AW1-Ausarbeitung
3D Interaktionen in Smart Homes

Edo Kriegsmann

Contents

1 Einleitung

”Die Technisierung des Lebensraums zur Steigerung von Aspekten wie Wohnkomfort und Werbeerfolg schreitet laufend voran. Ein Teilgebiet dieser Technisierung ist das Ermitteln und Bewerten von menschlichen Interaktionen und Emotionen. In diesem Kontext entstand im Jahr 2010 das ”Living Place” Projekt an der HAW, in welchem es unter anderem möglich ist, zukünftige Entwicklungen unter realen Bedingungen zu erproben (vgl. Rahimi;Voigt, 2010) . Zu diesen Entwicklungen gehören Bilderkennungssysteme, wie reguläre Farb-Kameras, als auch TTL-Kameras und seit einiger Zeit die Farb- und 3D-Kamera „Microsoft Kinect“ (vgl. Microsoft, 2011b) .” (vgl. Kriegsmann, 2011) Aus diesem Bereich entsteht der Kern dieser Ausarbeitung. So sollen die 3D- und Farbinformationen der Kinect Kamera verarbeitet und zur Ermittlung menschlicher Interaktion, Emotion sowie dessen Alters und Geschlechts genutzt werden. Die darauf aufbauenden Anwendungen sind in den Bereich von ”Ambient Intelligence”, ”Ubiquitous Computing” sowie ”Mensch-Companion-Interaktion” einzuordnen - Mindmap siehe Bild 1.1. ((vgl. GI-Jahrestagung, 2011) , (vgl. Weiser, 1991))

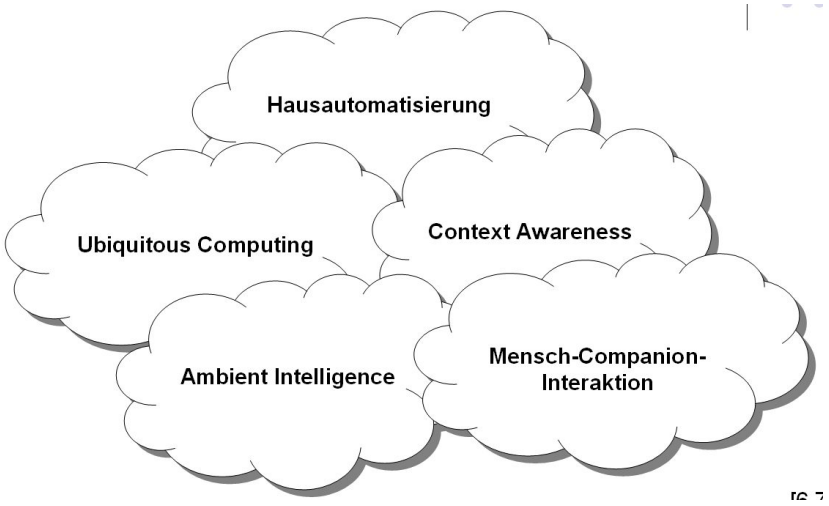

Figure 1.1: *Bereiche der Arbeit*

2 Zielsetzung

Die Zielsetzung der Aus- und Bewertung der Kinect-Kameradaten sieht vor, eine solide Datenbasis aufsetzender Anwendungen zu schaffen. So sollen die 3D-Tiefeninformationen zur Analyse der (Körper-)Bewegungen, darauf aufbauend zur Erkennung von Bewegungsmustern und -Abläufen, genuzt werden. Auch einfach mess- und quantisierbare Parameter, wie beispielsweise der Abstand zum System oder die Größe der erkannten Person können über diese Tiefendaten ermittelt werden. Weiterhin soll über die RGB-Farbbildinformationen eine Auswertung der Körpermerkmale wie Alter, Geschlecht und Emotion/Gemütszustand der erkannten Person erfolgen. Bei letzerem liegt gerade die Einordnung bzw. Bewertung der Emotionen im Aufgabenfeld dieser Arbeit.

2.1 Denkbare Szenarien

Folgende Szenarien wären mit den Informationen, welches dieses System bereitstellen könnte, denkbar:

Szenario eins: Ein mit diesem System ausgestatteter Fahrkartenautomat könnte etwa Rollstuhlfahrer erkennen und diesen einen barrierefreien Weg zum Zug vorschlagen. Zudem könnte der Bildschirminhalt des Fahrkartenautomaten dem Alter der bedienenden Person angepasst werden. Bei höherem Alter könnten Bedienelemente auf dem Touchscreen vergrößert werden, um ihnen die Bedienung zu erleichtern. Andererseits könnte jungeren Benutzern eine höhere technische Affinität unterstellt werden, was komplexere Bedienstrukturen erlauben würde. Selbstverständlich müssen die Anpassungen so erfolgen, dass auch bei fehlerhafter Erkennung die Bedienung durch die Anpassung nicht erschwert wird, oder gar der Benutzer sich subjektiv diskriminiert fühlt.

Szenario zwei: Eingesetzt in reaktiven Werbetafeln könnte dies System die Informationen einer dynamischen Anpassung der Tafel liefern (siehe Grafik 2.1). So könnten die Körpermerkmale vorbeigehender Personen analysiert und der Werbekontext entsprechend dieser angepasst werden. Eine Mutter mit Kind würde als solche erkannt und der Werbekontext daraufhin angepasst. Die Ermittlung des Alters und Geschlechts erfolgt über die Auswertung der Farb-Kamerainformationen. Während dem Kind auf Augenhöhe etwa eine kindgerechte Bonbonwerbung präsentiert wird, so suggeriert man der Mutter auf ihrer Augenhöhe, die Vorteile eines Kaufs des Produktes. Dies selbstverständlich mit entsprechend auf diese angepasste Werbeinformationen.

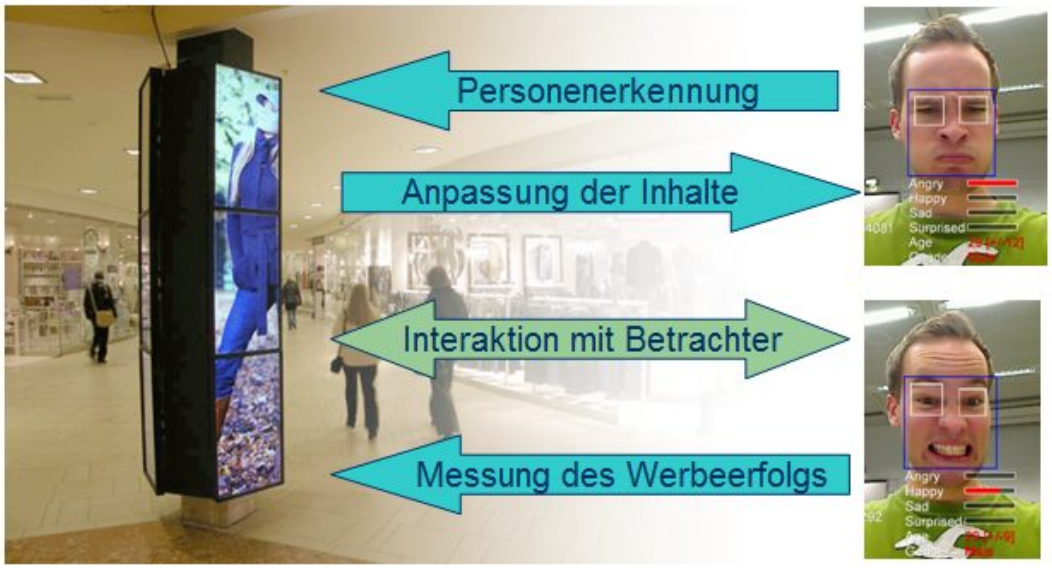

Figure 2.1: *Konzept Werbetafel*

3 Lösungsansätze

Grundlegend besteht die Ausarbeitung aus drei Kernpunkten: Zum einem dem Erfassen der Daten: Dies sind die 3D-Tiefen- und die RGB-Kamerainformationen. Der zweite Teil beschäftigt sich mit dem Bündeln und Vorverarbeiten dieser Daten, um aus diesen einen nutzbaren Informationspool zu erzeugen. Die letzte Schicht beschreibt dann den Gebrauch dieser Informationen durch aufgesetzte Anwendungen. Im folgendem Bild 3.1 ist dies grafisch aufbereitet dargestellt:

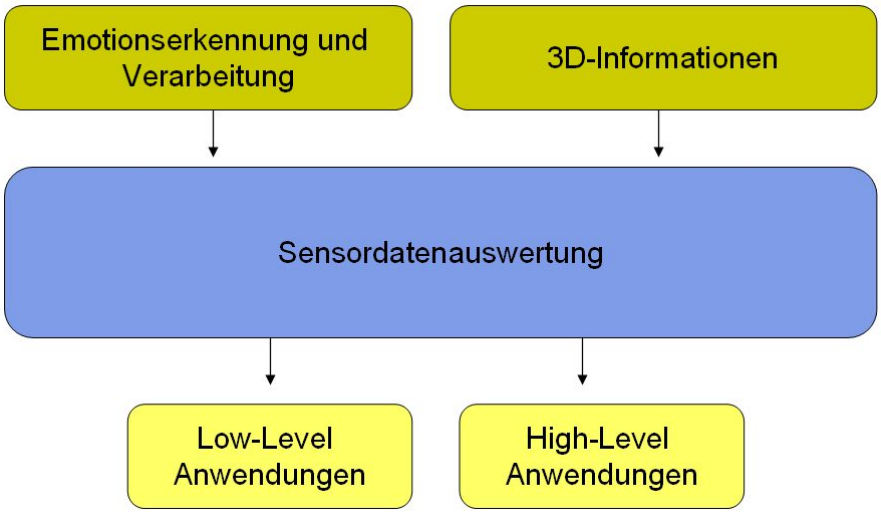

Figure 3.1: *Lösungsentwurf*

3.1 Erfassen der Daten

Zu allererst müssen die notwendigen Daten über die Sensorik ermittelt werden. Dies sind die RGBD-Daten[1], welche mittels der Microsoft-Kinect-Kamera (vgl. Microsoft, 2011b)

[1] RGBD - Red Green Blue Depth, gebündelte Farb- und Tiefenbildinformationen

aufgezeichnet werden. Diese Daten enthalten zu dem 3D-Tiefenbild auch die RGB-Bildinformationen, zur Auswertung der Emotionen und Körpermerkmale. Aufgrund vorhandener Bibliotheken - die NUI-Bibliothek[2] der Microsoft Kinect (vgl. Microsoft, 2011a, Seite 14) und der SHORE-Bibliothek des Frauenhofer Instituts (vgl. IIS, 2011) - wird sich dies voraussichtlich verhältnismäßig einfach gestalten. Die Bibliotheken werde ich im Abschnitt "Komponenten" auf Seite 10 noch etwas genauer vorstellen.

3.2 Bündeln, Reduzieren und Vorverarbeiten der Daten

Darauf folgend müssen die Informationen aus diesen Daten extrahiert und zu einer Basis gebündelt werden; erst dieser Schritt macht eine gemeinsame Weiterverarbeitung möglich. Die Einordnung schwer quantifizierbarer Messgrößen wie menschlicher Emotion stellt hier eine Herausforderung dar. So ist eine Eingrenzung auf wenige, signifikante Merkmale (Beispielsweise Freude und Wut) und eine Reduktion des Ergebnisraums (Beispielsweise: Viel Wut - neutral - wenig Wut) unausweichlich. Eine Grafik (siehe Bild 3.2) des Frauenhofer Instituts zeigt einen solchen Emotionsraum, welcher allerdings noch nicht reduziert ist. Auf Seiten der Software liegt die Schwierigkeit tendentiell in den unterschiedlichen Entwicklungsumgebungen, wenngleich hierfür sicher eine Lösung gefunden werden kann. Es ließen sich die Daten über einen asynchronen Nachrichtendienst, wie beispielsweise ActiveMQ, von einer Umgebung in die nächste transferieren.

3.3 Zweigeteilte Nutzung der Informationen

Zuletzt kann dann auf Basis dieses Datenpools eine Anwendung aufgesetzt werden, welche einen (Kunden-)Nutzen hervorbringt. Dabei gehen die Überlegungen hier in eine zweigeteilte Lösung. Zum einen die "High-Level"-Anwendungen, welche die Möglichkeiten dieses detailierten Informationspools voll und in Echtzeit auswerten können. Dies könnte eine, wie auf Seite 5 beschriebene Werbetafel sein. Der Nachteil dieser Anwendungen liegt in der erforderlichen Rechenleistung, welche ausschließt, dass einfache (mobile) Hardware-devices diese Informationen nutzen können. Abhilfe soll hier der weitere Entwicklungsweg, die "Low-Level"-Anwendungen darstellen. Eine Vorverarbeitung der Datenflut ermöglicht so auch verhältnismäßig einfacher Hardware die Nutzung dieser Informationen. Hier wären Beleuchtungssysteme vorstellbar, welche sich auf die Emotionen des Nutzers hin einstellen, oder wiederum versuchen eine Emotion zu hervorzurufen.

[2] NUI: natural user interface - Auswertung von Körpermerkmalen über das Tiefenbild

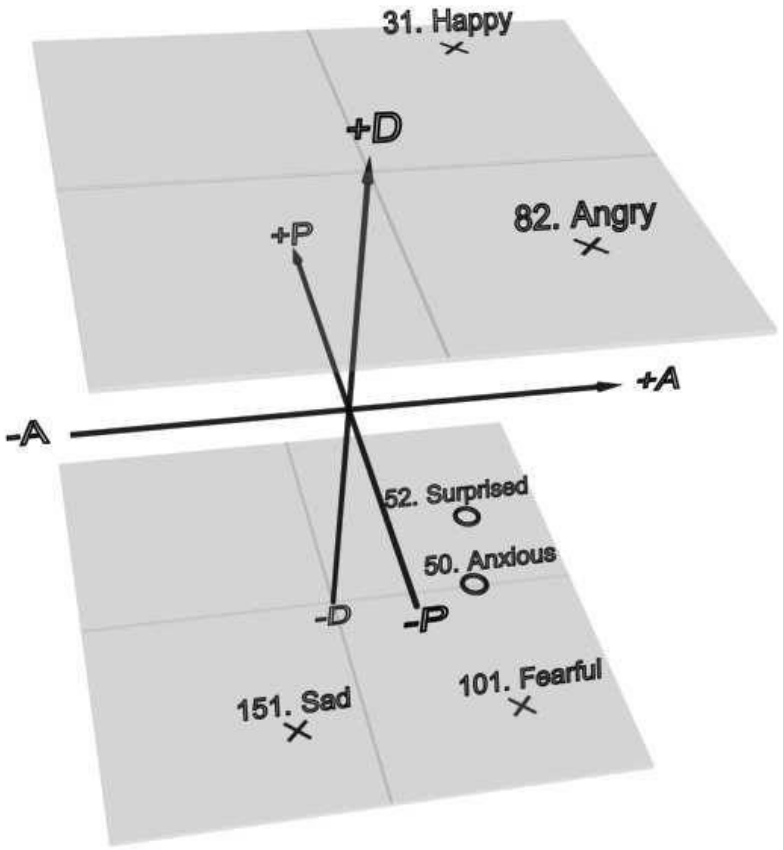

Figure 3.2: *Emotionsraum nach Frauenhofer IIS (vgl. Fraunhofer-IIS, 2011a)*

4 Komponenten

In diesem Abschnitt möchte ich auf die Komponenten zu sprechen kommen, welche ich voraussichtlich zur Ermittlung und Verarbeitung der Kameradaten nutzen möchte. Dies sind zum einen die Microsoft Kinect mit dem dazugerörigen SDK, zum anderen die Arbeit des Frauenhofer Instituts, welche die SHORE-Bibliothek entwickelte.

4.1 Microsoft Kinect - NUI-Bibliothek

Zur Ermittlung der RGBD-Daten kommt, wie bereits erwähnt, die Microsoft Kinect zur Verwendung. Diese besitzt zu einer regulären RGB-Farbkamera auch ein System, mit welchem man die Tiefeninformationen eines Bildes auslesen kann (vgl. Microsoft, 2011a, Seite 17) . Dazu wird eine Infrarot-Punktewolke in den zu analysierenden Raum geworfen und anhand des Versatzes dieser Punkte ein Tiefenbild errechnet. Dieses liegt dann in einer Auflösung von bis zu 640x480 Pixeln vor. (vgl. Microsoft, 2011a, Seite 18) Stereoskopie oder die Verwendung von TTL-Technik wird nicht genutzt.

Mit dem SDK, welches seit November 2011 in der zweiten Version vorliegt, ist eine umfangreiche Verwendung dieser Daten möglich. (vgl. Microsoft, 2012) Über die NUI-Bibliothek des SDKs erhält man Zugriff auf den Audio-, Depth- und Imagestream, wie in Grafik 4.1 zu sehen. (vgl. Microsoft, 2011a, Seite 14) Aus letzteren Informationen generiert das SDK zudem die ”Sceleton-Daten” mit Hilfe derer man die Position von 21 Skelettpunkten erkannter Personen ermitteln kann. (vgl. Microsoft, 2011a, Seite 20)

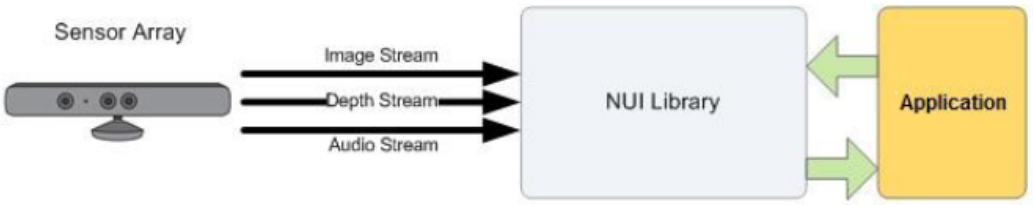

Figure 4.1: *Anbindung der Module in die Software (vgl. Microsoft, 2011a, Seite 14)*

Dieses System schafft so eine solide Basis zur Ermittlung und Verarbeitung der RGBD-Daten. Im Gegensatz zu Stereoskopie- und TTL-Kamerasystemen ist dieser Aufbau zudem sehr kostengünstig.

4.2 Frauenhofer Institut - SHORE-Bibliothek

Die vom Frauenhofer Institut entwickelte Engine names ”SHORE” - Sophisticated High-speed Object Recognition Engine (vgl. Fraunhofer-IIS, 2011b) stellt eine Bibliothek bereit, welche es ermöglicht in Farbbild-Informationen menschliche Gesichter und deren Emotionen zu erkennen. Zudem gehört auch die Erkennung von Alter und Geschlecht zum Funktionsumfang der Bibliothek. Diese Bibliothek nutzt 2D-Farbkamerainformationen mit einer minimalen Fläche von 24x24 Pixeln pro Gesicht zur Ermittlung der Informationen. (vgl. IIS, 2011, Seite 24, Abschnitt ”minFaceSize”) . In folgender Grafik 4.2 ist die Demonstrationssoftware der SHORE-Bibliothek zu sehen.

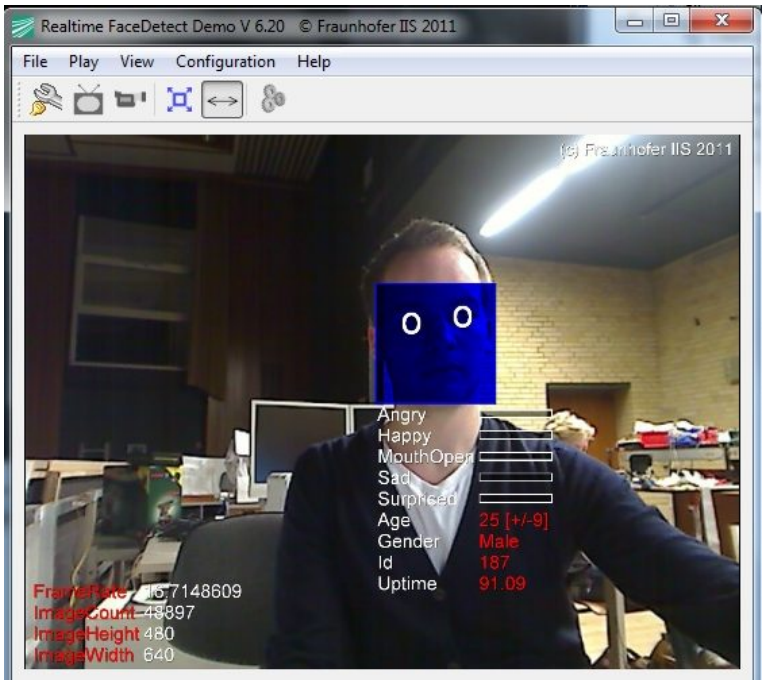

Figure 4.2: *Auswertung der Merkmale durch SHORE*

5 Vergleichbare Arbeiten

Im folgenden Abschnitt möchte ich kurz auf eine prägnante, vergleichbare Arbeit zu diesem Themengebiet eingehen. So beschäftigt sich die Arbeit mit der Auswertung von Kinect-Kamerainformationen zum Erlernen und Wiedererkennen von menschlichen Interaktionen.

Auch in diesem Kontext zu nennen, ist die Dissertation von Herrn Becker-Asano mit dem Titel "WASABI: Affect Simulation for Agents with Believable Interactivity" (vgl. Becker-Asano, 2008) . Die beschäftigt sich mit der Analyse und der Generierung von Emotionen. Eine genauere Betrachtung der Arbeit wäre aber an dieser Stelle zu umfangreich.

5.1 Human Activity Detection from RGBD Images

Die Arbeit von Jaeyong Sung, Colin Ponce, Bart Selman und Ashutosh Saxenai, beschäftigt sich mit dem Erlernen und (Wieder-)Erkennen von menschlichen Aktivitäten auf Basis der Microsoft Kinect (vgl. Sung u. a., 2011) . Das Team um Herrn Sung (vgl. Sung u. a., 2011, Seite 1) verwendet zur Auswertung der Daten noch die Open-Source-Lösung von PrimeSense (vgl. OpenNI, 2011) , während bei neueren Arbeiten inzwischen weitgehend das Microsoft SDK zur Verwendung kommt.

Des Weiteren zeigte diese Arbeit auch andere Lösungsvorschläge und Hilfsmittel zur Erkennung von menschlichen Aktivitäten auf. So wird auf die Nutzung von 2D-Kamerainformationen eingegangen, oder die Hilfe von RFID-Transpondern (vgl. Sung u. a., 2011, Seite 1) erwägt. Letzeren Lösungsansatz verfolgte ich bereits in meiner Abschlussarbeit (vgl. Kriegsmann) .

Figure 5.1: *Diverse Beispiele für ermittelbare Aktivitäten (vgl. Sung u. a., 2011, Seite 06)*

Der von dem Team um Jaeyong Sung entwickelte Lernalgorithmus ermöglichte es, mit einer Präzision von bis zu 86.5% bei bekannten Personen und bis zu 69% bei neuen Personen, eine erlernte Aktivität wieder zu erkennen (vgl. Sung u. a., 2011, Seite 6, Tabelle 1) . Dies zeigt eindrucksvoll die Möglichkeiten der Auswertung von 3D-Kamerainformationen.

Die Ausrichtung der Arbeit um Herrn Sung unterscheidet sich in soweit von dieser, als dass das Team um Herrn Sung zusätzlich zur Erkennung von beliebigen Bewegungsmustern einen Lernalgoritmus zur Erkennung wiederkehrender Aktionen implementiert hat.

6 Chancen, Risiken und Ausblick

Wie bereits im Abschnitt 3.2 beschrieben, stellt die Einordnung ermittelter Emotionen eine Schwierigkeit dar. Im Gegensatz zu einfach quantisierbaren Messgrößen wie der Körpergröße einer Person, sind menschliche Emotionen stark differenziert ausgeprägt und zudem kontextabhängig. Folgende Grafik zeigt deutlich diese Kontextabhängigkeit. Wärend auf dem linken Bild deutlich Wut zu erkennen ist, stellt das rechte Bild - wohlbemerkt, dass es sich um das exakt gleiche Gesicht handelt - eindeutig Ekel dar. Dieser Unterschied ist mittels einer Software, welche ausschließlich Farbbildinformationen eines Gesichts analysiert, nicht ermittelbar.

Figure 6.1: *Ein Gesicht - Zwei Bedeutungen.*

Weiterhin müssen noch präziser ausformulierte Szenarien entworfen werden, welche einen Nutzen aus den so gebündelten Informationen ziehen können. Wenn diese konkretisiert sind, liegen die Chancen der Ausarbeitung in genau dieser Informationsfülle. So sind ”alle” Informationen im Zusammengang mit einem Menschen, welche man einem RGBD-Bild entnehmen kann in einer Plattform gebündelt. Gerade auch die Tatsache, dass diese Informationen in Echtzeit zur Verfügung stehen, eröffnet vielfältige Nutzungsmöglichkeiten.

In dieser Ausarbeitung ist der grundlegende Überblick über die Thematik, sowie ein Einblick in vergleichbare Arbeiten geschaffen worden. So konnten sowohl die Teilziele der Masterarbeit definiert, als auch die dafür notwendigen Komponenten ausgewählt werden. Dies schafft die Basis für eine weitergehende theoretische wie praktische Einarbeitung in das Themengebiet.

Bibliography

[Becker-Asano 2008] BECKER-ASANO: *WASABI: Affect Simulation for Agents with Believable Interactivity*. 2008

[Fraunhofer-IIS 2011a] FRAUNHOFER-IIS: *Emotionsraum*. 2011. – URL http://www.iis.fraunhofer.de/en/bf/bv/ks/gpe/index.jsp. – Zugriffsdatum: 02.02.2012

[Fraunhofer-IIS 2011b] FRAUNHOFER-IIS: *Fraunhofer IIS Website*. 2011. – URL http://www.iis.fraunhofer.de/bf/bv/ks/gpe/. – Zugriffsdatum: 02.07.2011

[GI-Jahrestagung 2011] GI-JAHRESTAGUNG: *Companion-Systeme und Mensch-Companion-Interaktion*. 2011. – URL http://edu.cs.uni-magdeburg.de/EC/konferenzen-und-workshops. – Zugriffsdatum: 14.02.2012

[IIS 2011] IIS, Fraunhofer: *Fraunhofer-IIS - Shoore 1.4*. Fraunhofer IIS, 2011

[Kriegsmann] KRIEGSMANN, Edo: *Kaskadierbare berührungssensitive reaktive Flächen*. HAW-Hamburg. – URL http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/bachelor/kriegsmann.pdf

[Kriegsmann 2011] KRIEGSMANN, Edo: *Projektarbeitung 1 - 3D Interaktionen in Smart Homes*. 2011. – URL http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master2011-proj1/kriegsmann.pdf. – Zugriffsdatum: 18.02.2012

[Microsoft 2011a] MICROSOFT: *Microsoft Kinect SDK Programming Guide*. 2011. – URL http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/docs/ProgrammingGuide_KinectSDK.pdf. – Zugriffsdatum: 02.07.2011

[Microsoft 2011b] MICROSOFT: *Microsoft Kinect SDK Website*. 2011. – URL http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/. – Zugriffsdatum: 02.07.2011

[Microsoft 2012] MICROSOFT: *Microsoft Kinect SDK Website*. 2012. – URL http://www.microsoft.com/germany/msdn/aktuell/news/show.mspx?id=msdn_de_45012. – Zugriffsdatum: 12.02.2012

[OpenNI 2011] OPENNI: *OpenNI*. 2011. – URL http://www.openni.org/. – Zugriffsdatum: 02.07.2011

[Rahimi;Voigt 2010] RAHIMI;VOIGT: *HAW Living Place*. 2010. – URL http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master09-10-proj/rahimi-vogt.pdf. – Zugriffsdatum: 02.07.2011

[Sung u. a. 2011] SUNG, Jaeyong ; PONCE, Colin ; SELMAN, Bart ; SAXENAI, Ashutosh: *Human Activity Detection from RGBD Images*. Department of Computer Science Cornell University, Ithaca, NY, 2011. – URL http://dl.acm.org/citation.cfm?id=1377032.1377113

[Weiser 1991] WEISER, Mark: *The Computer for the Twenty-First Century*. 1991