Design of a Voice Controlled Robotic Gripper Arm using Neural Networks

Fariha Musharrat Haque
Department of Mechanical Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
farihahaque019@gmail.com

Asif Shahriyar Sushmit
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
sushmit0109@gmail.com

M. A. Rashid Sarkar
Department of Mechanical Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
rashid@me.buet.ac.bd

***Abstract*— The aim of this work is to propose a method to build an efficient Bangla voice controlled robotic gripper mechanism using Neural Networks. Robots are becoming an essential part of many industries and fields. Presently, various ways are used to control one. The most user friendly one of them is controlling it by voice commands. Though voice controlled robots are becoming a popular concept now, construction of Bangla voice controlled robots is still a new idea. Controlling the robot with voice commands along with visual feeds helps the robot to operate easily and more accurately. This robot consists of three modules: speech command recognition module, object classifier module and robotic gripper arm module. At first, the robot takes voice commands on which objects to grab and displace; then it finds the object using the object classifier module. And finally it grabs and displaces the object using the robotic gripper arm module. The speech recognition module and the object classifier module uses two distinct neural networks along with additional hardware to perform their tasks. This paper presents the design and fabrication process of the robot discussed so that robots can be made using this design that works under different situations. This robot can be used to perform tasks with a high efficiency on both industrial and domestic levels.**

***Index Terms*— Neural Networks, Robotic Gripper, Speech Recognition.**

I. INTRODUCTION

Speech recognition and voice recognition have gained momentum in the last few years thanks to neural networks and deep learning. Speech recognition deals with getting the information out of an audio file whereas voice recognition ensures receiving information as well as security since voice recognition can distinguish between audio samples from different users. This work focuses on speech recognition. The conventional speech recognition process is based on Hidden Markov Models(HMM). In this approach the input command is taken and then processed and verified against acoustic and language models. But at present since Deep Neural Nets are on the rise, using DNNs to do the pattern recognition is more convenient. Interested reader may check the work Fayek, Lech and Cavedon [1] where frame based formulation to speech emotion recognition was described using deep learning.

An object classifier is used to take the visual feed and relate to the voice commands. And then a Robotic Gripper Module is used to pick up the target object. To perform the last part with accuracy the work of Levine, Pastor, Quillen et al [2] can be regarded as pioneering. They used hand eye co-ordination for robotic grasping using deep learning. Their system uses a grasp prediction network to choose the motor commands for the robot that maximized the probability of a successful grasp. For the robotic gripper part, the work of Preseren, Augustin and Mravlje [3] can be referred to where they indicated the guidelines for designing the gripper arm. Different types of grippers are used for various purposes.

There have been very few works on Bangla voice controlled robots in the past. One mentionable work was done by Bhattacharjee, Khan, Haidar [4] on Bangla voice control robot for rescue operation in noisy environment. In their paper they generated a small codebook in order to implement from a short list of commands. They used traditional speech recognition algorithms in their process. Scheider, Sturn and Stachniss have shown [5] the use of tactile sensors to classify objects without any visual feed. Reference to the works by House, Malkin Bilmes[6] and RZhou, KPNg and YSNg[7] can also be made as the different models are closely related to the current work and can be further developed to build other innovative systems. We are discussing our model combining our ideas with some ideas from their works to develop a low cost

978-1-5386-1887-5/17/$31.00 ©2017 IEEE

II. SPEECH RECOGNITION MODULE

Instead of traditional algorithms, Recurrent Neural Networks (RNN) are better at speech recognition. The decision of choosing RNN over other ML or DL methods is that in a feed forward neural network, signal flow in only one direction from input to output. But in RNN, each step of the output is tied to the prior input.

So RNNs are more suitable while using speech recognition as the sample length is a variable of time.

The speech recognition module of our model consists of:

- Low noise audio input portion
- Computing portion

For the prototype version, at first the training input audios are taken by a microphone and then sampled at 16kHz and then quantized. After performing fast Fourier transformation and noise reduction on the data, it is plotted and after performing further processing, the data is fed to a recurrent neural network having 15 (or more, depending upon conditions) layers. This part of the model uses MATLAB to build the network.

For the command recognition, an average quality microphone is used that takes audio input and determines which of the preset commands were given at that instance. For the prototype four bangla words are being used to train the neural net. 20-50 audio files need to be used for each command while training the neural net.

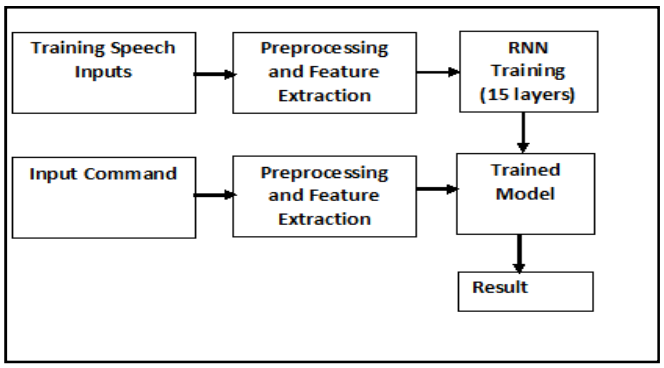

Diagram1: Speech Recognition Module

III. OBJECT CLASSIFIER MODULE

A simple object classifier is the second module here that uses image processing and a neural network. The classifier in the prototype version is designed to choose from three different objects to determine the position of the object that is to be picked up.

A portable 8MP camera in front of the gripping mechanism is used to take the visual feed. A separate hardware (a Raspberry-pi Model3) is used to do the computation and to ensure the fact that there is little or no delay. After the robot recieves and analyzes a voice command to pick up a specific object, the classifier searches for that specific object by taking a photo of the vision field and then the position of the object is known. After taking a photo, it is converted into a simplified photo by applying gradient maps and background cancellation. The image processing is done using opencv. We will be using opencv for further development too as it is open source. The classifier discussed here uses Tensorflow. And the programming language is Python.

The training images are to be taken from various angles under different lighting conditions to ensure maximum accuracy. Preferably. at least 150 pictures are needed to do the initial training under 5 lighting conditions for any efficient execution of this model. This increases the accuracy of detection and ensured greater rate of success at pinpointing the target object.

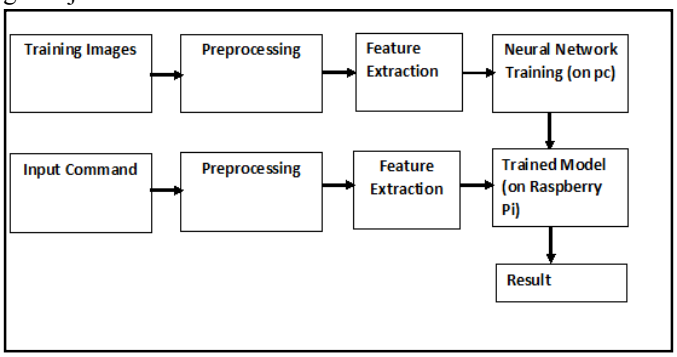

Diagram2: Object Classifier Module

IV. ROBOTIC GRIPPER ARM

For the gripper part, a custom made 3D printed robotic arm with parallel jaw grippers are to be used which is controlled by three servo motors. While designing the gripper arm using SolidWorks some features requires attention.

1. Configuration of the gripper footprint
2. Choosing the required shape and size for clamps, brackets and extension with the possibility of further modification into more flexibility to adjustment
3. Selecting the right gripper considering the application
4. Actuation shaft to enable mobility

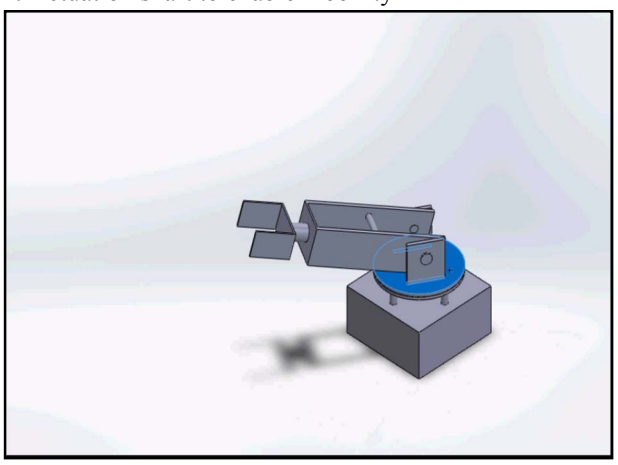

Fig 1: Gripper Arm that can be subjected to a 360 degrees rotation along the horizontal plane. (the placement of the motors not shown in the figure)

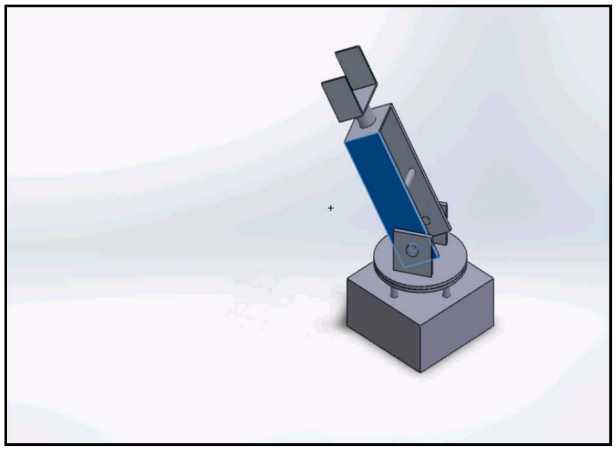

Fig 2: Gripper arm that can be subjected to a 180 degrees rotation along the vertical plane. (the placement of the motors not shown in the figure)

After the voice command is then activated to reach that position and pick up the object command activates the object classifier to find the target object, we get the position of the object. In the prototype, the positions are one of three predetermined positions. The servo motors are given definite instructions according to predefined points for object placement to make the model simple. Potential improvement on this section is discussed later.

The mechanism consists of the i) Object engagement & ii) Force locking. Then the object is released into a bucket placed at the right.

The release cycle works on the same procedure in the reverse sequence. This whole process is conducted using an Arduino Mega. After putting the desired object in the bucket, the gripper comes back at the initial position and the voice command module is activated again for the next command.

V. LIMITATIONS AND FUTURE WORK

What have been discussed here is just the blueprint of a flexible design. Using slight modifications on this general design model (using different training data and slightly different grippers) can lead to different specialized designs. We are working on building a prototype for a industrial grade specialized version Bangla voice controlled robotic gripping mechanism depending on this design.

The prototype so far recognizes only four bangla voice commands by different users. If NLP is incorporated with this model, then the area of its uses can be broaden. But no such significant work have already been done regarding such Bangla language processing that could be incorporated while developing this robot.

The gripping module here in this discussed design uses preset commands to grab the desired object that have been placed in one of three different preset positions. Incorporating hand-eye co-ordination [2] can lead to a significant amount of flexibility and accuracy. But this approach could not be included in this model due to lack of resources, both computational and financial. It is left for future works instead.

The prototype is under process to incorporate more voice commands and the classifier is undergoing process for recognizing a wide variety of objects. Different approaches are being tested to reduce the amount of computation while training so that this process can be done by the user using their version of this robot.

VI. APPLICATION

There is probably only a few sectors where the voice controlled arm cannot be implemented in this era of rapid increase of automation. This kind of robots are well suited in industries where the bots can be commanded by voice commands to continue performing a certain task. The second major aspect of this is that it can be used for remote control easily after slight modification. These robotic gripper mechanisms can also be used in the medical sector. Very low cost simple prosthetic arms can easily be developed as a further modification of this prototype model. With the emergence of robotic surgery this voice controlled model may serve to enhance the purpose of distant surgeries as well as increasing the efficiency by performing tasks during nanoscale operations. The most outstanding breakthrough is the restoration of mobility and independence of paralyzed patients by using neural network [8], [9]. Also, this gripper can serve as a rescue bot during natural disasters. Since the working principle is based on voice recognition, rescue operations can be easily performed since the bot is programmed to react to specific audio input and cancel out irrelevant noises. It can also be entrusted with household chores and work as an efficient assistant in laboratories. It can also serve as an intelligent pet like cozmo and act as a mood booster.

VII. CONCLUSION

Standing on the stairway of advancing Neural Network based modules in practically every aspect of modern lives, it is perhaps high time to replace the manual parts with highly efficient Neural Network incorporated machine parts. This voice controlled hand can potentially recognize any commands in Bangla and act accordingly if trained well. Rescue missions can be conducted in dangerous situations, remote surgeries can be performed, productivity in the industrial sections can be subjected to rapid growth and further modification of this module can render accessibility of usage in basically any sector. A low cost robotic system is very much needed in the context of the developing economy of Bangladesh where Bangla is the main spoken language.

References

[1] Haytham M. Fayek, Margaret Lech, Lawrence Cavedon "Evaluating deep learning architectures for Speech Emotion Recognition" in Elsevier Journal on Neural Networks, August 2017 special Ed, vol 92, pp 60-68.

[2] Sergey Levine, Peter Pastor, Alex Krizhevsky, "Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection" on 2016 International Symposium on Experimental Robotics, 21 March 2017, pp173-184.

[3] Greg Causey, "Guidelines for the design of robotic gripping systems" on Assembly Automation, vol 23, Issue 1, pp18-28.

[4] Arnab Bhattacharjee, Asir Intisar Khan, M.Z.Haider," Bangla Voice Controlled Robot for Rescue Operation in Noisy Environment" on Region 10 Conference (TENCON) 2016 IEEE.

[5] Alexander Schneider, Jugen Strum, Cyrill Stachniss, "Object Identification with Tactile Sensors using Bag-of-Features" on Intelligent Robots and Systems, 2009.

[6] Brandi House, Jonathan Malkin, Jeff Bilmes, "The VoiceBot: a voice controlled robot arm" on CHI '09, Proceedings of the SIGCHI Conference on Human Factors in Computing System, pp 183-192.

[7] R. Zhou, K.P.Ng, Y.S.Ng, "A Voice Controlled Robot Using Neural Network" on Intelligent Information Systems, 1994.

[8] Meel Velliste, Sagi Perel, M. Chance Spalding, "Cortical control of a prosthetic arm for self-feeding" on Nature 453, 1098-1101, 19 June 2008.