Gesture Recognition for Human-Robot Interaction: An approach based on skeletal points tracking using depth camera

Aravinth, Sivalingam Panchadcharam <me@aravinth.info>

Supervisor:
Dr. Yuan Xu

Technische Universität Berlin

Outline

- Introduction
- Goal
- Background
- Design
- Implementation
- Results
- Conclusion

Introduction

- Human-Robot Interaction
  - Communication between people and machines
  - Conventional interfaces are display, keyboard, mouse
- Natural Interaction
  - Human-to-human communication Modalities
  - Speech, Gestures, Facial Expressions
- Hand Gestures
  - Non-verbal communication using hands / arm
- Hand gesture recognition by modeling, training, classifying and recognizing based on computer vision and machine learning techniques.

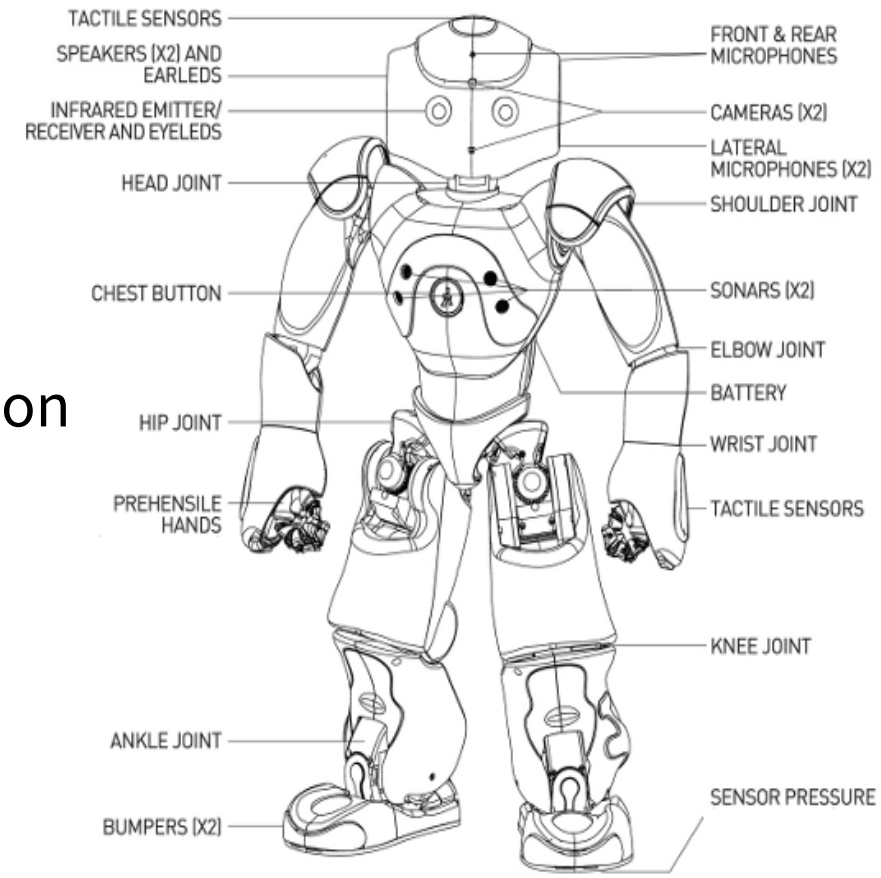
Goal

- Integrate depth camera into NAO
- Hand gesture recognition in real time, based on skeletal points tracking using depth image
  - Feature extraction, training, classification and prediction
- Human-robot interactions by
  - Gesture-To-Speech Translation
  - Gesture-To-Motion Translation
  - Gesture-To-Gesture Translation
- Graphical User Interface to visualize the interactions
- Test and evaluation of results
- Documentation

Background - Robot

- NAO
  - Humanoid Robot from Aldebaran Robotics
  - 25 Degrees of Freedom
  - Intel Atom @ 1.6 GHz
  - 1GB RAM
  - 32-bit Gentoo Linux
  - Real-time OS patched
  - NAOqi SDK in C++, Python

***Source**: Aldebaran Robotics*



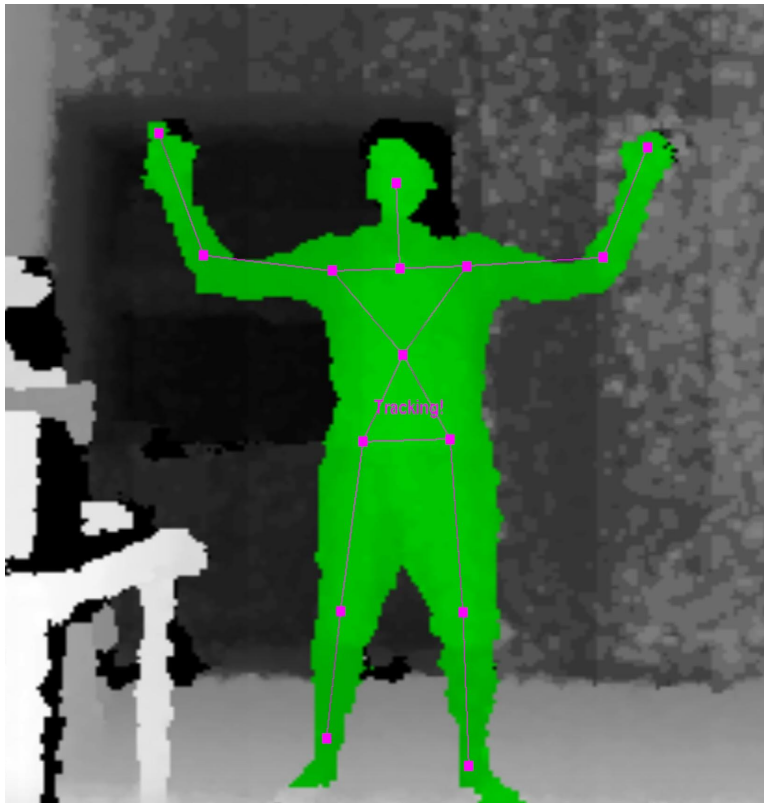
Background - Depth Camera

- Asus Xtion PRO LIVE
  - Infrared depth camera
  - 30 frames per second
  - RGB video
  - VGA (640x480): 30 fps
  - QVGA (320x240): 60 fps
  - OpenNI compatible
  - Light weight
  - USB powered

***Source**: Asus Inc.*

Background - OpenNI 2 & NiTE 2

- OpenNI 2 - Open Natural Interaction
  - Primesense driver for depth camera
- NiTE 2 - Natural Interaction Technology for End-user
  - OpenNI middleware
  - Human skeleton tracking
  - Hand tracking
  - Gestures detection
  - C++ Library



Background - GRT

- Gesture Recognition Toolkit (GRT)
  - Open source C++ library from MIT Media Lab
  - Machine Learning toolkit for real time gesture recognition
  - Classification and regression algorithms for static and temporal gestures
  - Flexible Object Oriented Gesture recognition pipeline with preprocessing, feature extraction, classification, post-processing modules
  - Classification Algorithms ANBC, SVM, MinDist, HMM, KNN, DTW

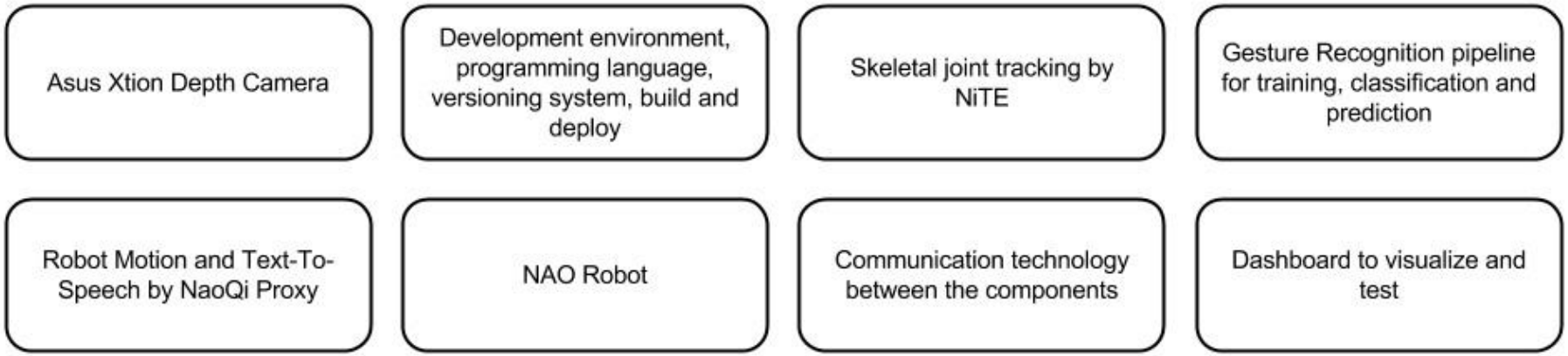
Background - ANBC

- Adaptive Naive Bayes Classifier (ANBC)
  - Based on Bayes Theory
  - N-dimensional input classification for basic and complex static gestures recognition
  - Gaussian distribution on input stream for real time prediction
  - Null rejection region threshold for non-gestures
  - Faster learning and prediction algorithm
  - Online training

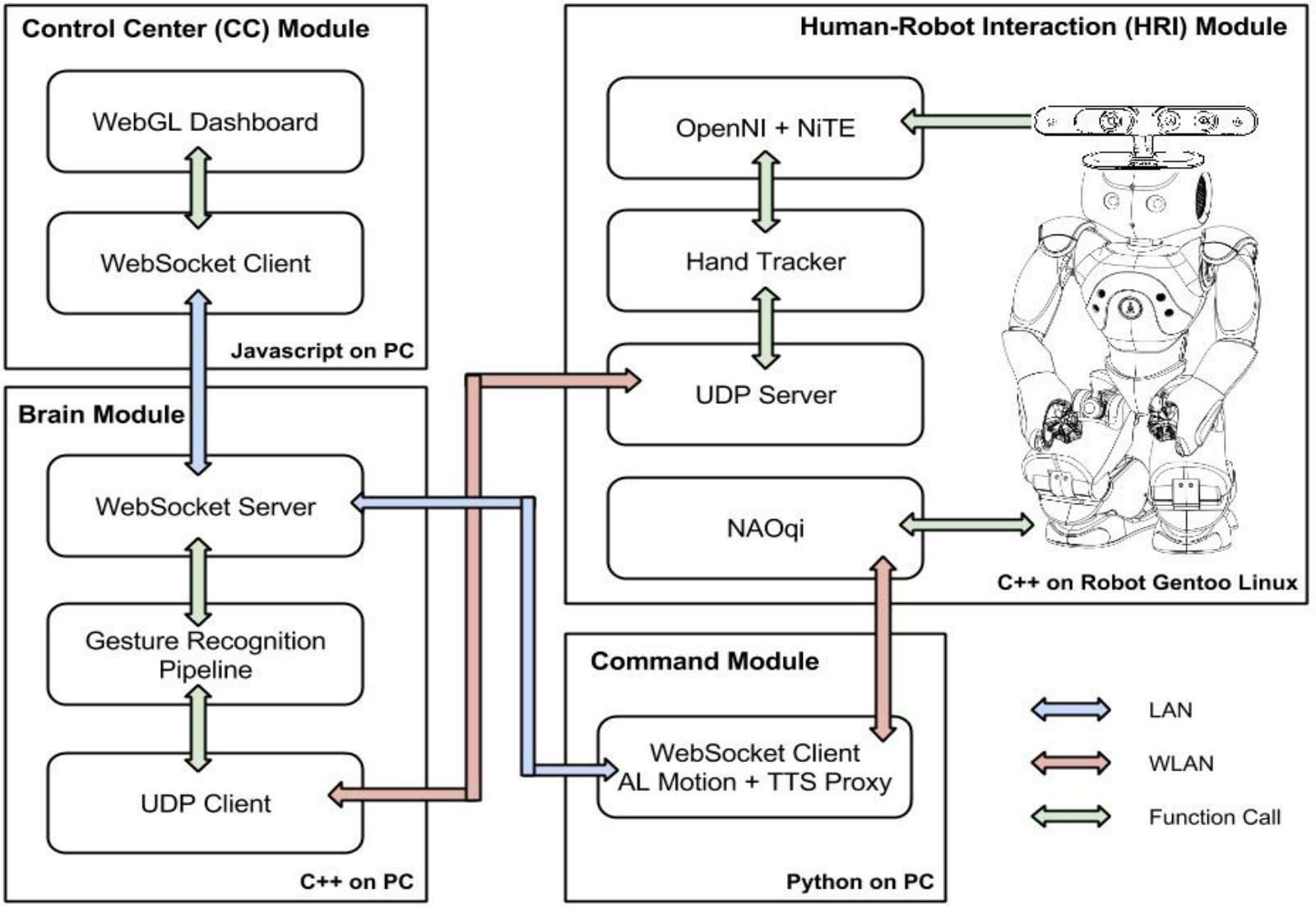
$$P(cl_i|g) = \frac{P(g_k|cl_i).P(cl_i)}{P(g_k)}$$

Design - Essential components

- Experimental designs
  - Everything On-Board
  - Extending NAO with Single Board Computer
  - Everything Off-Board
- The challenge is to find a solution that can integrate all these components into a robust system.

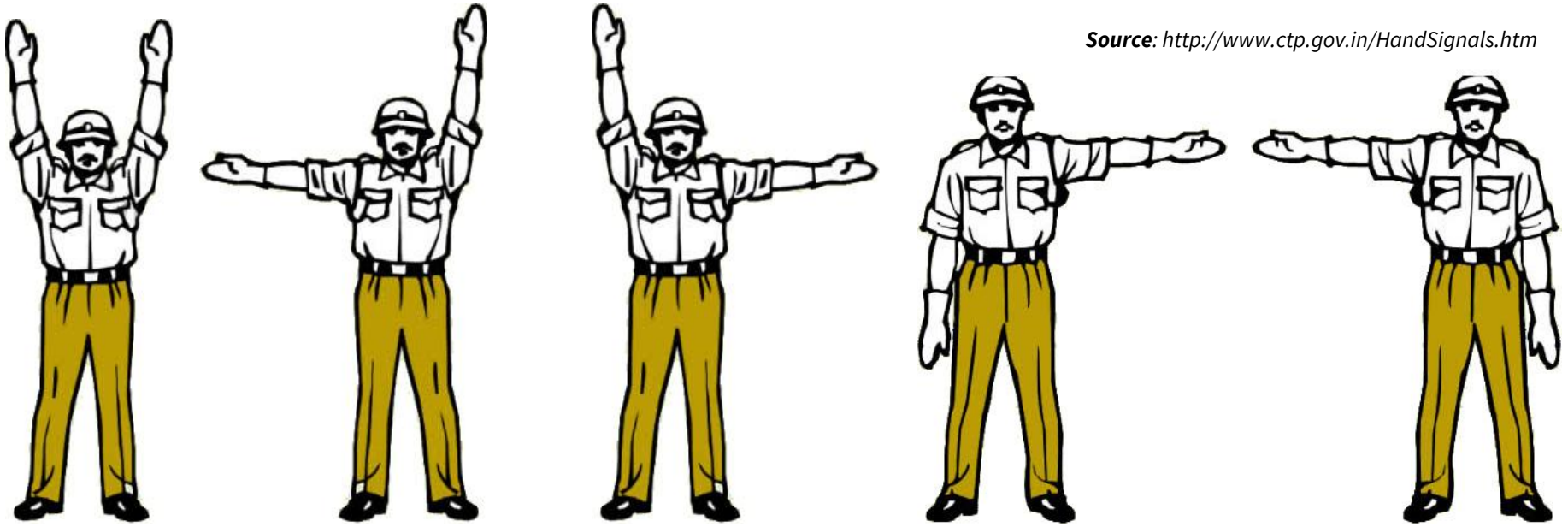


Design - Architecture



Implementation - Gesture Modelling

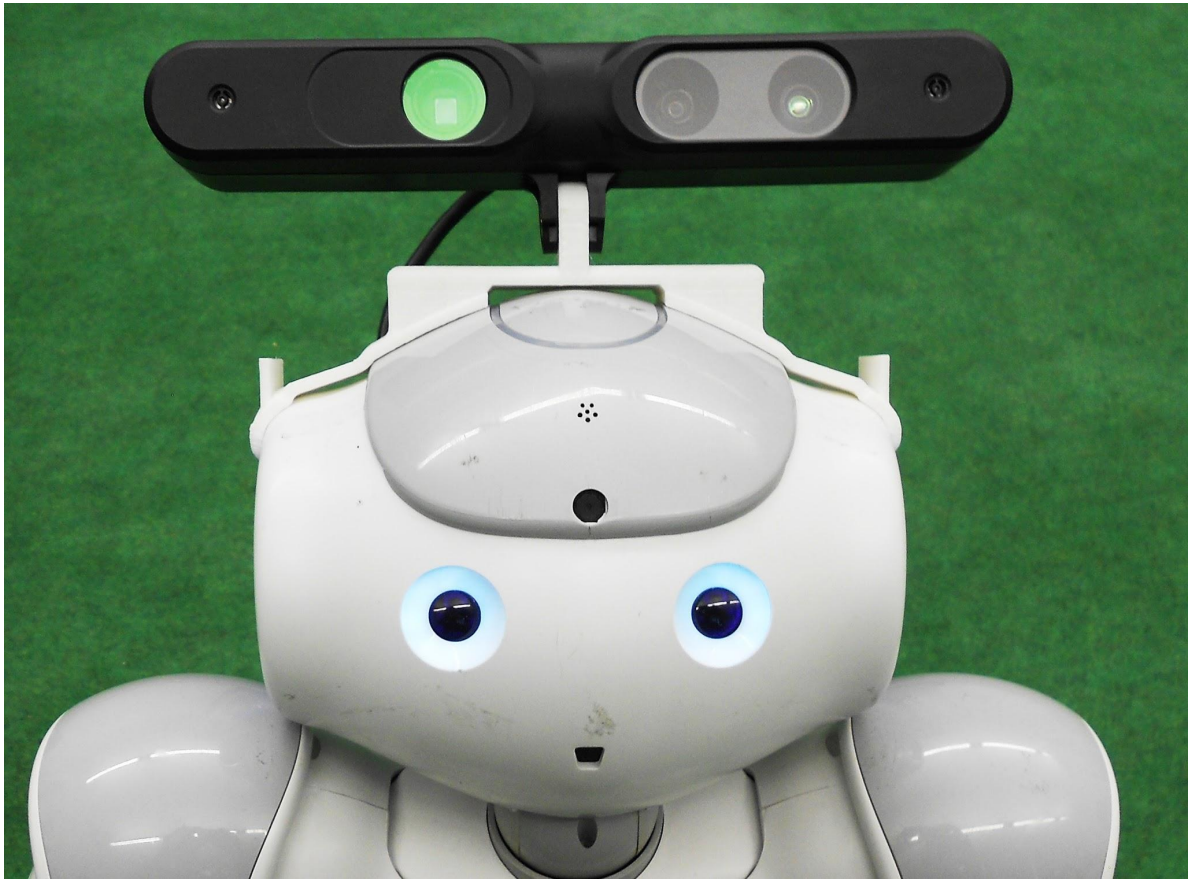
- Gestures based on skeletal points of left and right hand
- Five static gestures modelled based on traffic police hand signals - Walk, Turn Right, Turn Left, Move Right, Move Left



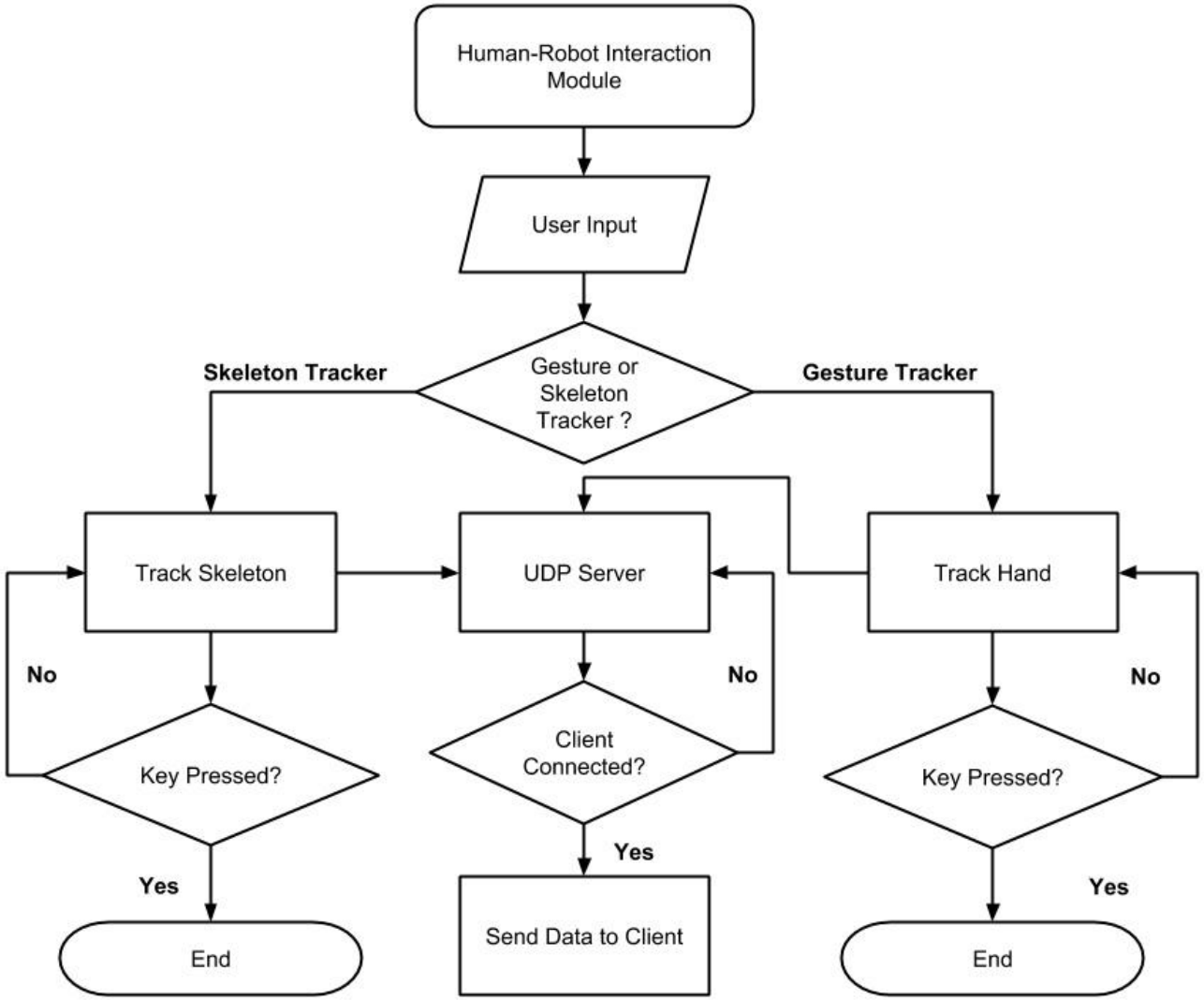
***Source**: http://www.ctp.gov.in/HandSignals.htm*

Implementation - NAO Depth Camera Mount

- 3D printed head mount for NAO to hold Asus Xtion



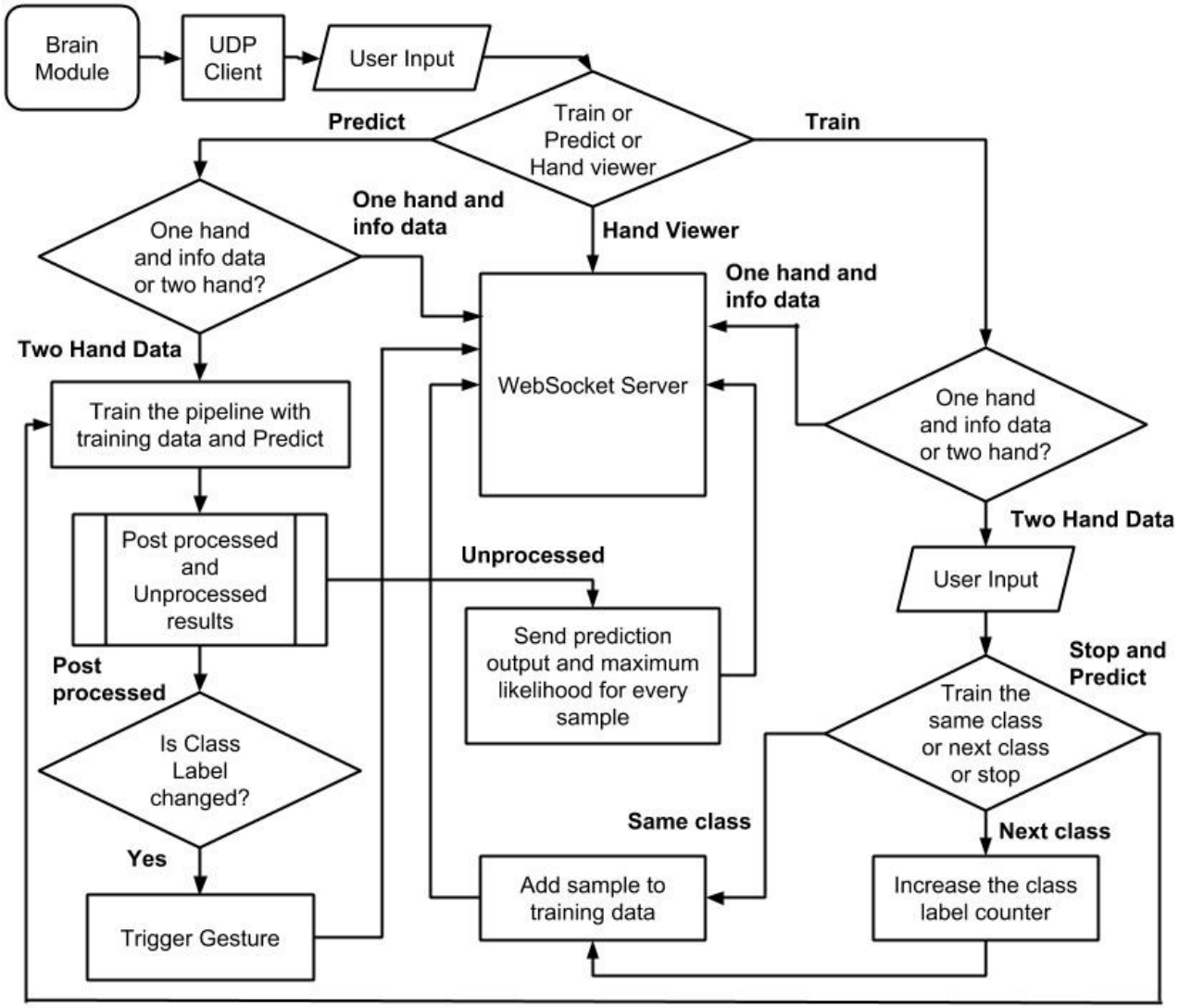
Implementation - HRI Module Flowchart



Implementation - HRI Module

- Human-Robot Interaction module
  - Accesses the depth camera via OpenNI 2
  - Starts skeleton / hand tracking using NiTE 2
  - Starts UDP Server to stream tracked joints to Brain module
  - Starts skeleton tracking using “Hands Raise” pose and hand tracking using “WAVE” focus gesture
  - When hand reaches the edge of field of view or hand is lost, informs the Brain module
  - Developed in C++ using Xcode on Mac OSX
  - Built using Clang for Mac OSX and Cmake GCC for 32-bit and 64-bit Linux
  - Uses Boost libraries such as Boost.Asio, Log, Thread

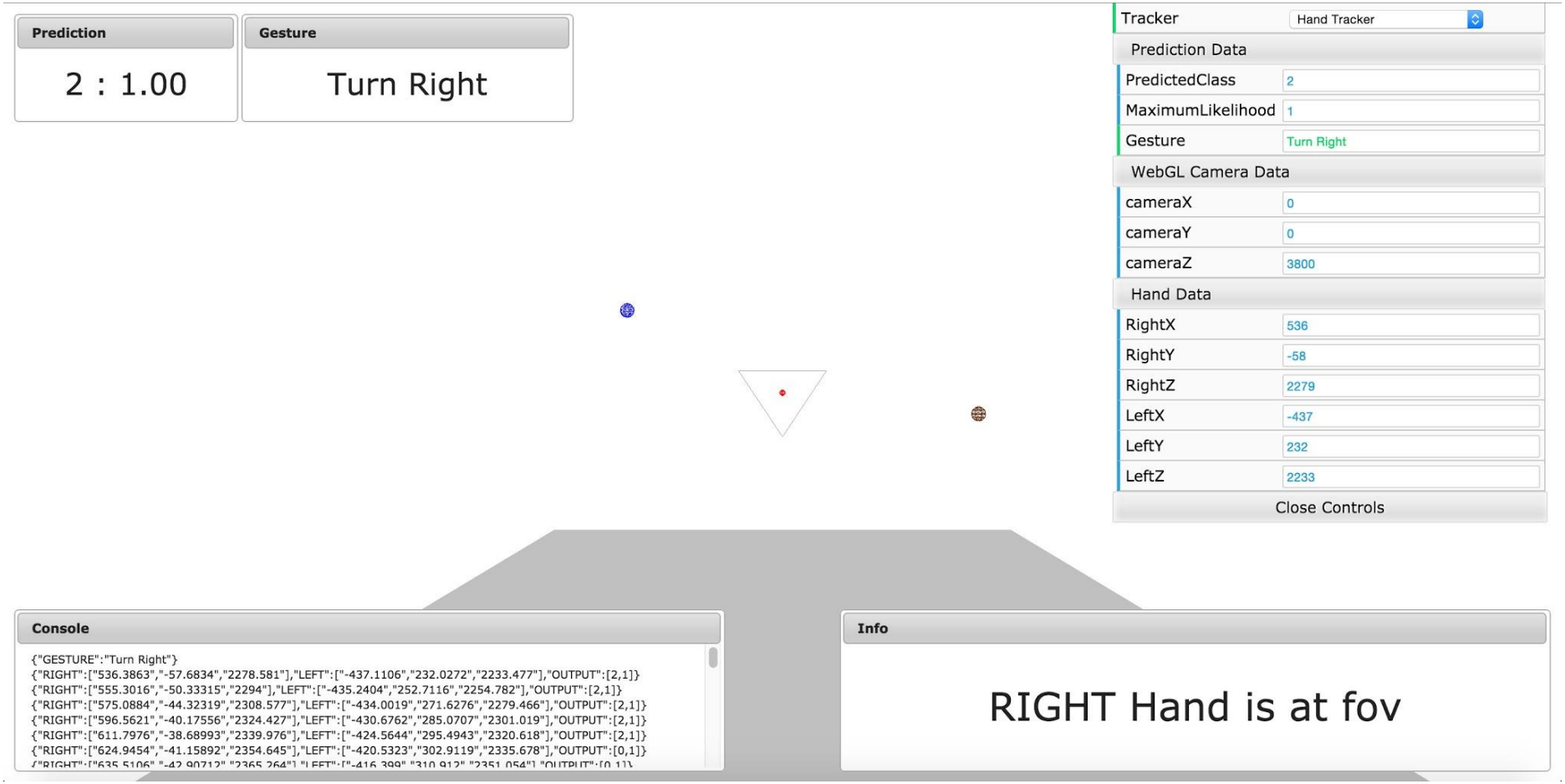
Implementation - Brain Module



Implementation - Brain Module

- Brain module
  - Starts UDP client to connect to HRI module
  - Starts WebSocket server to broadcast the results to Control Center and Command modules
  - Accepts 3 dimensional vector of hand
  - In training mode, stores input samples into training dataset for each class
  - In prediction mode, trains the classifier with the training data and performs real time prediction on the stream of input samples of left and right hand
  - Post-processes the prediction results and triggers output, when the gesture is gesticulated for more than one second
  - Developed in C++ using GRT, Boost, websocketpp

Implementation - CC Module



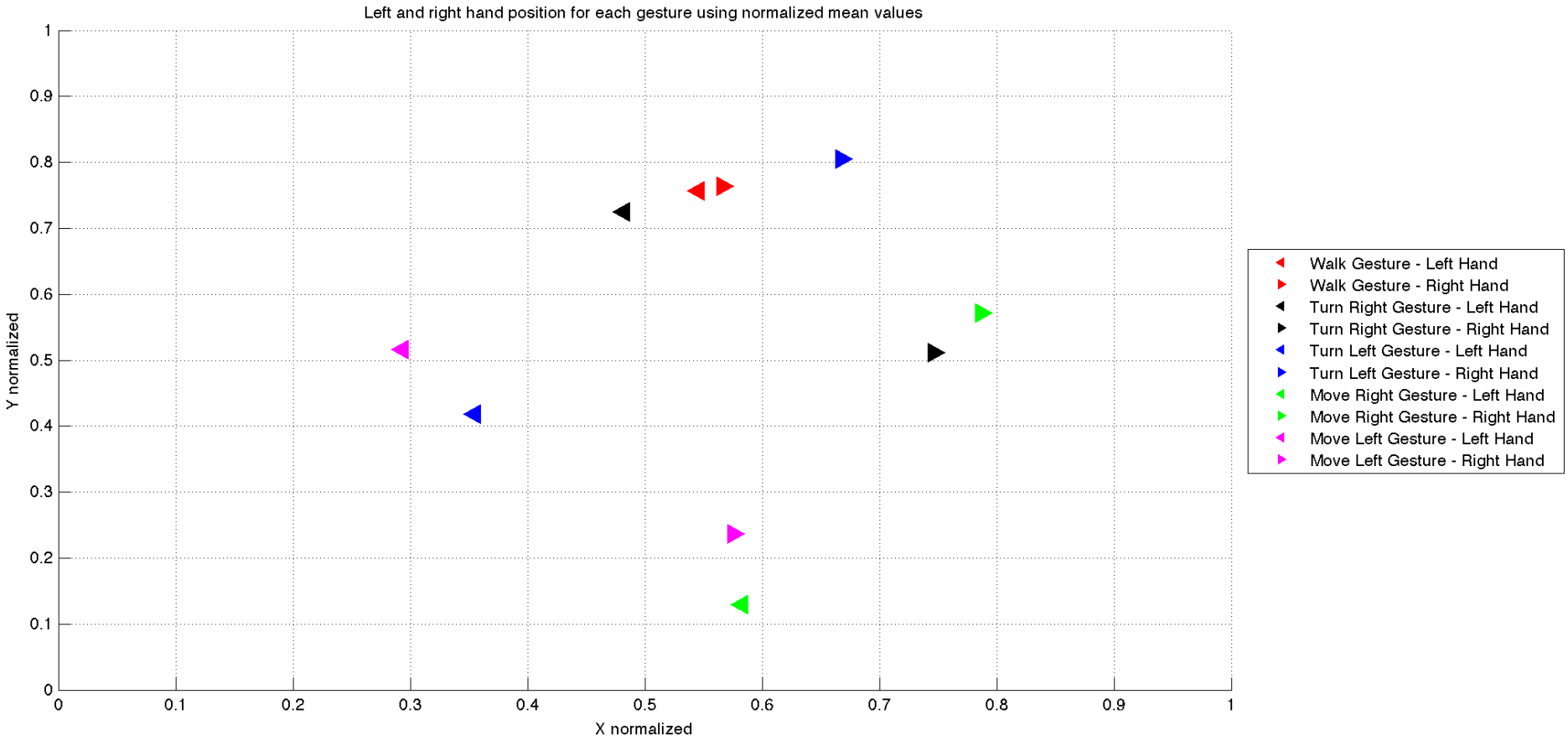
Implementation - CC Module

- Control Center module
  - Starts WebSocket client to connect to Brain module
  - Acts as the eye of the project to visualize the internal interactions between modules
  - Cross compatible app and needs just latest browser
  - Renders skeletal joint positions in 3D
  - Display prediction results and info messages
  - Developed in Javascript using WebStorm IDE on Mac OSX
  - Uses WebGL renderer
  - Uses libraries such as ThreeJS, RequireJS, jQuery, underscore and native JS websocket
  - Can replay from dumped data

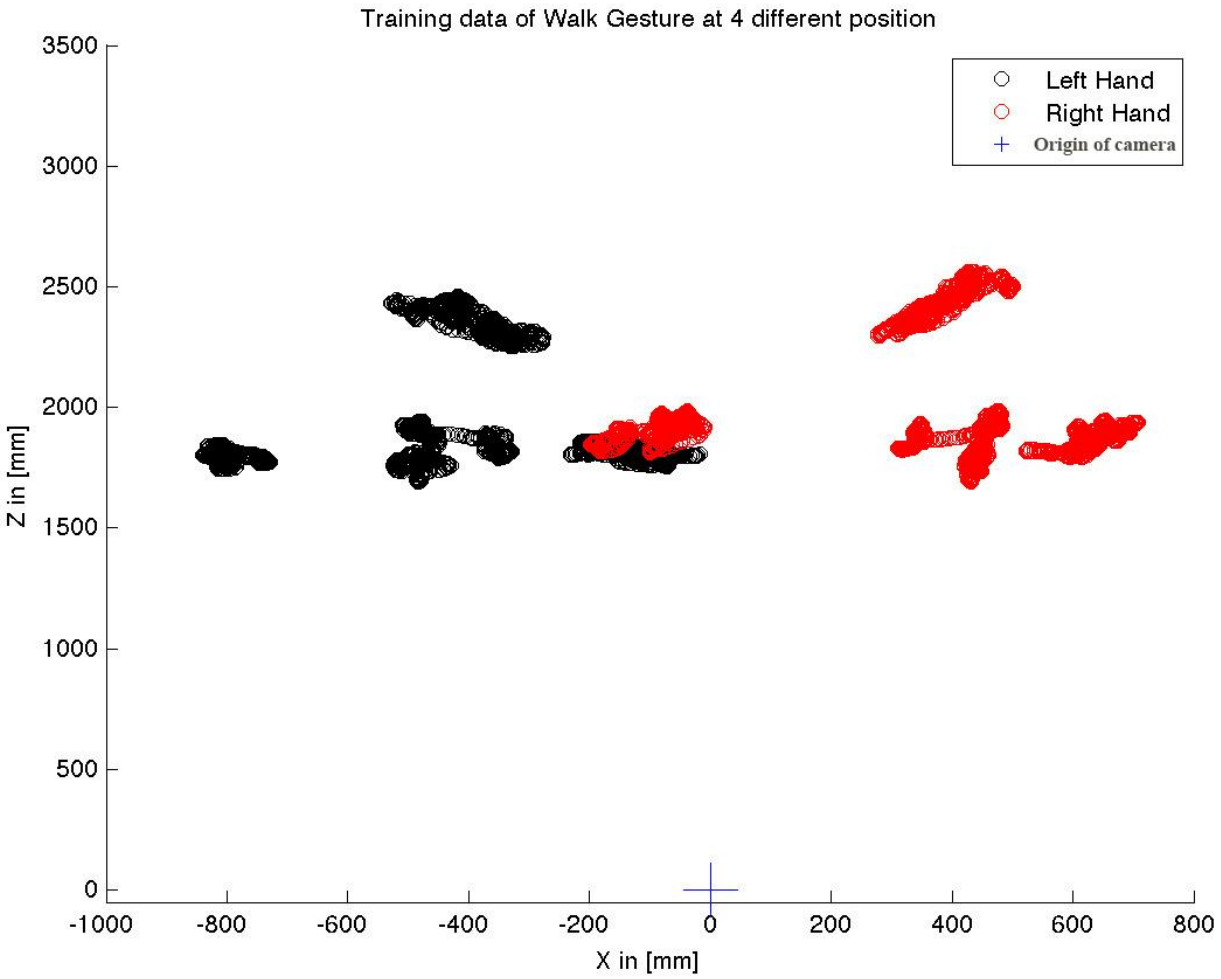
Implementation - Command Module

- Command module
  - Starts WebSocket client to connect to Brain module
  - Uses NAOqi SDK to proxy ALMotion, ALRobotPosture, ALTextToSpeech
  - Receives recognized gestures and info messages from Brain module
  - Commands NAO to do text-to-speech, locomotion and joint control tasks
  - Executes Gesture-to-Speech, Gesture-to-Motion and Gesture-to-Gesture translations
  - Developed in Python using PyCharm on Mac OSX
  - Completes the hand gesture recognition for Human-robot interaction

Training - Mean positions



Training - Min-Max Distance



Results

Demo

Results - Confusion Matrix, Precision, Recall

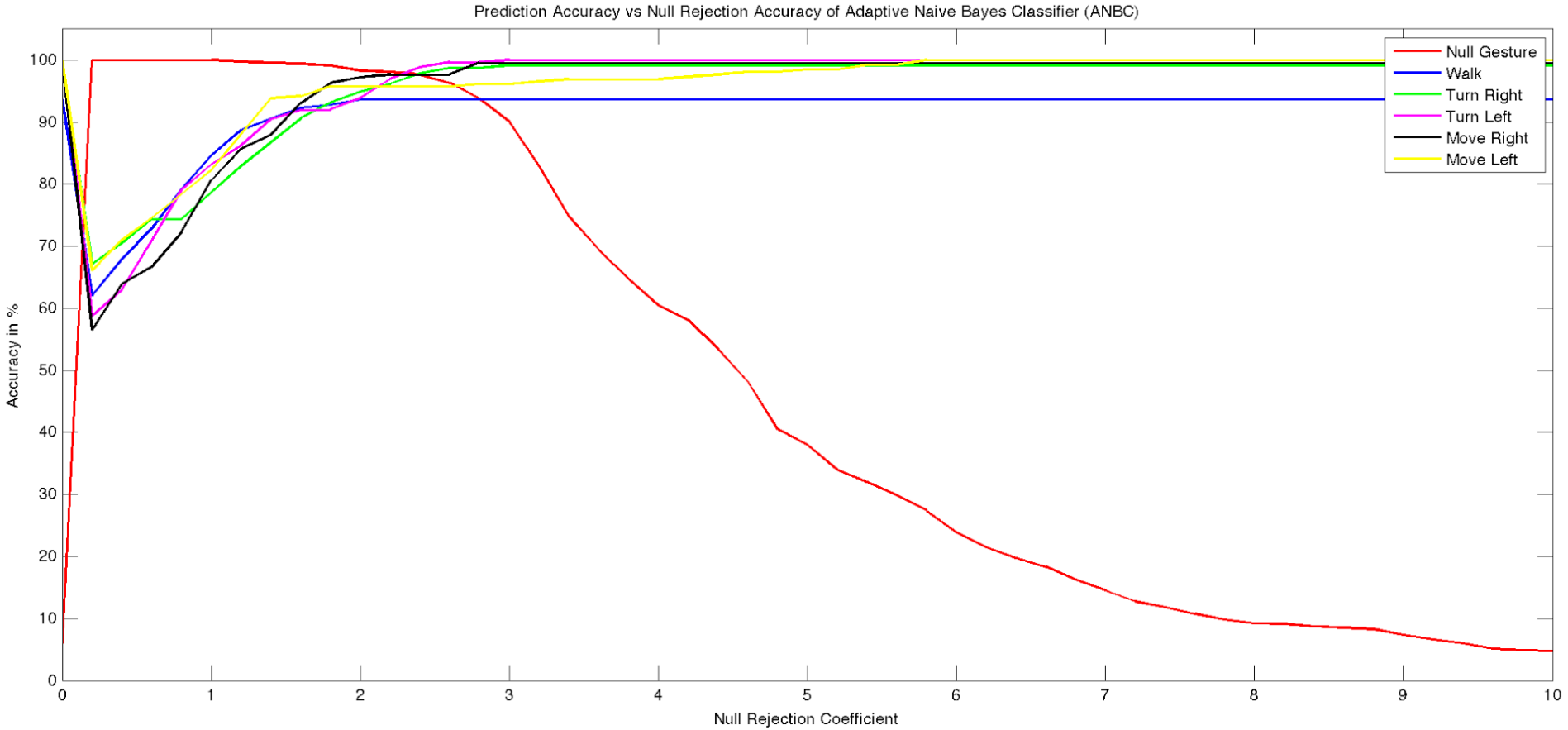
	Class 1	Class 2	Class 3	Class 4	Class 5
Precision	0.990	0.969	0.996	1.000	1.000
Recall	0.937	0.949	0.939	0.972	0.958
F-measure	0.963	0.959	0.967	0.986	0.978

Precision, Recall and F-Measure calculated by validating 10% of training dataset. ANBC Classifier trained with Null Rejection coefficient 2.0

	Non-Gesture	Class 1	Class 2	Class 3	Class 4	Class 5
Non-gesture	0.000	0.000	0.000	0.000	0.000	0.000
Class 1	0.032	0.937	0.032	0.000	0.000	0.000
Class 2	0.043	0.009	0.949	0.000	0.000	0.000
Class 3	0.061	0.000	0.000	0.939	0.000	0.000
Class 4	0.023	0.000	0.000	0.005	0.972	0.000
Class 5	0.042	0.000	0.000	0.000	0.000	0.958

Confusion Matrix calculated by validating 10% of training dataset. ANBC Classifier trained with Null Rejection coefficient 2.0

Results - Accuracy



Conclusion & Future work

- Results show that the implementation achieves the goal by building a robust system for NAO to facilitate human-robot interactions based on skeletal points tracking using depth camera.
- It can be further improved to recognize more static gestures by training more and dynamic gestures by extending it with classifiers such as Dynamic Time Warping or Hidden Markov Model which are readily available in GRT.
- Everything (source code, results, literatures, training data) that was done and gathered during this thesis are available at https://github.com/AravinthPanch/ gesture-recognition-for-human-robot-interaction

Conclusion

Thank you very much for your attention