis, and Machine Vision*. Thomson-Engineering, 2007. – 129 S. – ISBN 049508252X
- [Sutherland 1964] SUTHERLAND, Ivan E.: Sketch pad a man-machine graphical communication system. In: *DAC '64: Proceedings of the SHARE design automation workshop*. New York, NY, USA : ACM Press, 1964, S. 6.329–6.346
- [Teufel u. a. 1995] TEUFEL, Dr. S. ; SAUTER, Christian ; MÜHLHERR, Thomas ; BAUKNECHT, Prof. Dr. K.: *Computerunterstützung für die Gruppenarbeit*. Addison-Wesley, 1995. – 278 S
- [Tse u. a. 2006] TSE, Edward ; SHEN, Chia ; GREENBERG, Saul ; FORLINES, Clifton: Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In: *AVI '06: Proceedings of the working conference on Advanced visual interfaces*. New York, NY, USA : ACM, 2006, S. 336–343. – ISBN 1-59593-353-0
- [Vogel und Balakrishnan 2005] VOGEL, Daniel ; BALAKRISHNAN, Ravin: Distant freehand pointing and clicking on very large, high resolution displays. In: *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 2005, S. 33–42. – ISBN 1-59593-271-2
- [Wadlow 1981] WADLOW, Thomas A.: The Xerox Alto computer. In: *BYTE* 9 (1981), S. 58–68. – URL <http://www.guidebookgallery.org/articles/thexeroxaltocomputer>

- [Wahlster WS 2000/2001] WAHLSTER, Wolfgang: *SmartKom Leitvorstellung*. Webseite. WS 2000/2001. – URL <http://w5.cs.uni-sb.de/teaching/ws00/folien/sld197.htm>
- [Wang u. a. 2006] WANG, Jingtao ; ZHAI, Shumin ; CANNY, John: Camera phone based motion sensing: interaction techniques, applications and performance study. In: *UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology*. New York, NY, USA : ACM Press, 2006, S. 101–110. – ISBN 1-59593-313-1
- [Weiser 1991] WEISER, Mark: The computer for the 21st century. In: *SIGMOBILE Mob. Comput. Commun. Rev.* 3 (1991), Nr. 3, S. 3–11. – ISSN 1559-1662
- [Wilson 1991] WILSON, Paul: *Computer Supported Cooperative Work: An Introduction*. Intellect Books, Oxford, UK, 1991
- [Zhai u. a. 1993] ZHAI, Shumin ; MILGRAM, Paul ; DRASCIC, David: An evaluation of four 6 degree-of-freedom input techniques. In: *CHI '93: INTERACT '93 and CHI '93 conference companion on Human factors in computing systems*. New York, NY, USA : ACM Press, 1993, S. 123–125. – ISBN 0-89791-574-7

A. Anhang

A.1. Photogrammetrie

Photogrammetrie (Bildmessung) beschreibt eine Sammlung von Methoden, mit denen eine Rekonstruktion der räumlichen Lage bzw. der dreidimensionalen Form eines Objektes aus einer photographischen Abbildung möglich ist. Zum Einsatz kommen diese berührungslosen Verfahren beispielsweise bei der passiven Fernerkundung und Vermessung.

Die Abbildung der Raumpunkte auf die Bildebene erfolgt als Parallel- oder Zentralprojektion (siehe Abbildung TODO). Eine weitere Differenzierung erfolgt über die Lage der Ebenen zueinander, demzufolge gibt es vier Arten der Projektion:

Perspektive Kollineation Zentralprojektion zwischen Ebenen allgemeiner Lage, es gelten: Inzidenz, Doppelverhältnis

Perspektive Affinität Parallelprojektion zwischen Ebenen allgemeiner Lage, es gelten: Inzidenz, Parallelität, Teilverhältnis

Ähnlichkeit Zentralprojektion zwischen parallelen Ebenen, es gelten: Inzidenz, Parallelität, Teilverhältnis, Winkel

Kongruenz Parallelprojektion zwischen parallelen Ebenen, es gelten: Inzidenz, Parallelität, Teilverhältnis, Winkel, Flächeninhalt

Bei der Verwendung einer Kamera entsteht eine zentralperspektive Abbildung (perspektive Kollineation). Das Sensormodul der Kamera ist die Bildebene, senkrecht auf ihr steht die optische Achse, das Projektionszentrum (Hauptpunkt) ist der Schnittpunkt aller Abbildungsstrahlen. Die Objektebene befindet sich in beliebiger Lage zur Bildebene, veranschaulicht in [A.1](#).

Wie durch [Meisel \(2006\)](#) beschrieben, wird das Kameramodell der gesamten Projektion durch innere und äußere Faktoren bestimmt. Die **inneren Faktoren** sind kameraspezifisch und können nicht direkt gemessen werden. Die Bestimmung erfolgt indirekt durch eine Kalibrierung, mit aufwändigen Algorithmen wird das Bild einer Kalibriervorlage analysiert. Der Bildhauptpunkt, welcher gleichzeitig den Ursprung des Bildkoordinatensystems bilden soll,

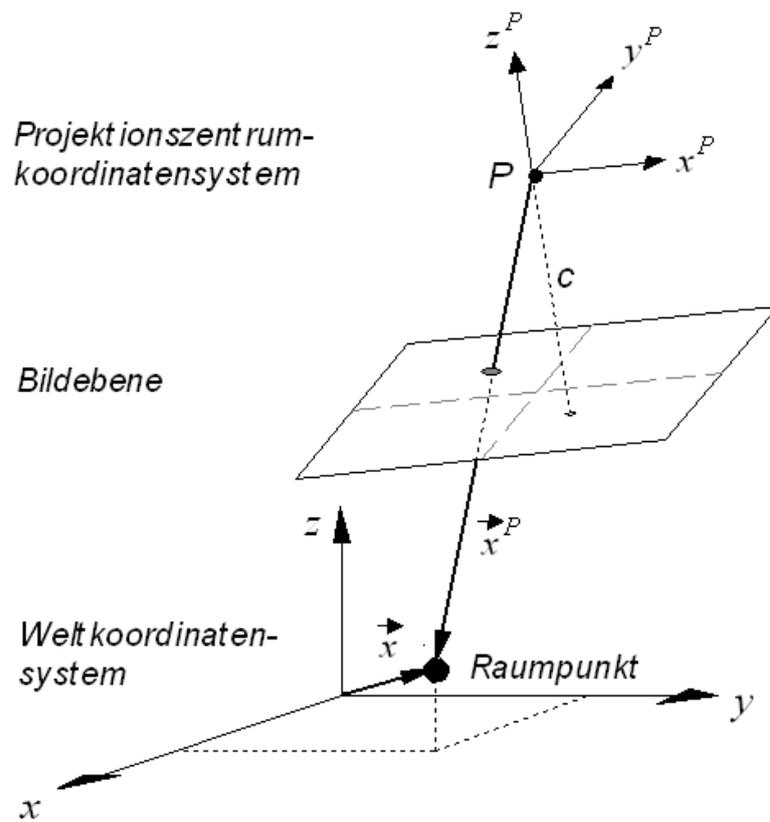


Abbildung A.1.: Weltkoordinatensystem und Kamerakoordinatensystem

wird idealerweise in der Mitte der Sensormatrix angenommen. Die Einbaulage der Sensormatrix und des Objektivs sowie die Asymmetrie des Objektivs sind Grund für eine unbekannte Abweichung vom Bildhauptpunkt, der **Bildhauptpunktverschiebung**. Die **Kammerkonstante**, bekannt als Bildweite, gibt den Abstand von Projektionszentrum zur Bildebene an. Weitere **Linsenverzeichnungsparameter** sind notwendig für die Korrektur eines realen, verzeichneten Bildpunktes in einen verzeichnungsfreien Bildpunkt. Die Korrektur erlaubt in der Weiterverarbeitung die Annahme, dass es sich wirklich um ein Resultat aus der Zentralperspektive handelt. Böhm (2007) zeigt in Abbildung A.2 ein verzeichnetes und ein korrigiertes Bild.



Abbildung A.2.: Verzeichnung und Korrektur

Die äußeren Parameter beschreiben die Orientierung und Position der Ebenen zueinander. Gemäß Abbildung A.1 wird die **Position der Kamera im Raum** durch den Translationsvektor \vec{t} gekennzeichnet und die **Orientierung der Kamera im Raum** beschreibt die Rotationsmatrix R (Verdrehung der beiden Koordinatensysteme zueinander über die primäre, sekundäre und tertiäre Achse).

$$R = \begin{pmatrix} \cos\beta \cos\gamma & \cos\alpha \sin\gamma + \sin\alpha \sin\beta \cos\gamma & \sin\alpha \sin\gamma - \cos\alpha \sin\beta \cos\gamma \\ -\cos\beta \sin\gamma & \cos\alpha \cos\gamma - \sin\alpha \sin\beta \sin\gamma & \sin\alpha \cos\gamma + \cos\alpha \sin\beta \sin\gamma \\ \sin\beta & -\sin\alpha \cos\beta & \cos\alpha \cos\beta \end{pmatrix}$$

Sind diese bekannt, können Punkte aus der Bildebene in den Objektraum und umgekehrt transformiert werden. Notwendig hierfür sind die allgemeinen Abbildungsgleichungen der

Zentralperspektive:

$$u = \Delta u_H - c * \frac{r_{11} * (x - t_x) + r_{12} * (y - t_y) + r_{13} * (z - t_z)}{r_{31} * (x - t_x) + r_{32} * (y - t_y) + r_{33} * (z - t_z)} \quad (\text{A.1})$$

$$v = \Delta v_H + c * \frac{r_{21} * (x - t_x) + r_{22} * (y - t_y) + r_{23} * (z - t_z)}{r_{31} * (x - t_x) + r_{32} * (y - t_y) + r_{33} * (z - t_z)} \quad (\text{A.2})$$

mit

- u: Bildkoordinate x
- v: Bildkoordinate y
- Δu_H : Bildhauptpunktverschiebung x
- Δv_H : Bildhauptpunktverschiebung y
- r_{11} - r_{33} : Elemente der Rotationsmatrix

Versicherung über Selbstständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit im Sinne der Prüfungsordnung nach §24(5) ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, 22. November 2007

Ort, Datum

Unterschrift