

Technische Universität Berlin

Fakultät IV - Elektrotechnik und Informatik

Fachgebiet AOT

Prof. Dr. Sahin Albayrak

Masterarbeitsproposal

Hand Gesture Recognition for Human-Robot  
Interaction

Sivalingam Panchadcharam Aravinth

Matrikel-Nummer 342899

**Betreuer** Dr. Yuan Xu

# Inhaltsverzeichnis

<b>Index</b>	<b>II</b>
<b>1 Abstract</b>	<b>1</b>
<b>2 Motivation</b>	<b>2</b>
<b>3 Background</b>	<b>3</b>
3.1 Robot and Artificial intelligence . . . . .	3
3.1.1 NAO Vision . . . . .	3
3.1.2 Extending NAO with 3D Camera . . . . .	3
3.2 Computer Vision . . . . .	4
3.3 Hand Gesture . . . . .	5
3.4 Gesture Recognition . . . . .	5
3.5 title . . . . .	5
<b>4 Goal</b>	<b>6</b>
<b>5 Taskset</b>	<b>7</b>
<b>6 Zeitplan</b>	<b>8</b>
<b>7 Organisatorisches</b>	<b>9</b>
<b>8 Anhang</b>	<b>10</b>

# 1 Abstract

Human-robot interaction (HRI) has been a topic of both science fiction and academic speculation even before any robots existed. HRI research is focusing to build an intuitive, and easy communication with the robot through speech, gestures, and facial expressions. The use of hand gestures provides an attractive alternative to complex interfaced devices for HRI. In particular, visual interpretation of hand gestures can help in achieving the ease and naturalness desired for HRI. This has motivated a very active research area concerned with computer vision-based analysis and interpretation of hand gestures. Important differences in the gesture interpretation approaches arise depending on whether static model of the gesture or non-static model of the gesture is used.

In this thesis, we attempt to do the method of modeling, analyzing, and recognizing gestures by using Computer Vision and Machine Learning techniques. Furthermore, Static (Gesture is formed by non-moving appearance of body parts) and non-static gestures (Gesture is formed by moving appearance of body parts), will be used to interact with robot and command the robot to execute certain actions.

We further hope to provide a platform to integrate Sign Language Translation to assist people with hearing and speech disabilities. However, further implementations and training data are needed to use this platform as a full fledged Sign Language Translator.

## 2 Motivation

Im Bereich ... (Warum muss es eine neue Lösung/ einen neuen Ansatz geben)

Huge influence of computers in society has made smart devices, an important part of our lives. Availability and affordability of such devices motivated us to use them in our day-to-day living.

Interaction with smart devices has still been mostly by displays, keyboards, mouse and touch interfaces. These devices have grown to be familiar but inherently limit the speed and naturalness with which we can interact with the computer.

The list of smart devices includes personal automatic and semi-automatic robots which are also playing a major role in our household. For an instance, Smart Vacuum Cleaner is an robot that automatically cleans the floor and goes to its charging station without human interaction. Usage of robots for domestic and industrial purposes have been continuously increasing. Thus in recent years there has been a tremendous push in research toward an intuitive, and easy communication with the robot through speech, gestures, and facial expressions.

Tremendous progress had been made in speech recognition, and several commercially successful speech interfaces have been deployed. However, there has been an increased interest in recent years in trying to introduce other human-to-human communication modalities into HRI. This includes a class of techniques based on the movement of the human arm and hand, or hand gestures. The use of hand gestures provides an alternative mode of communication for Human-robot interaction (HRI) than the conventional interfaced devices.

## 3 Background

### 3.1 Robot and Artificial intelligence

Proper vision is of the utmost importance for the function of any vision based autonomous robot. Areas of artificial intelligence deal with autonomous planning or deliberation for robotical systems to navigate through an environment. A detailed understanding of these environments is required to navigate through them. Information about the environment could be provided by a computer vision system, acting as a vision sensor and providing high-level information about the environment and the robot.

In this thesis, we will focus on the hand gesture recognition using computer vision techniques for a humanoid robot name as NAO.

Nao is an autonomous, programmable humanoid robot developed by Aldebaran Robotics. The Nao Academics Edition was developed for universities and laboratories for research and education purposes.

#### 3.1.1 NAO Vision

Two identical video cameras are located in the forehead. They provide a up to 1280x960 resolution at 30 frames per second. They can be used to identify objects in the visual field such as goals and balls, and bottom camera can ease NAO's dribbles. NAO contains a set of algorithms for detecting and recognizing faces and shapes. NAO can recognize who is talking to it or find a ball or, eventually, more complex objects.

#### 3.1.2 Extending NAO with Depth Camera

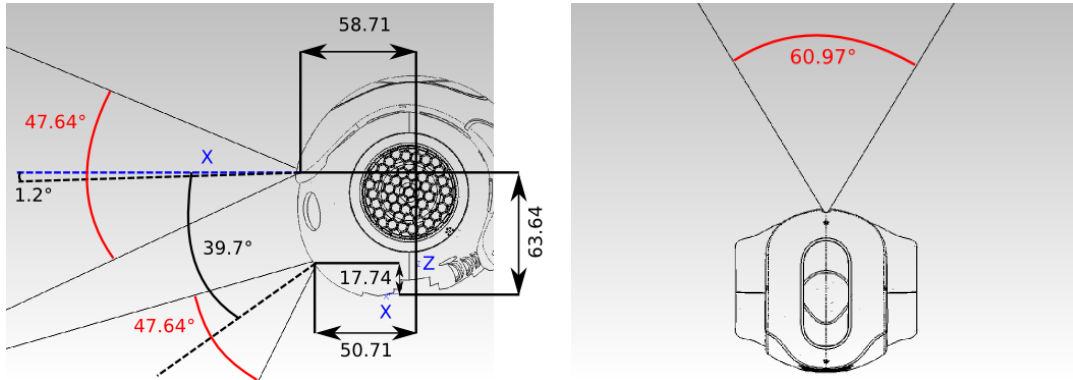
The computational limitations of the NAO robot make it impossible to process the entire image in real time. Therefore, different methods of capturing the real time gesture must be used. Hence, it is been proposed to use a separate 3D camera such

### 3 Background

Tabelle 3.1: Nao's hardware specification

Height	58 centimetres (23 in)
Weight	4.3 kilograms (9.5 lb)
Autonomy	60 minutes (active use), 90 minutes (normal use)
Degrees of freedom	21 to 25
CPU	2 x Intel Atom @ 1.6 GHz
Built-in OS	Linux
Compatible OS	Windows, Mac OS, Linux
Programming languages	C++, Python, Java, MATLAB, Urbi, C, .Net
Vision	2 x HD 1280x960 cameras
Connectivity	Ethernet, Wi-Fi
Sensors	4 x directional microphones 1 x sonar rangefinder 2 x IR emitters and receivers 1 x inertial board 9 x tactile sensors 8 x pressure sensors

as Microsoft Kinect or Asus Xtion. Microsoft's Kinect sensor is the first large-scale, commercial release of a depth camera device. Microsoft's novel use of the device as a game controller is driven by the company's own software for analyzing its 3D data, including proprietary algorithms for feature selection, scene analysis, motion tracking, skeletal tracking and gesture recognition.



## 3.2 Computer Vision

Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world

in order to produce numerical or symbolic information, e.g., in the forms of decisions.

## 3.3 Hand Gesture

Human hand gestures are a means of non-verbal interaction among people. They range from simple actions of using our hand to point at to the more complex ones that express our feelings and allow us to communicate with others. To exploit the use of gestures in HRI it is necessary to provide the means by which they can be interpreted by robots. The HCI interpretation of gestures requires that dynamic and/or static configurations of the human hand, arm, and even other parts of the human body, be measurable by the machine.

## 3.4 Gesture Recognition

Initial attempts to recognize hand gestures, resulted in electro-mechanical devices that directly measure hand and/or arm joint angles and spatial position using sensors. Glove-based gestural interfaces require the user to wear such a complex device that hinders the ease and naturalness with which the user can interact with the computer controlled environment.

Even though such hand gloves are used in highly specialized domain such as simulation of medical surgery or even the real surgery, the everyday user will be certainly deterred by such a sophisticated interfacing devices. As an active result of the motivated research in HCI, computer vision based techniques were innovated to augment the naturalness of interaction.

## 3.5 title

## 4 Goal

Im Rahmen ... (Was will ich ueberhaupt mit meiner Abreit erreichen? Etwas verbessern, entwickeln, vergleichen...) As described earlier, HRI research is focusing to build an intuitive, and easy communication with the robot through speech, gestures, and facial expressions. The use of hand gestures provides the ease and naturalness with which the user can interact with robots.

In this thesis, we attempt to implement a feature for NAO to recognize hand gestures and execute a predefined action based on the gesture.

In order to achieve this, NAO will be extended with an external Depth Camera, that will enable NAO to recognize static as well as dynamic gestures. This 3D camera will be mounted on the head of NAO, so that it can scan for gestures in the horizon.

With the hand gesture recognizing feature, NAO will be available to the users in two modes a) Command mode b) Translation mode.

In command mode, a gesture will be recognized by NAO and related task will be executed. Even though the gesture based Interaction is real time, NAO can not be interrupted or stopped by using any gesture while it is executing a task. However, part of NAO intelligence such as voice commands can be used in such situation to stop or interrupt the ongoing task execution.

In translation mode, NAO will be directly translating the meaning of the trained gestures. To achieve this, text-to-speech SDK of NAO will be used and translated gesture can be said out loud using the integrated speaker. This feature of NAO will be allowing it to translate a sign language to assist people with hearing and speech disabilities.

In this thesis, we planned to train NAO with few very simple gestures due to the reason that NAO has limited resources. Gestures will involve both the hands or single hand to interact with the robot.



# 5 Taskset

Ausgehend von der Zielbeschreibung werden folgende Arbeitspakete definiert:

- Einarbeitung in das Themenbereich. Dazu gehört die Evaluierung relevanter Paper sowie das Berücksichtigen der bereits verfügbaren Komponenten und Implementierungen. Das Arbeitsergebnis ist eine Zusammenfassung relevanter Arbeiten.
- Entwicklung eines Ansatzes ...
- Analyse, Design und Entwurf einer Architektur
- Definition von geeigneten Testszenarien
- Umsetzung der Architektur, Implementierung
- Dokumentation der Architektur und des Programmcodes

## 6 Zeitplan

Die Bearbeitung dauert maximal 6 Monate. (Am besten eignet sich ein Gannt Diagramm)

# 7 Organisatorisches

- Sprache der Diplomarbeit:
- Textverarbeitungssystem:
- Programmiersprache: Java 1.5
- Betreuer:
- Gutachter: Prof. Dr. Sahin Albayrak, Dr.- Ing. Stefan Fricke

## 8 Anhang