



Gesten- und Sprachsteuerung für einen mobilen Roboter mittels Kinect for Windows unter Java

Studienarbeit

für die Prüfung zum

Bachelor of Science

der Angewandte Informatik

an der Dualen Hochschule Baden-Württemberg Karlsruhe

von

Simon Ebner, Volker Werling

Januar 2013

Bearbeitungszeitraum
Matrikelnummer, Kurs
Gutachter

12 Wochen
5837963 7012192, TAI10B2
Prof. Hans-Jörg Haubner

Erklärung

Wir erklären hiermit ehrenwörtlich:

1. dass wir unsere Studienarbeit mit dem Thema *Gesten- und Sprachsteuerung fuer einen mobilen Roboter mittels Kinect for Windows unter Java* ohne fremde Hilfe angefertigt haben;
2. dass wir die Übernahme wörtlicher Zitate aus der Literatur sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen innerhalb der Arbeit gekennzeichnet haben;
3. dass wir unsere Studienarbeit bei keiner anderen Prüfung vorgelegt haben;
4. dass die eingereichte elektronische Fassung exakt mit der eingereichten schriftlichen Fassung übereinstimmt.

Wir sind uns bewusst, dass eine falsche Erklärung rechtliche Folgen haben wird.

Karlsruhe, Januar 2013

Simon Ebner, Volker Werling

Zusammenfassung

Inhaltsverzeichnis

1	Einleitung	7
2	Aufgabenstellung	9
3	Stand der Technik	10
3.1	Hardware	10
3.1.1	Kinect	10
3.1.2	Roboter - Ausblick	10
3.2	Software	10
3.2.1	Microsoft Kinect SDK	10
3.2.2	Java	10
3.2.3	GUI	10
3.3	Mustererkennung	10
4	Kinect-Framework	11
4.1	Microsoft Kinect SDK	11
4.2	Java Frameworks	11
4.2.1	OpenNI	11
4.2.2	OpenKinect	12
4.2.3	jnect	12
4.2.4	Scoring-Modell	12
4.2.5	Analyseergebnis	12
5	Konzeption	13
5.1	Technische Vorgaben	13
5.1.1	Vorgaben durch Kinect	13
5.1.2	Vorgaben durch jnect-Framework	13

5.1.3	Service-Architektur	13
5.1.4	Abhängigkeit zu Robotertechnik	13
5.2	Fachliche Vorgaben	13
5.2.1	Vorgaben durch Mensch-Computer-Interaktion	13
5.2.2	Klassifikation der Anwendung	13
5.2.3	Analyse der Gesten- und Sprachsteuerung	13
6	Modelle zur Gesten- und Spracherkennung	14
6.1	Hidden-Markov-Modell	14
6.1.1	ToDo	14
6.2	Baum-Weich-Algorithmus	14
6.2.1	ToDo	14
6.3	ToDo	14
7	Gesten	15
7.1	Kreisbewegung	15
7.1.1	ToDo	15
7.2	Vorwärtsbewegung	15
7.2.1	ToDo	15
7.3	Haltesignal	15
7.3.1	ToDo	15
8	Sprachbefehle	16
8.1	ToDo	16
9	Implementierung	17
9.1	OSGI-Bundles	17
9.1.1	Gesten	17
9.1.2	Sprachbefehle	17
9.1.3	Roboterschnittstelle	17
9.1.4	Ausführbare Aktionen	17
9.2	Oberfläche	17
9.2.1	RCP-Anwendung	17
9.2.2	Grafische Darstellung	17
9.3	Ablaufbeschreibung	17

10 Ausblick - weitere Arbeiten	18
Abbildungsverzeichnis	19
Tabellenverzeichnis	20
Quellcodeverzeichnis	21
Literaturverzeichnis	22
Glossar	23

1 Einleitung

Gestik im Sinne von kommunikativen Bewegungen stellt einen wesentlichen Teil der nonverbalen Kommunikation dar. Die Definition nach Kurtenbach und Hulteen (1990) besagt:

A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key on a keyboard is neither observed nor significant. All that matters is which key was pressed.

Als solches bildet eine Geste abstrakte Strukturen und bildhafte Vorstellungen unmittelbar ab. Sie ist somit eine Körperbewegung, die Informationen enthält. Informationen, die herkömmliche Eingabegeräte, darunter Tastatur und Maus, nicht in dieser Form wiedergeben können. Ein Umstand, der sich besonders im Bereich der Mensch-Computer-Interaktion auswirkt. Das Ziel der Mensch-Computer-Interaktion ist es, die Kommunikation intuitiv und unmittelbar zu gestalten. Genau an dieser Stelle setzen Gesten an, in dem sie eine bessere Schnittstelle zwischen Mensch und Maschine darstellen.

Die Spannweite von Realisierungen von Mensch-Maschine-Schnittstelle ist groß. Frühe Ansätze zeigen bereits das Potential, Gesten zur Steuerung und Nutzung von diversen Geräten zu verwenden. Das *Theremin* - ein 1919 erfundenes elektornisches Musikinstrument, das mit Händen berührungslos durch Beeinflussung eines elektromagnetischen Feldes gesteuert wird und dadurch Töne erzeugt - ist eine der ersten technischen Lösungen für eine Gestensteuerung. Eine weitere Form, der *Datenhandschuh*, 1977 von Electronic

Visualization Laboratory Labs entwickelt ¹, erregte großes Interesse und wurde unter anderem als *Power Glove* von Mattel vertrieben. Datenhandschuhe werden bis heute in den verschiedensten Bereichen eingesetzt. Eine moderne Variante der Gestenerkennung ist die Bewegungsdetektion. Dabei wird über optische Sensoren der Körper oder einzelne Körperteile des Nutzers erkannt und somit die Steuerung durch Gesten ermöglicht. Der populärste Vertreter ist *Kinect* mit über 18 Millionen verkauften Exemplaren ². Im folgenden wird mit jener Kinect gearbeitet.

In dieser Arbeit wird mit Hilfe der Kinect eine Gesten- und Spracherkennung entwickelt um eine Steuerung für einen mobilen Roboter zu implementieren. Aufgrund des Umfangs dieser Studienarbeit wird die Erarbeitung dieser Anwendung in zwei Teile aufgeteilt. Der erste Teil der Arbeit beschäftigt sich vorrangig mit der Realisierung der Gesten- und Spracherkennung. Die eigentliche Steuerung für einen mobilen Roboter wird vorerst durch eine Schnittstelle und einer Testumgebung ersetzt. Der Fokus liegt hierbei in erster Linie auf der verwendeten Technik, in diesem Fall der Kinect zur Bewegungsdetektion, den verwendeten Gesten und Sprachbefehlen, sowie der Umsetzung der Software und den darin verwendeten Technologien. Der zweite Teil wid an diese Arbeit anknüpfen und im weiteren die Umsetzung der Steuerung für und den Einsatz eines mobilen Roboters beschreiben.

Die folgende Ausarbeitung spiegelt die Umsetzung des ersten Teils dieser Arbeit wieder.

¹Sturman, D.J., Zeltzer, D. (January 1994). „A survey of glove-based input“. IEEE Computer Graphics and Applications 14 (1): 30–39

²Takahashi, Dean (Januar 9, 2012). „Xbox 360 surpasses 66M sold and Kinect passes 18M units“. venturebeat. Abgerufen Januar 10, 2012

2 Aufgabenstellung

3 Stand der Technik

3.1 Hardware

3.1.1 Kinect

3.1.2 Roboter - Ausblick

3.2 Software

3.2.1 Microsoft Kinect SDK

3.2.2 Java

3.2.3 GUI

3.3 Mustererkennung

4 Kinect–Framework

4.1 Microsoft Kinect SDK

asdf

4.2 Java Frameworks

adsf

4.2.1 OpenNI

Beschreibung

adsf

Technische Daten

SWOT–Analyse

4.2.2 OpenKinect

Beschreibung

Technische Daten

SWOT–Analyse

4.2.3 jnect

Beschreibung

Technische Daten

SWOT–Analyse

4.2.4 Scoring–Modell

4.2.5 Analyseergebnis

adsasfd

5 Konzeption

5.1 Technische Vorgaben

5.1.1 Vorgaben durch Kinect

5.1.2 Vorgaben durch jnect-Framework

5.1.3 Service-Architektur

5.1.4 Abhängigkeit zu Robotertechnik

5.2 Fachliche Vorgaben

5.2.1 Vorgaben durch Mensch-Computer-Interaktion

5.2.2 Klassifikation der Anwendung

5.2.3 Analyse der Gesten- und Sprachsteuerung

6 Modelle zur Gesten- und Spracherkennung

6.1 Hidden-Markov-Modell

6.1.1 ToDo

6.2 Baum-Weich-Algorithmus

6.2.1 ToDo

6.3 ToDo

7 Gesten

7.1 Kreisbewegung

7.1.1 ToDo

7.2 Vorwärtsbewegung

7.2.1 ToDo

7.3 Haltesignal

7.3.1 ToDo

8 Sprachbefehle

8.1 ToDo

9 Implementierung

9.1 OSGI–Bundles

9.1.1 Gesten

9.1.2 Sprachbefehle

9.1.3 Roboterschnittstelle

9.1.4 Ausführbare Aktionen

9.2 Oberfläche

9.2.1 RCP–Anwendung

9.2.2 Grafische Darstellung

9.3 Ablaufbeschreibung

10 Ausblick - weitere Arbeiten

Abbildungsverzeichnis

Tabellenverzeichnis

Quellcodeverzeichnis

Literaturverzeichnis

Glossar

Datenhandschuh

Der Datenhandschuh ist ein Eingabegerät in Form eines Handschuhs. Durch Bewegungen der Hand und Finger erfolgt eine Orientierung im virtuellen Raum. 6

Mensch-Computer-Interaktion

Die Mensch-Computer-Interaktion (englisch *Human-Computer Interaction, HCI*) als Teilgebiet der Informatik beschäftigt sich mit der benutzergerechten Gestaltung von interaktiven Systemen und ihren Mensch-Maschine-Schnittstellen. 5, 6, 11

Mensch-Maschine-Schnittstelle

Die Benutzerschnittstelle (nach Gesellschaft für Informatik, Fachbereich Mensch-Computer-Interaktion auch Benutzungsschnittstelle) ist die Stelle oder Handlung, mit der ein Mensch mit einer Maschine in Kontakt tritt. 6