



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	10
3	Stand der Technik	11
3.1	Kinect	11
3.1.1	Hardware	12
3.1.2	Software	15
3.2	Roboter - Ausblick	16
3.3	Java	16
3.3.1	GUI	16
3.4	Mustererkennung	16
4	Kinect-Framework	17
4.1	Microsoft Kinect SDK	17
4.2	Java Frameworks	17
4.2.1	OpenNI	17
4.2.2	OpenKinect	18
4.2.3	jnect	18
4.2.4	Scoring-Modell	18
4.2.5	Analyseergebnis	18
5	Konzeption	19
5.1	Technische Vorgaben	19
5.1.1	Vorgaben durch Kinect	19
5.1.2	Vorgaben durch jnect-Framework	19
5.1.3	Service-Architektur	19

5.1.4	Abhängigkeit zu Robotertechnik	19
5.2	Fachliche Vorgaben	19
5.2.1	Vorgaben durch Mensch-Computer-Interaktion	19
5.2.2	Klassifikation der Anwendung	19
5.2.3	Analyse der Gesten- und Sprachsteuerung	19
6	Modelle zur Gesten- und Spracherkennung	20
6.1	Hidden-Markov-Modell	20
6.1.1	ToDo	20
6.2	Baum-Weich-Algorithmus	20
6.2.1	ToDo	20
6.3	ToDo	20
7	Gesten	21
7.1	Kreisbewegung	21
7.1.1	ToDo	21
7.2	Vorwärtsbewegung	21
7.2.1	ToDo	21
7.3	Haltesignal	21
7.3.1	ToDo	21
8	Sprachbefehle	22
8.1	ToDo	22
9	Implementierung	23
9.1	OSGI-Bundles	23
9.1.1	Gesten	23
9.1.2	Sprachbefehle	23
9.1.3	Roboterschnittstelle	23
9.1.4	Ausführbare Aktionen	23
9.2	Oberfläche	23
9.2.1	RCP-Anwendung	23
9.2.2	Grafische Darstellung	23
9.3	Ablaufbeschreibung	23

10 Ausblick - weitere Arbeiten	24
Abbildungsverzeichnis	25
Tabellenverzeichnis	26
Quellcodeverzeichnis	27
Literaturverzeichnis	28
Glossar	29

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 weitere 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. Eine frühe Lösung ist der sogenannte *Videoplace* ², entwickelt von Myron Krueger Mitte der 70'er Jahre ³. Ausgelegt als ein *Labor für künstliche Realität* war eine Person, durch ihn umgebende Projektoren und Videokameras, in der Lage seine Umgebung zu beeinflussen. Moderne Lösungen haben das Konzept der Detektion durch Kameratechnik weiterentwickelt und einer weiten Öffentlichkeit zugänglich gemacht. Darunter auch das *PLAYSTATIONEye* ⁴. Der populärste Vertreter ist *Kinect for XBOX* mit über 18 Millionen verkauften Exemplaren ⁵. 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.

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

²Beschreibung der Einrichtung „VIDEOPLACE“. medienkunstnetz.de. Abgerufen Dezember 22, 2012

³Myron Krueger. Artificial Reality 2, Addison-Wesley Professional, 1991. ISBN 0-201-52260-8

⁴„PLAYSTATIONEye Brings Next-Generation Communication to PLAYSTATION3“. us.playstation.com. Sony Computer Entertainment America. Abgerufen Dezember 21, 2012

⁵Takahashi, Dean (Januar 9, 2012). „Xbox 360 surpasses 66M sold and Kinect passes 18M units“. venturebeat. Abgerufen Dezember 20, 2012

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

2 Aufgabenstellung

3 Stand der Technik

3.1 Kinect

Kinect ist eine Gerät zur Dedektion von Bewegungen. Entwickelt wurde es von PrimeSense als Hardware zur Steuerung der Videokonsole XBOX 360 von Microsoft Corp. Nach der Ankündigung im März 2010 ¹ war die Erweiterung mit dem Erscheinungsdatum 10. November 2010 in der Ausführung *Kinect for XBOX 360* in Europa erhältlich ². Nach einem sehr erfolgreichem Verkaufsstart ³ und anhaltender Nachfrage⁴. kündigte Microsoft am 9. Januar 2012 ein weiteres Modell der Kinect, die sogenannte *Kinect for Windows* für den 1. Februar 2012 an ⁵.

Darüber hinaus veröffentlichte Microsoft bereits am 16. Juni 2011 eine erste Version ihrer *Kinect for Windows SDK*⁶. Mit diesem Software Development Kit ist es möglich auf einer Windows 7 Plattform Anwendungen zu entwickeln, die eine Kinect als Eingabegerät verwenden.

Mit der Verfügbarkeit einer Kinect-Variante, die für den Einsatz am PC ausgelegt ist und der frei zugänglichen einer breiten Öffentlichkeit als Forschungsgegenschaft zugänglich,

¹Pressemitteilung, der Veröffentlichung. Siehe Link. Microsoft.com. Abgerufen Dezember 20, 2012

²Erscheinungsdatum der *Kinect for XBOX 360*. Siehe Link. BBC UK. Abgerufen Dezember 20, 2012

³„Am schnellsten verkauftes Peripheriegerät für Spiele“. Guinnessworldrecords.com. Abgerufen Dezember 22, 2012

⁴Takahashi, Dean (Januar 9, 2012). „Xbox 360 surpasses 66M sold and Kinect passes 18M units“. venturebeat. Abgerufen Dezember 20, 2012

⁵„Ankündigung der *Kinect for Windows*“. blogs.msdn.com. Abgerufen Dezember 22, 2012

⁶„Microsoft Releases Kinect for Windows SDK Beta for Academics and Enthusiasts“. Microsoft.com. Abgerufen Dezember 21, 2012

was auch der ausschlaggebende Punkt für den Einsatz der Kinect in dieser Studienarbeit ist.

3.1.1 Hardware

Die beiden Kinect-Modelle unterscheiden sich in einigen Details⁷. Der bedeutendste Faktor ist das sogenannte *Near Mode* Feature:

Near Mode enables the depth sensor to see objects as close as 40 centimeters and also communicates more information about depth values outside the range than was previously available. There is also improved synchronization between color and depth, mapping depth to color, and a full frame API.⁶

Neben dem aktualisierten Tiefensensor unterscheiden sich die beiden Varianten auch in einer höheren Auflösung der RGB-Kamera, die für eine Gestenerkennung relevant sein kann. Aus diesem Grund wird für diese Arbeit die *Kinect for Windows* genutzt.

Technische Daten

Eine Kinect enthält innerhalb des Gehäuses, folgende Sensoren¹:

- Eine RGB-Kamera mit einer Auflösung von 1280x960. Dies ermöglicht eine Farbbilderfassung
- Ein Infrarot (IR) Emitter und ein IR Tiefensensor. Der Emitter emittiert infrarote Lichtstrahlen und der Tiefensensor erfasst die an den Sensor reflektierten Strahlen. Die reflektierten Strahlen werden in Tiefeninformation umgewandelt, in dem der

⁷„Informationsseite über Unterschiede der Kinect Versionen“. Microsoft.com. Abgerufen Dezember 22, 2012

¹„Kinect for Windows Sensor Components and Specifications“ msdn.microsoft.com. Abgerufen Dezember 21, 2012

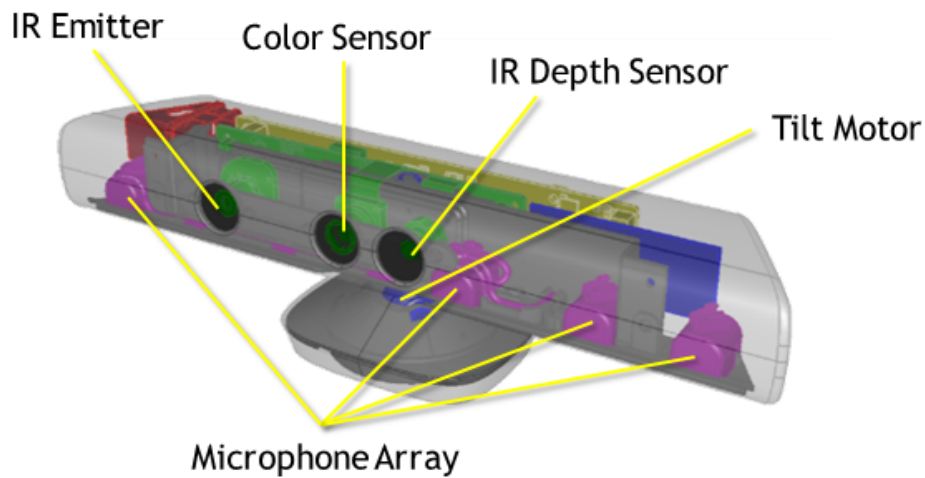


Abbildung 3.1: Schematische Ansicht der Sensoren einer *Kinect for Windows* (Quelle: Microsoft Corp.¹⁾)

Abstand zwischen Objekt und Sensor bestimmt wird. Dies ermöglicht die Erfassung von Tiefenbildern

- Ein Mikrofonarray, das vier Mikrofone zur Soundaufnahme enthält. Die Anzahl der Mikrofone ermöglicht nicht nur die Aufzeichnung von Audiodaten, sondern auch die Lokalisierung der Soundquelle und die Richtung des Audiosignals
- Ein 3-Achsen-Beschleunigungssensor, konfiguriert für einen Bereich der zweifachen Erdbeschleunigung, um die gegenwärtige Ausrichtung der Kinect zu bestimmen
- Ein Kippmotor, zur automatisierten Justierung der Sensoren

Weitere Details der technischen Spezifikation einer *Kinect for Windows* sind in Tabelle 3.1 aufgelistet:

Tabelle 3.1: Kinect - technische Spezifikation ¹

Kinect	Spezifikation
Blickwinkel	43° vertikales, 57° horizontales Blickfeld
Vertikaler Neigebereich	±27°
Bildwiederholrate (Farb und Tiefensignal)	30 Bilder pro Sekunde (FPS)
Audioformat	16-kHz, 24-bit mono pulse code modulation (PCM)
Audioeingang	Ein Vier-Mikrofonarray mit 24-Bit Analog-Digital-Wandler (ADC)
Datensignal-Tiefensensor	640x480 16-bit, 30 Bilder pro Sekunde
Datensignal-RGB-Kamera	1280x960 16-Bit, 12 Bilder pro Sekunde 640x480 16-Bit, 30 Bilder pro Sekunde
Tiefensensorreichweite	0,4 – 4 m
<i>Skelett Tracking System</i>	Erkennung von bis zu sechs Benutzern, zwei davon trackbar/verfolgbar Verfolgung von 20 Gelenken pro aktivem Nutzer

¹„Kinect for Windows Sensor Components and Specifications“ msdn.microsoft.com. Abgerufen Dezember 21, 2012

3.1.2 Software

Durch die Veröffentlichung eines Software Development Kits ist es möglich, die Kinect in eigene Programme einzubinden und neue Anwendungsfälle zu bearbeiten.

Kinect for Windows SDK

Das *Kinect for Windows SDK* steht aktuell in der Version 1.6 ⁸ bereit. Dabei kann direkt in den Programmiersprachen C++, C# , und Visual Basic auf einer Windows 7 Plattform entwickelt werden. al;kdjfa;lkdjf;lekjf;lakdjf;alskjf

Weitere Frameworks

Da Microsoft die Nutzungsmöglichkeiten seines SDK hinsichtlich verwendeter Programmiersprache und Plattform einschränkt, begannen Forscher eigene Frameworks und Treiber zur Nutzung der Kinect zu entwickeln⁹. PrimeSense selbst veröffentlichte Treiber und Middleware für die Kinect ¹⁰. Ein Teil dieser Ausarbeitung ist es, ein für die Aufgabenstellung und Zielsetzung der Studienarbeit passendes Framework zu bestimmen.

⁸Downloadseite des SDK. Microsoft.com. Abgerufen Dezember 21, 2012

⁹„Open Source Kinect contest has been won“. hackaday.com. November 11, 2010. Abgerufen Dezember 21, 2012

¹⁰Mitchell, Richard (Dezember 10, 2010). „PrimeSense releases open source drivers, middleware for Kinect“. Joystiq. Abgerufen Dezember 22, 2012

3.2 Roboter - Ausblick

3.3 Java

3.3.1 GUI

3.4 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

3.1	Caption for LOFd	13
-----	----------------------------	----

Tabellenverzeichnis

3.1 Caption for LOF 14

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