

GESTURE RECOGNITION INTERPRETATION AND TRANSMISSION (GRIT)

A Report

submitted by

Akshat Sharma (13BCE1009)

Antariksh Narain (13BCE1017)

in partial fulfillment for PBL

of

**CSE-323
Computer Networks**

School of Computing Science and Engineering



November – 2015

ACKNOWLEDGEMENT

It gives us great pleasure to bestow my profound sense of gratitude and reverence to Prof. Pradeep K.V., School of Computer Science Engineering, for his unequivocal supervision during the course of the project work. We feel privileged to have opportunity to work under her unfailing and astute guidance.

(Akshat Sharma-13BCE1009)

(Antariksh Narain-13BCE1017)

CONTENTS

TOPIC	PAGE NO.
CERTIFICATE	2
ACKNOWLEDMENT	3
ABSTRACT	5
PROBLEM DEFINITION	6
SYSTEM REQUIREMENTS	7
INTRODUCTION	8
SYSTEM DESIGN	9
IMPORTANT CODE	11
GRAPHS	12
IMPLEMENTATION	13
CONCLUSION	15
REFERENCES	16

ABSTRACT

The objective of this project is to recognize simple gestures such as alphabets and numbers. The gesture would be converted to text. This text can then be sent as a message using Networking. The motive is to control most of the functionality of the messenger using gestures.

Features such as the server checking that the receiver client is available or not, notification on receiving of the message will be part of the project. If the receiver is connected then the message would be sent else the message will be stored on the server. When the user returns the message will be delivered to him/her.

For understanding the gestures we are using image processing on multiple frames of the gestures video and passing the processed frames as input to a neural network which would classify them.

The window is fully controllable by gestures, such as swiping between windows, writing text to be send and to use clickable buttons.

PROBLEM DEFINITION

There are many application in the digital market to exchange text, image, audio and video messages. Currently there are hundreds of companies in the market competing for the top position by adding new services now and then. Features like group chat, broadcast, stickers etc. have been added which keeps them in the race. People are looking for cool features in these application which makes it fun to use and hassle free.

We plan to develop a prototype in MATLAB. The system will recognize different gestures and interpret them according as actions or words. The message generated will be send to appropriate receiver on the network. In case of unavailability of the client on receiver's end, the message will be stored in the server's user database. When the user comes online the message will be delivered.

As the application is gesture based, people who are **visually impaired** can also take the benefit of it. As no physical touch is involved in the system with the device, sharing messages becomes even easier, one just needs a display and gesture recognition camera around.

SYSTEM REQUIREMENTS

Recommended:

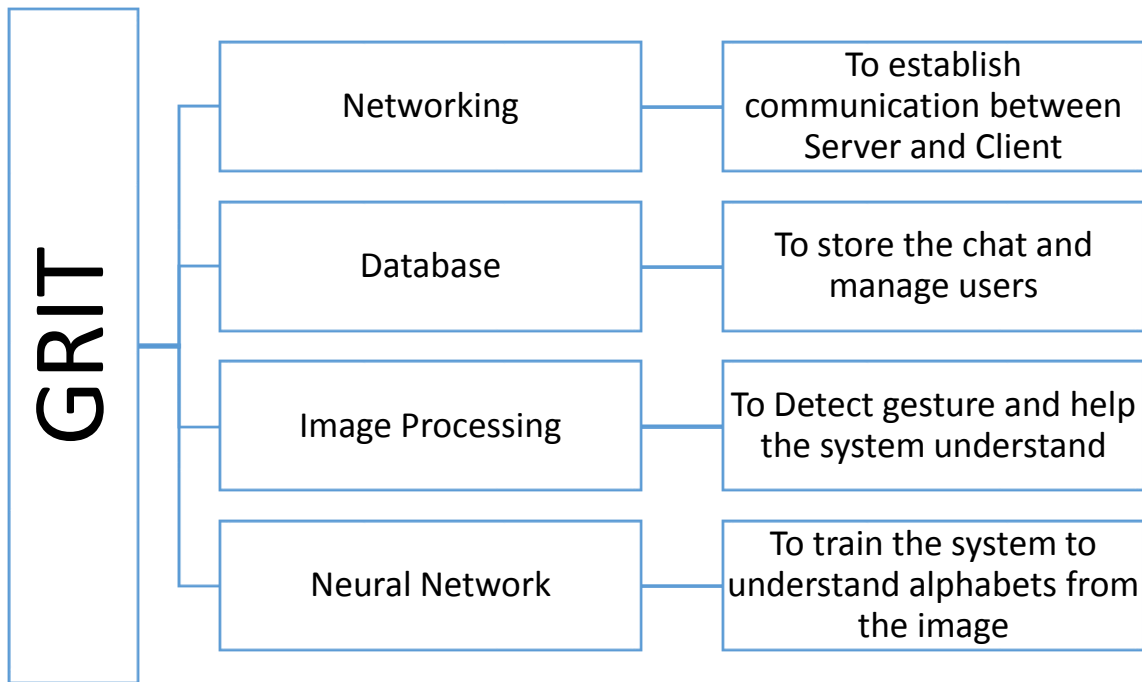
- 2.4 GHz Dual Core Processor.
- 8 GB RAM
- Windows/Linux Operating System
- MATLAB 2014a +
 - Webcam Hardware Package
 - ODBC Database connector for MySQL
 - Image Processing Toolbox
 - Database Toolbox
 - Parallel Processing Toolbox
- MySQL 5.6.*

A System is required which acts as a server and N systems which connect to the server as clients.

INTRODUCTION

The project includes Networking, Image Processing, Machine Learning and Database Management.

Hence it has been sub divided into four sections



Each part has its own importance. Integrating these multiple system into a single system makes **GRIT (Gesture Recognition Interpretation and Transmission)** a complete project.

This project does not aims to overcome the drawbacks of existing system but instead it focuses on making the existing system more fun to use.

Imagine you are watching TV and your phone is in your room. Consider you receive a message, this message using networking will be displayed on the TV and using gesture camera create a message and send it back.

SYSTEM DESIGN

Network:

- Creating a server which has a dedicated connection to the database.
- Server should be concurrent in nature so multiple clients can connect to the server at a single instant of time and request can be processed in parallel.
- Message should be received and send only from a registered user.
- Constant log of a user should be maintained.
- The IP Address and Port of each client needs to be stored, so data can easily be directed to a client.
- TCP/IP protocol should be used for communication as it fast and secure.
- Following functions need to be created:
 - New_server – which creates a new independent connection to a client on request.
 - Set_user – send user details to the User table
 - Get_Data – to get data from user's Data table and send it to the connected client.
 - Get_Chat – to get chat between two users and send it to the connected client.
 - Send_user – to send message to a specified user, i.e. to add to Chat table.

Database:

- The server will be using MySQL database.
- A database named Project should be created to which the server connects and excess the tables.
- A connection should be created on a port for the server to connect to for sending and receiving SQL queries.
- Following tables need to be created:
 - User – to store username, password, IP Address and Port
 - Data_* – to store unread messages of each user.
 - Chat_*_* - to store the complete chat history of client A and client B.

Image Processing:

- A program that tracks red color.
- Use the above program to detect gestures and words.
 - To detect motion of finger and classify it as sliding up/down or left/right action.
 - To write a word using a pointer, create a frame that stores the position of finger with some thickness.
 - Segment the frame such that each alphabet is separated out.
 - Send the frame formed to the neural network to classify each alphabet from image to text.
- Send the text to the networking console to be forwarded.
- Concatenate multiple images to create a panorama. Program that creates a sliding effect with a panoramic image and integrate with the gesture. Only a part of the panorama should be visible on the screen.
- A program needs to be created for converting text to image for the chat window screen.

Neural Network:

- A dataset with at least 50 images per character.
- Train the neural network with the classifier on the images in the dataset.
- Using the classifier, classify each image character to its corresponding text. Choose the character with maximum probability as the answer.
- Combined all the letters to make a word.
- Send this word to the dictionary, and using Bayes's Probability calculate the probability of being a word in the dictionary. Send the word with the maximum Bayes's Probability as the expected word and send it to the network.

IMPORTANT CODES

Detect Red Color:

```
cam=webcam;
data = fliplr(snapshot(cam));
ycbcr=rgb2ycbcr(data);
diff_im=ycbcr(:, :, 3);
diff_im = medfilt2(diff_im, [3 3]);
diff_im = im2bw(diff_im, 0.65);
diff_im = bwareaopen(diff_im, 300);
bw = bwlabel(diff_im, 8);
stats = regionprops(bw, 'BoundingBox', 'Centroid');
```

TCP/IP Connection

Server:

```
t = tcpip('0.0.0.0', Port, 'NetworkRole', 'server');
fprintf(t, '%s', Id);
while t.bytesAvailable==0
end
tmp=fread(t, t.bytesAvailable);
```

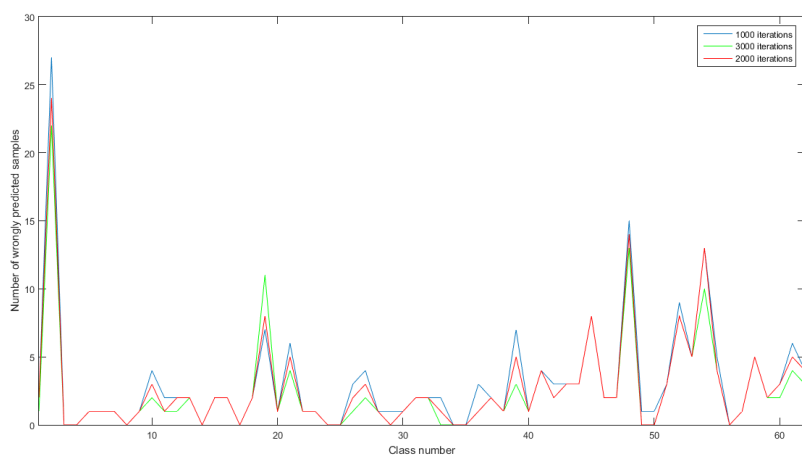
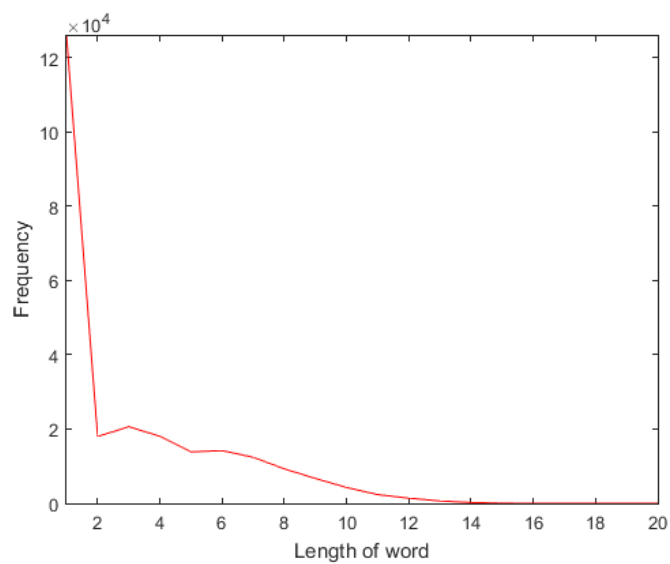
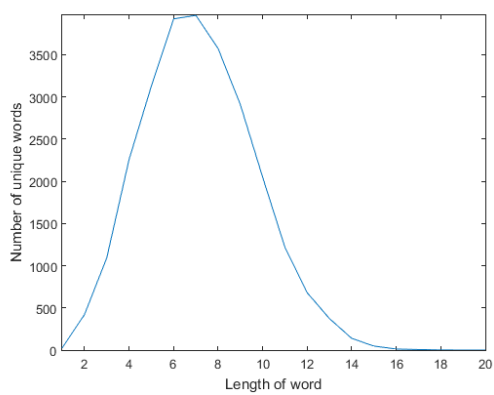
Client:

```
t = tcpip('Server IP', Port, 'NetworkRole', 'client');
fprintf(t, '%s', Id);
while t.bytesAvailable==0
end
tmp=fread(t, t.bytesAvailable);
```

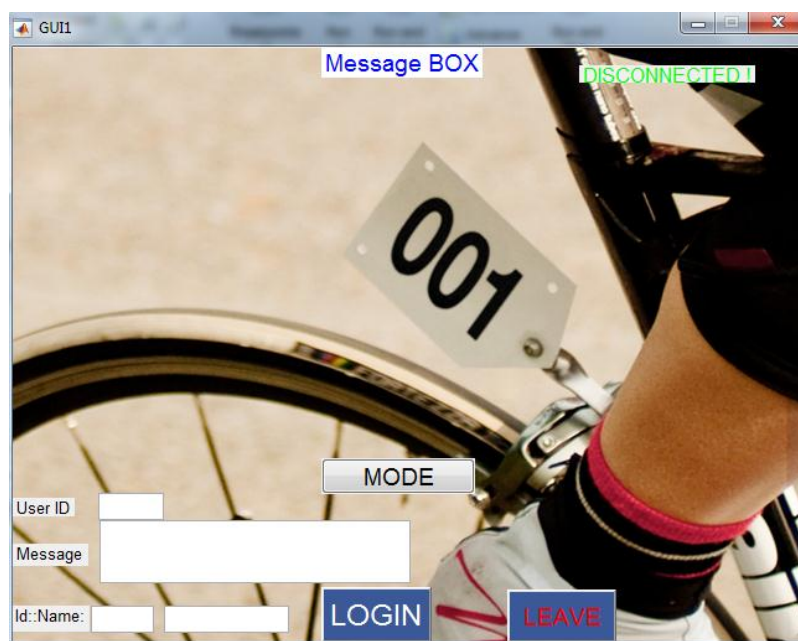
Bayesian Probability:

```
function [maxword]=findword(b,plist)
max=0.0;
maxword="";
load c;
for i=1:size(plist{2},1)
    if size(b,2)==size(plist{1}{i},2)
        x=prob(b,plist{1}{i},c(plist{1}{i}));
        if max<=x
            max=x;
            maxword=plist{1}{i};
        end
    end
end
return
```

GRAPHS



IMPLEMENTATION



INPUT:

Holmes

OUTPUT:

```
>> a=imread('ho.png');  
>> [s,k]=wordsplit(a);  
>> [p,h]=predict(Theta1,Theta2,s);  
>> toword(p')
```

ans =

H0zMe6

```
>> w=findword(h',plist)
```

w =

holmes

SPLIT

WORD



The System Detect it as 'holmes'.

CONCLUSION

We understand that it is going to take time for gesture based means to communicate with machines to take over other means of interaction, mainly because gesture based systems are still in their infancy stage. However that fact the gesture based systems are more powerful than the presently available systems, cannot be understated.

It is clear that a combination of naïve based model with a neural network is a more versatile system than a regular neural network to detect words from a language. Since there are 2 datasets involved we can have one dataset determining what the characters look like and the other containing words of the language. This increases the system accuracy on the whole.

Such a gesture based system can be used by the visually impaired and by the regular public for more control over the Digital World.

With a coming up of technologies like Microsoft Holo Lens, the need for such system cannot be denied.

REFERENCES

- norvig.com/spell-correct.html
- www.mathworks.in/help/database/ug/mysql-jdbc-windows.html
- <https://www.coursera.org/learn/machine-learning/home/welcome>
- <http://in.mathworks.com/matlabcentral/fileexchange/>
- www.mathworks.com