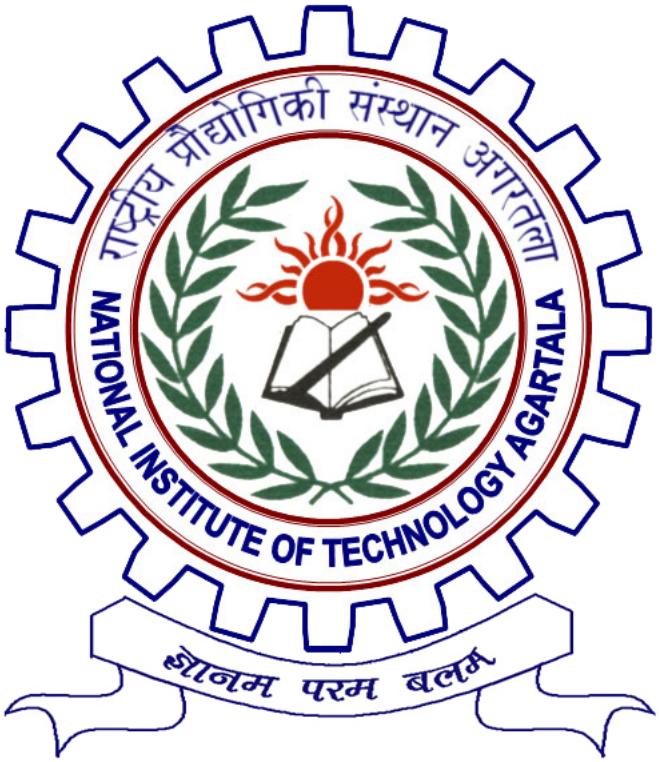
Exploratory study on humanoid robot NAO for designing enhanced HMI(Human Machine Interaction)



NIKHIL SHARMA

COMPUTER SCIENCE AND ENGINEERING
SHRI MATA VAISHNO DEVI UNIVERSITY, KATRA
INDIA- 182320

1 Acknowledgement

I would like to take this opportunity to express my deep sense of gratitude to all who helped me directly or indirectly during this dissertation work. Firstly, I would like to thank my guide, ***Mr Suman Deb****, for being a great mentor and the best adviser I could ever have. His advise, encouragement and critics are source of innovative ideas, inspiration and causes behind the successful completion of this dissertation. The confidence shown on me by them was the biggest source of inspiration for me. It has been a privilege working with Suman sir. Sir is really is legend personality, his enthusiasm and will to give back to the society always encourages me. He is the most hardworking and dedicated faculty member I have ever meet. I would also thank all my seniors, especially Tutan sir and Barnita mam, who helped me a lot and made this summer training a memorable experience for me. Apart from training and academics , NIT Agartala's peaceful environment also helped me to come with success of my summer training and joyful experience.*

NIKHIL SHARMA

Contents

2 Introduction

Human is considered as most intelligent animal on this planet earth.Can human intelligence be defeated by some machine ?

From past several years scientist has developed most advanced computational machines , some of them can also look like humans,they can act like humans, performs daily routine tasks. They can be used in industries as well as in homes. These are the ROBOTS. Robots can be social as well as industrial. The most social and human like robots are called Humanoid robots.

The NAO robot, which is used in the current work, is a humanoid robotic platformmanufactured by Aldebaran Robotics. It is mainly used for education and research. The robot itself is about 58 cm high, capable of movement and speech, and possesses a range of sensor types and on-board computational capabilities (1.6 GHz CPU, 1 GB RAM, 2 GB flash memory, 8 GB micro SDHC). Connecting to the robot is possible by Ethernet and Wi-Fi.

The most important sensors for the current work are the two front cameras able to take pictures in a resolution range from ***160 X 120*** *up to* ***1280 X 960****. It is important to note that the cameras are not actually placed in the "eyes" of the robot, but one on the forehead and the other in position of the "mouth", as can be see in figure, enabling vision downwards without moving the head. While the cameras are able to produce a continuous image stream, only single, static images are usable for object recognition in the current implementation.*

2.1 Scope and Motivation

Machine learning and deep learning algorithm are continuously improving the lifestyle and helping the mankind. The combination of artificial intelligence and mechatronics give rise to Robotics,which brings the intelligence more physically into the machines.

The application of robotics can be seen in many fields like medical science and defence , where robots can perform those difficult tasks more accurately and efficiently which are beyond human capabilities. A tangible robot or machine can be controlled intangibly through human gestures and human voice. With current rapid progress in the development of autonomous robots for lots of possible applications in all fields of activity, it is only natural that intelligent systems need effective ways for not only perceiving, but also understanding their surroundings through the means of robotic or, more general, computer vision.

The goal is to understand the functioning of robots and develop intelligent ways by which robots can interact with humans. Our motto is to eliminate the dependency for interacting with robots but to develop fully autonomous robots,which can interact with human by their own.

2.2 Objective

Learning the principles of robotics and understanding functioning of humanoid robot, its very fundamental operations. Analyzing the existing systems and methods to interact with human Understanding of kinematics and vision of NAO robot and improving it with the help of machine learning models , and hence enhancing HMI(Human Machine Interaction)

- *To understand the fundamental functioning of Humanoid robot, NAO and finding the scope of implementation of deep learning algorithms to enhance robot's vision and kinematics.*
- *To identify and extract learning outcome from implementation model*

3 Literature Review and Related Works

3.1 Literature Review

3.1.1 Human and Humanoid

Human has a beautiful ,well structured body equipped with lots of sensors and receptors and also a brain to take actions based on those receptors data. The humanoid robot NAO also have many receptors and sensors like camera, microphone ,speaker, tectile sensor, sonar, and IR sensors.

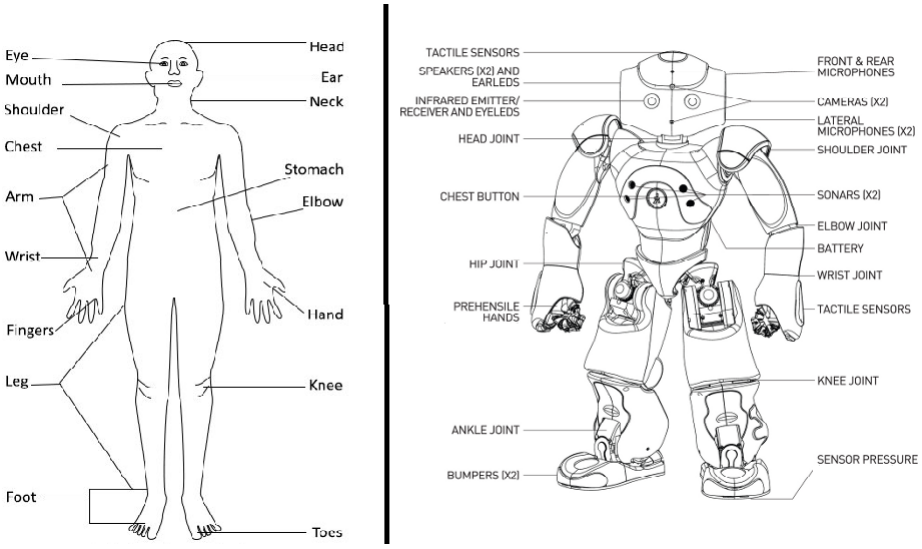


Figure 1: body structure of Human and NAO

Human body has several joints and each joint has different types of degrees of freedom(DOF). A human body has 244 DOFs with different movements and a Nao robot has 25 DOFs. Like humans NAO is also equipped with advanced fall detection manager.It can recognize human face and voice and based on sound localization system turn its head in the direction of sound and human face.

3.1.2 Object Recognition in NAO

Many machine learning and deep learning algorithms are used for the detection of human face ,voice and objects. Face and object detection are part of computer vision technology and voice recognition uses natural language processing. NAO can be trained by implementing deep learning models moreover CHOREGRAPH , which is a main framework in interaction with NAO , also allow NAO to learn a human face and further detection and recognize it. similarly, human voice command can be feed into humanoid robot as text and it can recognize it by using ALTextToSpeech API. An object's dimensions and features can be feed to NAO ,manually to detect that object to an extend. Although it is not the efficient way to make NAO objects ,faces and voice.

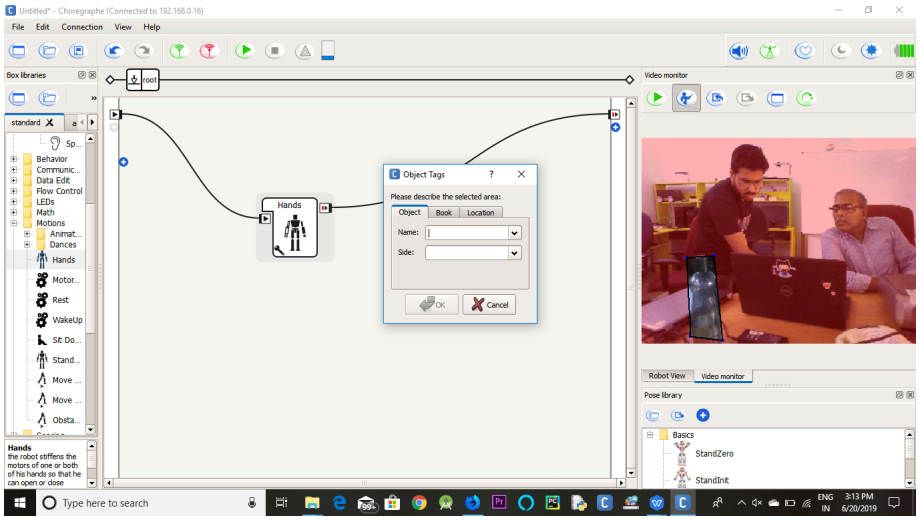


Figure 2: learning of objects by bounding box manually through NAO camera

3.2 Related works

1. ***Implementing deep learning object recognition on NAO*** *: In this bachelor's thesis, a way is shown to implement a ready-to-use object recognition implementation on the NAO robotic*

platform using Convolutional Neural Networks based on pre-trained models. Furthermore, the implementation offers a graphical user interface with several options to adjust the recognition process and for controlling movements of the robot's head in order to easier acquire objects in the field of view.

2. ***A Deep Learning Approach for Object Recognition with NAO Soccer Robot :*** *In this paper, they propose a method for validating the results provided by color segmentation approaches. In particular, they present a novel algorithm for merging an adaptive segmentation procedure with a Deep Learning based validation stage.*

4 Fundamentals

4.1 NAO

The Aldebaran NAO robot, which is used in the current work, is a humanoid robotic platform manufactured by SOFTBANK ROBOTICS . It is mainly used for education and research. The NAO robot is about 58 cm high, capable of movement and speech, and possesses a range of sensor types and on-board computational capabilities (1.6 GHz CPU, 1 GB RAM, 2 GB flash memory, 8 GB micro SDHC). Connecting to the robot is possible by Ethernet and Wi-Fi.

The most important sensors for the current work are the two front cameras able to take pictures in a resolution range from 160 X 120 up to 1280 X 960. It is important to note that the cameras are not actually placed in the "eyes" of the robot, but one on the forehead and the other in position of the "mouth", enabling vision downwards without moving the head. While the cameras are able to produce a continuous image stream, only single, static images are usable for object recognition in the current implementation.

The field of view for each camera is 47.64 vertically and

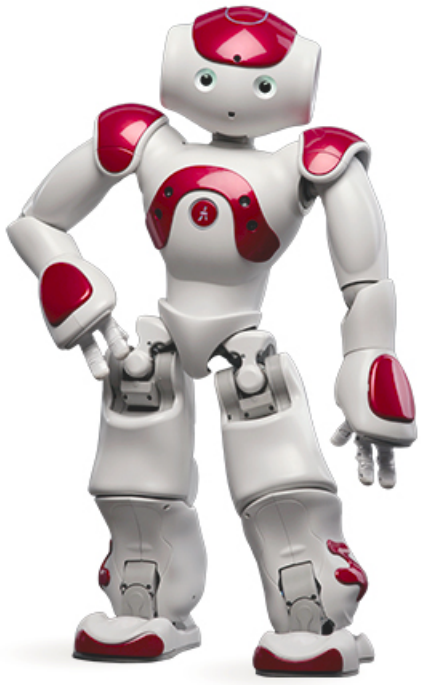


Figure 3: NAO

60.97 horizontally . Downwards inclination is 1.2 for the top camera, 39.7 for the bottom one. Other hardware components of immediate importance are the microphones and the loud speakers, used for speech recognition in the dialogue system 6.5 and returning the object recognition's results acoustically in addition to console output. The head joint relevant for the movement controls in 6.4.1 offers two degrees of freedom.

The programming languages can be used in designing and functioning of varoius operations in NAO are C++, MATLAB, JAVA, LabVIEW and Python. In the current work ,we have used python for writing programs for NAO.

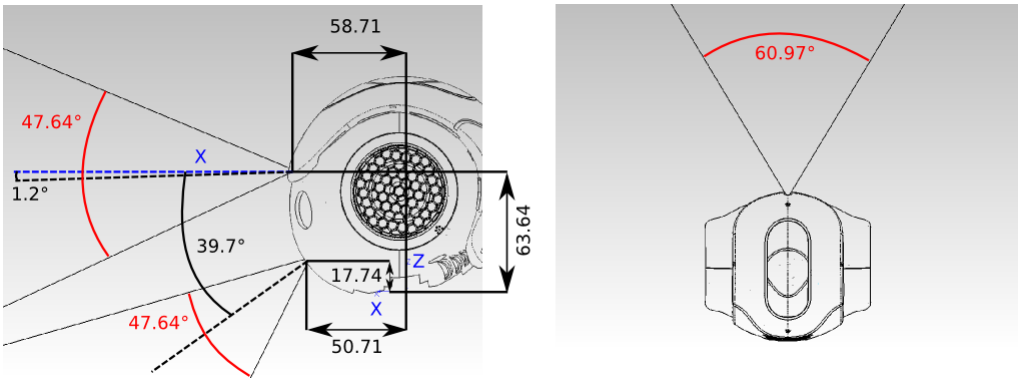


Figure 4: field view of NAO camera

Python script, when provided with NAO's IP address and port number, can directly run programs on NAO.

NAO is equipped with many sensors which include Infrared, Sonar and FSR(Force sensing resistors which used in proving Dynamics and kinematics of humanoid robot.

Because NAO may fall, we taught him to stand up. But we went even further by establishing a system to detect falls: it protects with her arms before touching the ground. NAO can therefore fall, he will not hurt. By default, the fall manager is active on NAO.

ALMemory: *ALMemory is a centralized memory used to store all key information related to the hardware configuration of your robot.*

More specifically, ALMemory provides information about the current state of the Actuators and the Sensors. For further details, see: NAO - Actuator Sensor list.

4.1.1 Choregraph

Choregraph is a Aldebaran's framework which work on the NAOqi (operating system of NAO) . It is a most flexible method to interact with NAO. We can inbuilt methods to perform a particular task. It always give a real time structural view of NAO. Any program or function can be simulated on Choregraph. If we does not have a real robot with us, we can use VIRTUAL ROBOT for the testing of function.

The from very basic operations like movement related commands to the complex face detection and speech recognition can be easily performed by NAO through Choregraph. We can also use VIDEO MONITOR to see through NAO's camera, and hence apply face and object detection techniques.

The main application of Choregraph is , we can create complex behavior by using drag and drop functions. By this a person with no technical knowledge can operate NAO easily.

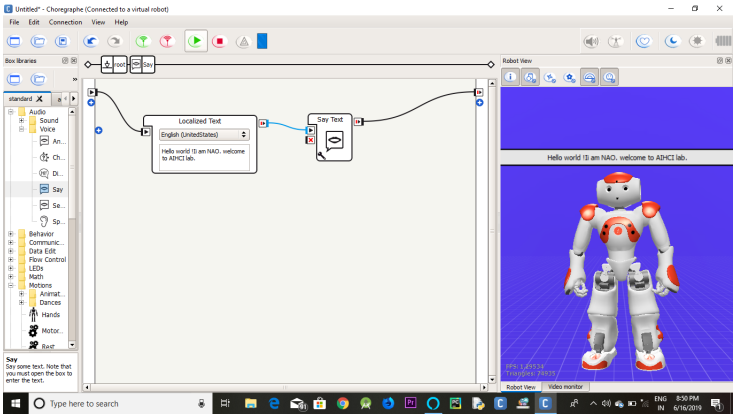


Figure 5: Simulation in choregraph

4.1.2 Monitor

Monitor is dedicated to give you an elementary feedback from your robot and a simple access to its camera settings.

Its modular architecture allows you to load plugins in different mobile widgets, each of them being connected to the robot of your choice. That is to say you can connect several robots at a time.

Mainly 2 plugins are available:

1. *Camera viewerYou can use Monitor's camera viewer to see through the robot's eyes. You can also configure the camera settings from there (brightness,contrast...), and finally understand why everything seems to go wrong.*
2. *memory viewer This plugin allows you to watch the data held by the ALMemory module of a given NAOqi. Number-like data evolution can also be plotted. This is very useful to diagnose behaviors through the internal data they use. You can also follow hardware data like those you can see on your NAO Web page in the Advanced ¿ Hardware section.*

    A. Viewer setting

    B. Camera setting

    C. Action button

    D. Video recorder tab

4.1.3 V-REP

The robot simulator V-REP, with integrated development environment, is based on a distributed control architecture: each object/model can be individually controlled via an embedded script, a plugin, a

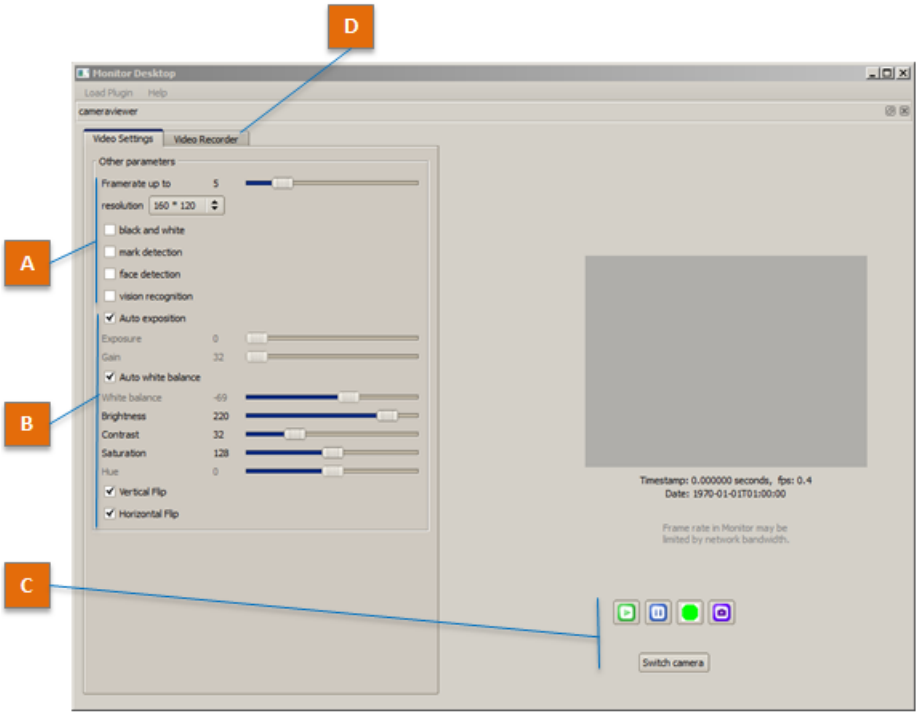


Figure 6: Camera view plugin in monitor

ROS or BlueZero node, a remote API client, or a custom solution. We can create a world,similar to real surrounding, for NAO in V-REP, where it interact with animated objects and perform various tasks. .

4.1.4 WEBOTS

Webots is an open-source robot simulator. It is widely used in industry, education and research.

It's a 3D robot simulator for creating worlds for robots. We can design 3D objects and able to work with Nao for those objects Webots uses a fork of the ODE (Open Dynamics Engine) for detecting of collisions and simulating rigid body dynamics. The ODE library allows one to accurately simulate physical properties of ob-

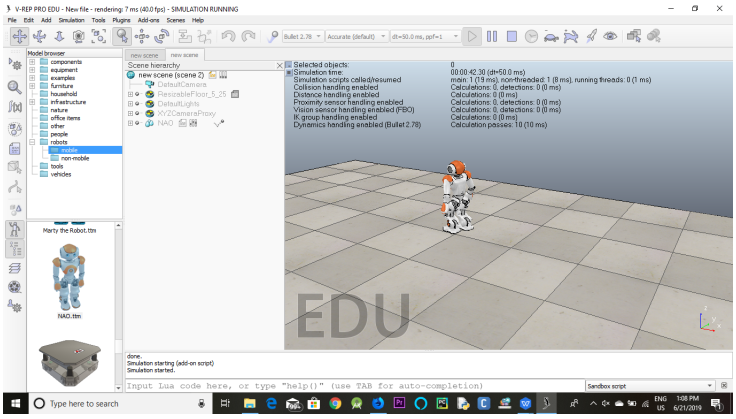


Figure 7: Simulation in v-rep

jects such as velocity, inertia and friction.

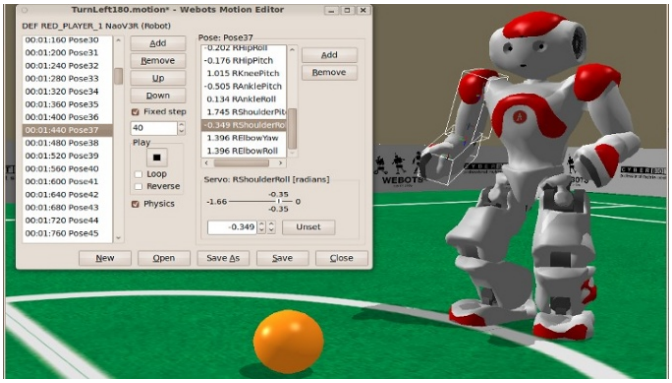


Figure 8: Simulation in webot

5 System Architecture

5.1 NAO vision

NAOqi has various proxy modules for the different functioning of the Robot. The module which work on the vision of the NAO robot

is called "ALvision" and the module which deals with its motion is called "ALmotion". This section describes the vision components of your robot.

NAOqi currently contains the following vision modules:

ALBacklightingDetection: enables to check if the image of the camera is backlit.

ALDarknessDetection: gives NAO the ability to detect if it is in a dark environment.

ALFaceDetection: gives NAO the ability to detect and recognize human faces.

ALLandmarkDetection: gives NAO the ability to detect specific visual landmarks.

ALMovementDetection: gives NAO the ability to detect some movement and tell where it comes from.

ALPhotoCapture: provides tools to take pictures and save them on disk.

ALRedBallDetection: gives NAO the ability to detect red and circular objects.

ALVideoDevice: manages video inputs.

ALVideoRecorder: provides video recording tools.

ALVisionRecognition: provides learning and recognition capabilities of visual patterns.

ALVisualCompass: gives NAO the ability to use an image as a compass.

ALVisionToolbox (deprecated): provides tools to record / analyze camera images. Use ALBacklightingDetection, ALDarknessDetection, ALPhotoCapture and ALVideoRecorder instead.

The ALVisionRecognition module is part of the NAOqi framework. It provides functionalities for learning to identify specific objects, which have to be labeled by hand. However, the capabilities offered by this module can be seen as the opposite of what the current work tries to implement. While ALVisionRecognition enables the user to teach the robot specific object instances by using visual key points. Visually different objects can not be identified, if they have not been specifically learned beforehand. Additionally, ALVisionRecognition, as provided by NAOqi, has no concept of object classes at all.

In contrast, by using CNNs for object recognition , it is possible to recognize objects never seen before, if the object category they belong to is part of the training set underlying the model in use. On the other hand, in its current form, the implementation does not have the ability to recognize whether a specific object has been encountered before.

5.2 HMI in NAO

The ability of robots to interact in a socially intelligent manner with humans is the core of human-robot interaction (HRI). The quality of this interaction is typically measured in terms of how it is engaging to the users either reflected in duration of time users spend interacting with a robot, or their self-reports on engagement during the interaction.

Currently,NAO detects the face of human and react to it, it also works on sound localization, i.e. NAO turns its head it the direction from which the sound is coming.It can also track a human by its face when move around the NAO.

The learning ,detection and recognition of face is not much effective because it does not have much processing power to process the image of face effectively. And the object recognition is also not proper, it needs to be manually locate the object and provide it the required parameters for each object.

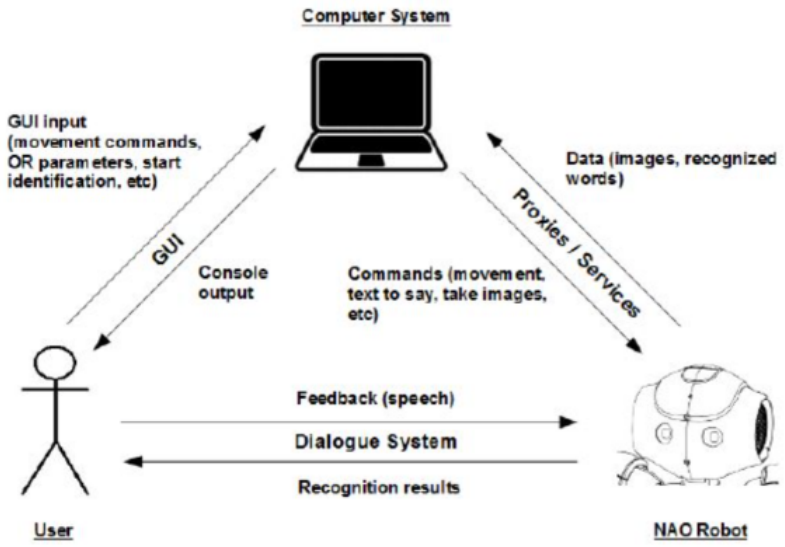


Figure 9: Human machine Interaction in NAO

NAO is completely controlled by a another machine to which it is connected over a router. Robot can be given commands and used to run the programs written on that computer.

5.3 Machine Learning Models

Machine learning and Deep learning algorithms are continuously improving over the existing technologies and finding the more effective and efficient manner.Taking into consideration the Object identification and recognition, which currently can be done manually and also not a effective approach, replaced by deep learning methods using models like CNN and ***YOLO****(you only look once).These models work on a given data-set and identifies and recognizes objects even it has not seen them before. These work require object identification and classification, to know from which class it belongs.*

As these models require deep neural network and high computational power for training of data set, witch is not available

in case of NAO, training can be done using GPU on remote system connected to NAO.

As the NAO robots are used in ***ROBOSOCCER*** *game, they need to recognize the position of ball and its co players and identify its teammates. This is only possible when it has ability to process real time data and extract the feature out of it. For this purpose we use* ***YOLO****, which is one of the best real time video processing model for extracting features.*

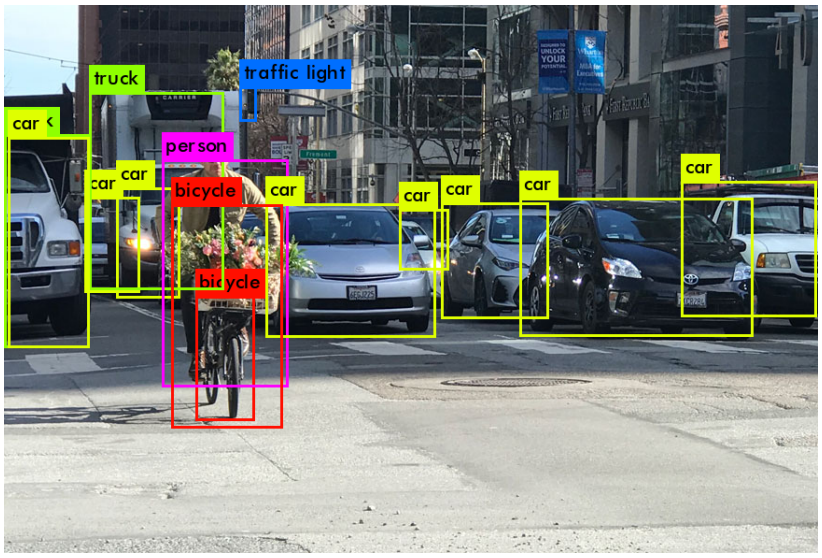


Figure 10: Real time object detection with YOLO

To develop a better human machine interaction, machine should identify the basic the features like face recognition , object recognition and voice recognition and continuously improving it with the optimization of algorithms and improvement of hardware.

6 Conclusion and Future Work

NAO, the humanoid robot can find many applications in our home. It is a social humanoid robot, we can always control its action through

a remote machine to which the NAO is connected.The motion of this robot can also be enhanced using more smoother and human like structures.

The detection of objects could also be enhanced through some additions to Multibox, like suppressing duplicate detections, or using a leaner and faster solution for region proposals altogether, if one exists. As for future work, object recognition opens up a lot of possibilities. It enables the robot to better analyze its environment and interact with it. An interesting addition would be a system capable of remembering classifications of objects and their position in the world. The data base could be similar to ontologies used in artificial intelligence. For example, actions possible with the object class "bottle", like lifting one without spilling the content and pouring it into a glass, are probably applicable to most kinds of bottles, regardless of them being wine or water bottles, and could be carried out after successful recognition. While this might seem far-fetched, it only goes to show the possibilities offered by effective and robust object recognition in combination with robotic platforms.

The further work in HMI(Human Machine Interaction) find ways to control NAO using human gestures and this too using NAO vision,to give it more human like characters.

7 References

Here we have some references which used in the current work.

1. Aldebaran NAO documentation : http://doc.aldebaran.com/2-1/home$_n$ao.html

2. NAOqi documentation: http://doc.aldebaran.com/2-5/index$_d$ev$_g$uide.html

3. Choregraph tutorial video series: https://www.youtube.com/watch ?v=SSD8DsVSrj4list=PLma9tC7VHPpi8Bz4i2P5FuMlfMhjZ3nJ-

4. Choregraph tutorial documentation: http://doc.aldebaran.com/2-1/software/choregraphe/tutos/index.html

5. Webot download guide: https://cyberbotics.com/webots

6. Implementing deep learning Objecet recognition on NAO: https://hdms.bsz-bw.de/frontdoor/deliver/index/docId/5394/file/ $Implementing_Deep_Learning_Object_Recognition_on_NAO.pdf$

7. A Deep Learning Approach for Object Recognition with NAO Soccer Robot :https://link.springer.com/chapter/10.1007/978-3-319-68792-6$_3$3