

Masterseminar
3D Interaktionen in Smart Homes

Edo Kriegsmann

Contents

1	Einleitung	3
2	Vorarbeiten	4
2.1	Skelett-Erkennung und Verarbeitung	4
2.2	Kinect-Audioverarbeitung	4
2.3	Verarbeitung der SHORE-Daten	5
2.4	Fazit der Vorarbeiten	5
3	Ziele	7
3.1	Kernszenario	8
3.2	Weitere Anwendungsmöglichkeiten	9
3.3	Hardwarenahe Nutzung der Informationen	10
4	Risiken	11
4.1	Risiken in der Realisierung	11
4.2	Risiken aus der Benutzersicht	13
5	Zusammenfassung und Ausblick	14
	Bibliography	15

1 Einleitung

Im Masterseminar beschäftigte ich mich mit der Microsoft Kinect-Kamera und dem Potential der Daten, welche man mit dieser gewinnen kann. Aus den RGBD-Informationen der Kinect-Kamera sollen sowohl die 3D-Daten extrahiert, als auch die 2D-Farbbildinformationen mithilfe der Frauenhofer-SHORE-Bibliothek untersucht werden. Die Vereinigung dieser Daten kann dann beispielsweise dazu beitragen, Hausautomatisierungen individueller arbeiten zu lassen und damit intelligente Wohnräume zu ermöglichen. Die folgende Grafik 1.1 zeigt schemenhaft die Bereiche der Arbeit. ((vgl. [GI-Jahrestagung, 2011](#)) , (vgl. [Weiser, 1991](#))) In den vorangegangenen Kursen konnte die Grundlage für diese Aufgabenstellung geschaffen werden. So wurden in "Anwendungen 2" vergleichbare Arbeiten betrachtet und die Überschneidungen zu diesem Projekt herausgearbeitet. In den Projektteilen 1 und 2 wurden daraufhin die Komponenten ausgewählt und weitgehend getestet.

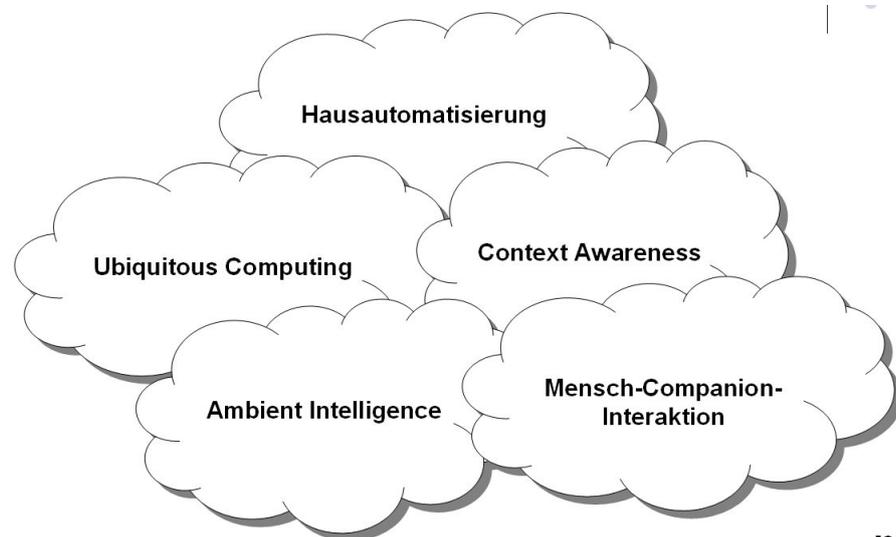


Figure 1.1: Bereiche der Arbeit

2 Vorarbeiten

In Masterprojekt 2 konnte der Großteil der in Projekt 1 offenen Punkte gelöst werden. (vgl. [Kriegsmann, 2011](#)) Mit der so geschaffenen Basis steht sowohl eine direkte Nutzung der 3D-, als auch der Emotions-Daten in einer Entwicklungsumgebung zur Verfügung. Im folgenden Abschnitt möchte ich kurz auf diesen Projektteil eingehen.

2.1 Skelett-Erkennung und Verarbeitung

Die Verarbeitung der 3D-Daten erfolgt mithilfe des Kinect-SDKs und funktioniert mit der neusten SDK-Version einwandfrei. (vgl. [Microsoft, 2013](#)) Ein Mapping des Koordinatensystems der Skelettinformationen auf das Koordinatensystem der RGB-Bildinformationen ist von Beginn an möglich. Dies erlaubt es, in der Anwendungsebene augmented-reality-Elemente¹ einzubringen. Auch lassen sich die wichtigsten XYZ-Koordinaten des menschlichen Skeletts, welches mithilfe der Kamera über die erkannten Personen gelegt wird, erkennen und verarbeiten.

2.2 Kinect-Audioverarbeitung

Mit dem Update des Kinect-SDKs stand im Projekt 2 die Kinect-Audioverarbeitung zur Verfügung, welche im Projekt 1 noch nicht betrachtet wurde. Die Demonstrationsprogramme des SDKs zeigten eindrucksvoll die Möglichkeiten dieses Instruments. So konnte zum einen präzise der Winkel einer Audioquelle in Relation zur Kinect ermittelt, als auch Sprache erkannt werden. Dies zeigt vielfältige Anwendungsmöglichkeiten und könnte als zusätzliche Informationsquelle für den Datenpool genutzt werden. Folgende Bildschirmgrafik zeigt die Software, welche das Wort "rot" aus einer Auswahl möglicher Wörter erkannte und zudem die Richtung der Geräuschquelle lokalisierte (blaue Linie).

¹Augmented Reality ist eine Kombination aus wahrgenommener und vom Computer erzeugter Realität

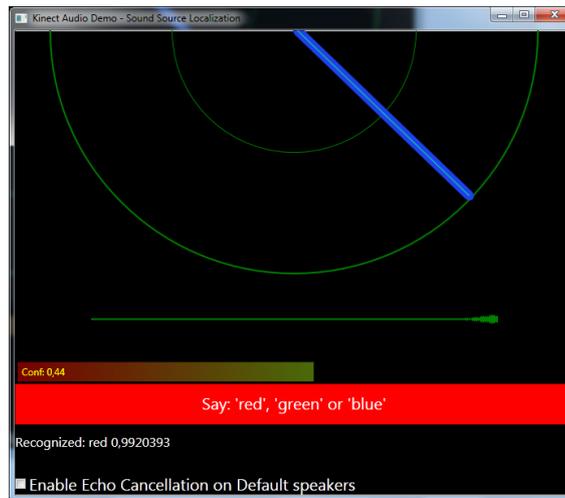


Figure 2.1: Richtung einer Geräuschquelle inkl. Spracherkennung

2.3 Verarbeitung der SHORE-Daten

Die Erfassung der Daten erfolgt wie in den vorangegangenen Projektteilen über die Bibliothek des Fraunhofer Instituts, der "SHORE"-Engine (Sophisticated High-speed Object Recognition) (vgl. IIS, 2011) . Über einen Active-MQ-Wrapper ist es in Projekt 2 gelungen, die Daten dieser Bildauswertung auf einem Active-MQ-Server abzulegen und zur Weiterverarbeitung bereitzustellen. Die Einordnung von Daten wie Alter und Geschlecht funktioniert im gegebenen Rahmen tadellos - die Emotionsdaten werden prozentual zu der Wahrscheinlichkeit der erkannten Emotion übergeben. Im Beispiel-Bild Bild 2.2 wird das Alter durch die SCORE-Bibliothek ermittelt, mithilfe eines Active-MQ-Systems übertragen und in C# wieder ausgegeben.

2.4 Fazit der Vorarbeiten

Insgesamt war es zum Ende des Masterprojekts 2 möglich, die verknüpften Daten in einer Entwicklungsumgebung zu nutzen. In folgendem Beispiel (vgl. Kriegsmann, 2012) wurden die 3D-Daten zur Generierung der Koordinaten eines "Laserschwertes" genutzt, diese in das 2D-Farbbild gemapped und die Einfärbung des "Laserschwertes" in Abhängigkeit des erkannten Geschlechts variiert:

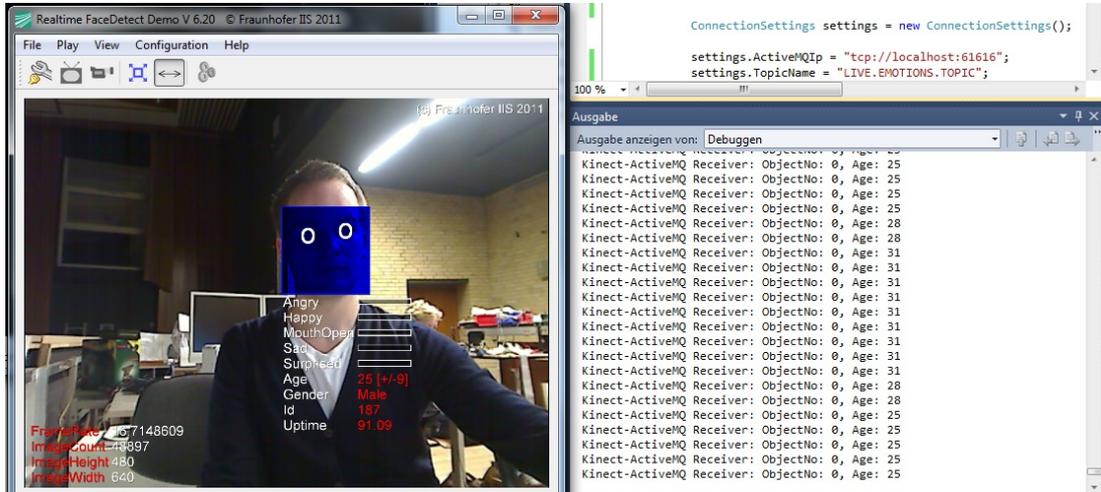


Figure 2.2: SHORE-Ausgewertete Daten in C#

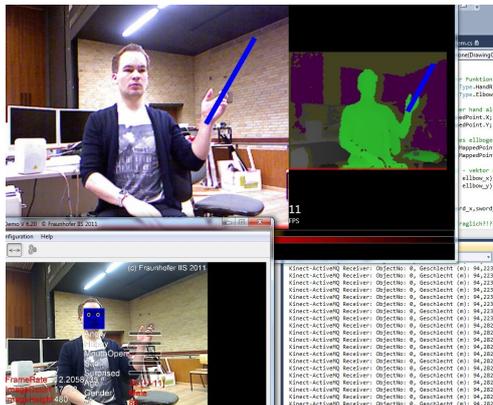


Figure 2.3: Laserschwertfärbung Mann

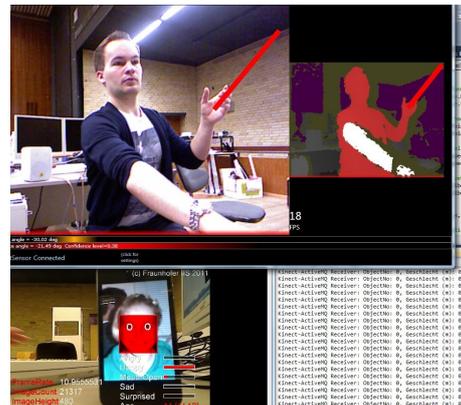


Figure 2.4: Laserschwertfärbung Frau

3 Ziele

Das Ziel der Masterarbeit ist es nun, die gesammelten Informationen aufzubereiten und in eine Anwendung zu überführen, welche dem Anwender einen Nutzen bereitstellt. Hierzu muss eine Sensordatenauswertung der 3D- und Emotionsdaten erfolgen, die dann in einer nachfolgenden Anwendung verarbeitet wird. Dies können zum einen Anwendungen auf Hochsprachenbasis sein, als auch Hardwarenahe Aktoren, welche vorverarbeitete Datenpakete nutzen. Das folgende Schaubild zeigt die Komponenten des Projekts.

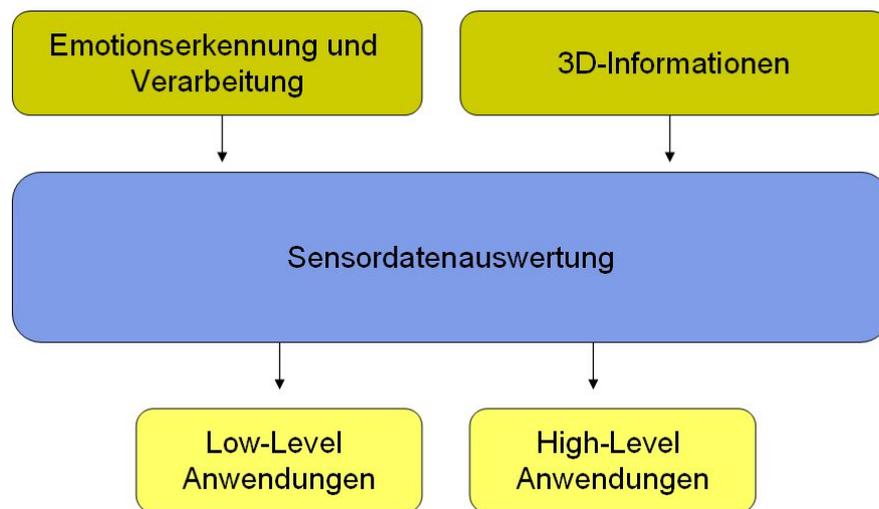


Figure 3.1: Struktur

Hierbei ist das im folgenden Absatz beschriebene Szenario ein gutes Beispiel für die praktische Anwendung dieser Technik und wird zum Kernbereich der Abschlussarbeit werden.

3.1 Kernszenario

Der Bewohner eines mit diesem System ausgestatteten Hauses betritt nach Feierabend seine Wohnung.

Ist es möglich, durch Erweiterung dieses Datenpools um 3D-Informationen und die Emotionserkennung, das Verhalten und die Stimmung des Nutzers vorherzusagen und damit im passenden Moment (Bewohner setzt sich auf das Sofa und geht keiner anderen Aufgabe nach) den Fernseher einzuschalten, sowie den der Stimmung des Nutzers angepassten Fernsehkanal zu wählen?

Das im der Wohnung integrierte System erfasst bereits zentral eine Reihe von Daten, wie die lokalen Wetterinformationen, den persönlichen Terminkalender, das gewohnte Fernsehprogramm, die alltäglichen Gewohnheiten und den Schlafrhythmus des Bewohners - "Context Awareness". (vgl. [Rahimi, 2010](#)) Beim Betreten der Wohnung erfolgt zudem eine Emotionserkennung durch dieses System.

Diese gesammelten Daten könnten in Hinblick auf die Fragestellung unter der Berücksichtigung folgender Aspekte ausgewertet werden.

- Wie viele Termine hatte der Bewohner, handelte es sich um einen anstrengenden Tag? Dieser Aspekt spielt dabei eine Rolle, ob der Bewohner lieber "leichte Kost" konsumieren möchte oder ob er noch in der Verfassung ist einen langen Spielfilm zu schauen. Hierbei könnten auch die Termine des folgenden Tages in Hinblick auf die Länge einer Serie bzw. eines Spielfilms berücksichtigt werden.
- Zu welcher Entscheidung kommt die Emotionserkennung? Hierbei könnten Verknüpfungen zwischen den alltäglichen Fernsehgewohnheiten und der dabei erkannten Emotion gebildet werden. Schaut der Benutzer bei guter Laune etwa gerne einen Musiksender, so könnte bei erkannter guter Laune beim Betreten der Wohnung dieser Musiksender direkt eingeschaltet werden.
- Auch könnte die alltäglichen Gewohnheiten sowie Fernsehgewohnheiten (mithilfe des EPG) in der Art ausgewertet werden, dass vom Nutzer präferierte Sendungen direkt - auch ohne dass er es manuell einstellen muss - aufgezeichnet werden, sollte er nicht zuhause sein. Wenn der Nutzer kurz nach Beginn der Sendung das Haus betritt, so könnte diese Sendung direkt abgespielt werden, wenn der Bewohner sich auf das Sofa setzt und seine durch die Emotionserkennung erkannte Stimmung dafür spricht.

- Wird durch die Emotionserkennung beim Betreten der Wohnung schlechte Laune wie beispielsweise Wut erkannt und spricht auch die Auswertung des Terminkalenders dafür, dass der Bewohner einen anstrengenden Tag hatte, so sollte das System die Entscheidung treffen, den Fernseher gar nicht erst einzuschalten.

Wie aus den oben stehenden Beispielen hervorgeht, handelt es sich hierbei um ein komplexes Problem, bei dem viele Faktoren und auch die Gewohnheiten des Bewohners einzubeziehen sind.

3.2 Weitere Anwendungsmöglichkeiten

Neben diesem Kernszenario, welches das Ziel der Masterarbeit darstellen wird, sind aber auch Anwendungsmöglichkeiten in der Werbebranche gut vorstellbar. Werbetafeln könnten sowohl alters- und geschlechtsspezifische Werbung präsentieren, als sich aber auch über die Anreicherung mit Emotionsdaten der Stimmung der Betrachter anpassen (siehe Grafik 3.2). Hier offenbart sich allerdings auch schnell eines der Risiken dieses Systems: Der Datenschutz. Bei einem Feldversuch im Jahr 2010 wurden 27 individualisierte Werbetafeln in einer Japanischen U-Bahn aufgebaut, woraufhin ähnliche Kritik geäußert wurde (vgl. [Digital-Signage-Promotion-Project, 2010](#)). Die Betreiber dieses Kunstprojektes versicherten jedoch, keine Daten dauerhaft zu speichern und nur die Merkmale einer Gruppe von Personen auszuwerten.

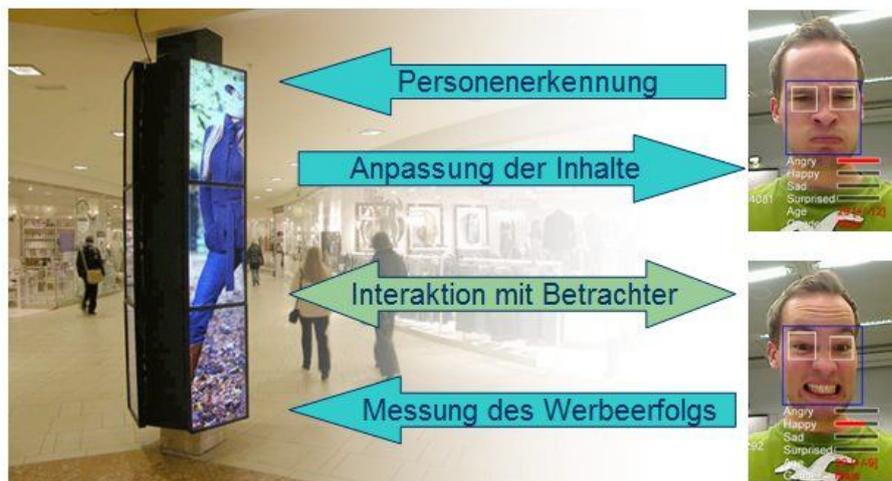


Figure 3.2: Auswertung der Körpermerkmale in Werbetafeln

3.3 Hardwarenahe Nutzung der Informationen

Ein weiteres Ziel der Masterarbeit ist die hardwarenahe Nutzung des Datenpools. Hier ist eine Reduzierung und Vorverarbeitung der Daten von Nöten, welche es Aktoren mit geringer Rechenleistung erlaubt, diesen 2D-/3D-/Emotions-Datenpool zu nutzen. Ein einfaches Beispiel wäre eine dezentral arbeitende LED-Beleuchtung, welche über den Zugriff auf vorverarbeitete Emotionsdaten die Beleuchtung - Lichtfarbe - eines Raumes der Stimmung der Bewohner anpasst. Die Bewegungserkennung könne dabei das Licht gezielt auf den Punkt lenken, an dem dies Benötigt wird, ohne den Benutzer dabei zu blenden.

Im Projekt 2 wurde hier noch das Entwicklungsboard FEZ-Spiter favorisiert (vgl. [Electronics, 2012](#)) . "Mit einem 72 MHz 32Bit-Prozessor verfügt das FEZ-Board eine deutlich größere Rechenleistung als beispielsweise die Arduino-Plattform mit 16 MHz. Dies erlaubt auch komplexere Informationen in Echtzeit zu verarbeiten." (vgl. [Kriegsmann, 2012](#), Seite 12) Inzwischen geht hier aber die Präferenz hin zu einer Eigenentwicklung, welche im Gegensatz zu einer allgemeinen Entwicklungsplattform auf diese Anwendung massgeschneidert werden könnte.

4 Risiken

Im folgenden Abschnitt möchte ich auf die Risiken im Zusammenhang mit der Aufgabenstellung dieser Ausarbeitung eingehen. Dies sind zum einen die Risiken in der Umsetzbarkeit der Technik, zum anderen die möglicherweise fehlende Akzeptanz der Lösung beim Benutzer und nicht zuletzt auch die rechtlichen Risiken im Bereich des Datenschutzes.

4.1 Risiken in der Realisierung

Das Kernproblem dieser Aufgabe ist die sinnvolle Einordnung und Reduzierung der gesammelten Daten sowie deren Verarbeitung. Sempel quantisierbare Messgrößen wie Entfernung der Person zur Kinect-Kamera sowie das Tracking der Körperpunkte sind problemlos möglich. Komplexität gewinnt diese Aufgabe bei der Einordnung von Emotionen. Herr Becker-Asano beschäftigt sich in seiner Doktorarbeit ausführlich mit diesem Thema und stieß auf den gleichen Sachverhalt (vgl. [Becker-Asano, 2008](#), Seite 23 bis 31). Folgende Grafik 4.1 aus der Arbeit von Herrn Becker-Asano zeigt die Vieldimensionalität des Emotionsraums, woraus schnell ersichtlich wird, dass eine simple Überführung in boolesche Variablen schlicht unmöglich ist.

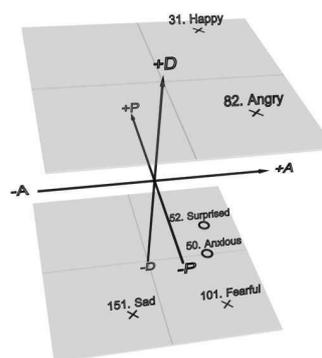


Figure 4.1: Vieldimensionaler Emotionsraum

Ein Beispiel für die Uneindeutigkeit der Emotionserkennung wäre in diesem Szenario die Unterscheidung von Wut und Ekel. Folgende Grafik zeigt diese Kontextabhängigkeit von Emotionen anschaulich. Während auf dem linken Bild der Herr Wut vermittelt, so ist die Emotion des rechten Bildes - wohlgemerkt, dass es sich um das exakt gleiche Gesicht handelt - eher dem Ekel zuzuordnen. Mittels einer Software, welche ausschließlich Farbbildinformationen eines Gesichts analysiert, ist dieser Unterschied nicht ermittelbar. So könnten Fehlerkennungen des Systems zu Fehlentscheidungen führen, welche ein Nutzer eines Produktivsystems nicht lange tolerieren würde.



Figure 4.2: Ein Gesicht - Zwei Bedeutungen. (vgl. [el Kaliouby R. Picard R.W. Hoque, 2009](#))

4.2 Risiken aus der Benutzersicht

Risiken bzw. Schwierigkeiten diese neuartige Technik "in den eigenen vier Wänden" anzunehmen, können sich aus Sicht der Benutzers auf zwei Ebenen ergeben.

Psychologischer Aspekt

Beim Einsatz des Systems kommt es zwangsläufig dazu, dass dem Nutzer Entscheidungen, wie das Einschalten des Fernsehers und der Auswahl des Fernsehprogrammes, beim Betreten der Wohnung abgenommen werden. Auf diese Weise könnte aufgrund der Tatsache, dass ein Computer eine Entscheidung für den Nutzer trifft - welche unter Umständen auch unzutreffend sein könnte - zu einem Gefühl des Kontrollverlustes kommen. Statt einer beabsichtigten Erleichterung für den Benutzer und der Schaffung von mehr Komfort könnte die gefühlte Abgabe von Kontrolle zu einer Ablehnung des Systems führen.

Technischer Aspekt

Neben der eben erläuterten Problematik des Kontrollverlustes ist es bei technisch weniger affinen Personen denkbar, dass sie eine Überforderung durch die technische Komplexität des Systems erleiden. Solange das System die gegebenen Informationen richtig auswertet, spielt dieser Faktor keine Rolle. Lediglich bei einer falschen Erkennung oder Deutung der Daten, welche wiederum beispielsweise zum Einschalten des falschen Fernsehprogrammes führt, müsste der Benutzer das Ergebnis händisch korrigieren. Dies wiederum erfordert den direkten Kontakt zum System, widerspricht damit dem "Ubiquitous Computing" und könnte beim Benutzer ebenfalls dazu führen, dass er bei mehrfachen Falschentscheidungen durch das System, welche durch ihn nicht ohne weiteres zu korrigieren sind, dem System ablehnend gegenüber steht.

Beide Risiken, welche beim Benutzer zur Ablehnung des Systems führen könnten, könnten dadurch verringert werden, dass das die Zahl von falschen Entscheidungen möglichst gering ist. Eine gut funktionierende Erkennung und Auswertung der Ergebnisse führt zu einem erhöhtem Komfortgefühl beim Nutzer, er müsste keine Korrekturen manuell durchführen und das System würde sich unauffällig in seine Umgebung integrieren - "Ambient Intelligence".

Neben den möglichen Risikofaktoren beim Einsatz des Systems in Wohnräumen könnte als Risiko bei der Nutzung in Werbetafeln aufgeführt werden, dass eine Auswertung von Körpermerkmalen grundsätzlich auf Ablehnung stoßen könnte. Zum einen könnte beim Passanten ein Gefühl des "beobachtet werdens" entstehen, zum anderen gerät man - auch im Falle, dass Daten nicht dauerhaft gespeichert werden - bei der Aufzeichnung der Daten in die komplexen Gefilde des Datenschutzes.

5 Zusammenfassung und Ausblick

Die technischen Grundlagen zur Ermittlung, Kollaboration und Bereitstellung der über die Kinect-Kamera gesammelten Daten konnte erfolgreich entwickelt werden. Die Schwierigkeit ist nun die Verarbeitung und Einordnung dieser Daten, wie in Abschnitt 4.1 beschrieben. Hier muss untersucht werden, ob ein Vorgehensmodell aus der Kombination von Emotion und Gestik geschaffen werden kann und die Nutzung einen quantifizierbaren Mehrwert für den Benutzer schafft. Hierbei könnte vor allem die Verknüpfung zwischen den Gewohnheiten und Vorlieben des Benutzers, sowie den mithilfe der Kinect-Kamera gesammelten Daten zu einem erfolgsversprechenden Modell führen - "Mensch-Companion-Interaktion" (vgl. [GI-Jahrestagung, 2011](#)) . Ob das System vom Benutzer angenommen wird, gilt es es dann in einem nächsten Schritt in einer Testumgebung zu überprüfen.

Ziel der Masterarbeit ist es nun dieses Verhaltensmodell aus den gesammelten Daten zu erstellen und dessen Funktion zu validieren.

Bibliography

- [Becker-Asano 2008] BECKER-ASANO: *WASABI: Affect Simulation for Agents with Believable Interactivity*. 2008
- [Digital-Signage-Promotion-Project 2010] DIGITAL-SIGNAGE-PROMOTION-PROJECT: *Tokyo trials digital billboards that scan passers-by*. 2010. – URL <http://www.google.com/hostednews/afp/article/ALeqM5iDd1xzYx7CaahlxkLnvo4Xtcksug?docId=CNG.0b52dcd07e90b3e1c23517e10acff14c.1d1&index=0>. – Zugriffsdatum: 12.01.2013
- [Electronics 2012] ELECTRONICS, GHI: *FEZ Spider Starter Kit*. 2012. – URL <http://www.ghielectronics.com/catalog/product/297>. – Zugriffsdatum: 02.02.2012
- [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 Website*. 2011. – URL <http://www.iis.fraunhofer.de/bf/bv/ks/gpe/>. – Zugriffsdatum: 02.07.2011
- [el Kaliouby R. Picard R.W. Hoque 2009] KALIOUBY R. PICARD R.W. HOQUE, M. E. el: *When human coders (and machines) disagree on the meaning of facial affect in spontaneous videos*. 2009
- [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
- [Kriegsmann 2012] KRIEGSMANN, Edo: *Projektarbeitung 2 - 3D Interaktionen in Smart Homes*. 2012. – URL <http://users.informatik.haw-hamburg.de/~ubicomp/projekte/master11-12-proj2/kriegsmann.pdf>. – Zugriffsdatum: 12.01.2013

- [Microsoft 2013] MICROSOFT: *Microsoft Kinect SDK Website*. 2013. – URL <http://www.microsoft.com/en-us/kinectforwindows/develop/>. – Zugriffsdatum: 14.02.2013
- [Rahimi 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
- [Weiser 1991] WEISER, Mark: *The Computer for the Twenty-First Century*. 1991