A Project Report

on

An Optimum Approach to Indian Sign Language Recognition using Efficient Convolutional Neural Networks

Submitted in partial fulfillment of the requirement for the award of the degree of

Computer Science Engineering

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Name of Supervisor:
Mr. Suresh N Kumar
Professor

Submitted By:

NIKUNJ BHARDWAJ

(18SCSE1010408)

RITIK KUMAR

(18SCSE1010646)

]

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING, GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"SIGN LANGUAGE RECOGNITION"** in partial fulfillment of the requirements for the award of the Btech submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Name… Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Nikunj Bhardwaj18SCSE101048

Ritik Kumar 18SCSE1010646

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor Name
Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Nikunj Bhardwaj 18SCSE1010408, Ritik Kumar 18SCSE1010646. has been held on _____ and his/her work is recommended for the award of Btech

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: May, 2021
Place: Greater Noida

Acknowledgements

I would like to acknowledge the great efforts and precious time given by our project guide Mr. Suresh Kumar N Assistant Professor Galgoltias university Greater Noida, Uttar Pradesh School of Computing Science and Engineering.

We were able to complete our job thanks to great advice and criticism. We'd also want to thank our parents for their constant support and care. Last but not least, without our team, we would not have been able to build this project and learn so much in the process.

ABSTRACT

Sign Language is the only formal way of communication for the mute persons and the hearing impaired. Developing systems that can recognize these signs provide an interface between signers and non–signers by which the meaning of sign can be interpreted. The aim of this project work is to design a user independent framework for automatic recognition of Indian Sign Language which is capable of recognizing various one handed dynamic isolated signs and interpreting their meaning. The proposed approach consists majorly of three steps: pre-processing, feature extraction and recognition. In the first step, i.e., the pre-processing, skin color detection is done followed by face elimination and key frame extraction. A novel method for key frame extraction is proposed which finds out the distinguishing frames from the input video. This algorithm speeds up the system by finding out the most important frames for feature selection. In feature extraction phase, various hand shape features like circularity, extent, convex deficiency and hand orientation are considered. Finally, in the recognition phase, Multiclass Support Vector Machine is used to classify the signs and recognize them. Experimentation with vocabulary of 22 signs from ISL is conducted. Results prove that the proposed method for recognition of gestured sign is effective and having high accuracy.

Contents

List of Figures

Chapter 1

INTRODUCTION

The recent developments in Human Computer Interaction (HCI) have contributed to the growth of Machine Learning. With this push towards growth, technology has made a major contribution to social applications such as awareness of sign language. The prime means of communication between a signer and a non-signer is sign language. It is the only structured communication medium through which mute individuals and hearing impaired individuals use body movement or facial expressions to express themselves. To communicate information, sign language recognition systems mainly use hand gestures. Hand gesture recognition is another widely growing area which has many applications in various areas such as entertainment industry including interactive virtual tourisms and computer games, controlling industry including 2D and 3D pointing, control of household appliances, objects grasping, command descriptions. Most of the gestures area unit performed with the hand however conjointly with the face and therefore the body. Hand form along with its movement and position with respect to the opposite body elements, forms a hand gesture. Hand gestures area unit employed in each facet of human communication to accompany speech, or alone for communication in howling surroundings and area unit the foremost vital facet of the facet of linguistic communication as most of the data is communicated through hand by the signers.

Sign Language Recognition (SLR) is a hybrid research field involving pattern recognition, natural language processing, computer vision and linguistics. Sign Language Recognition Systems can be used as an interface between human beings and computer systems to create algorithms and techniques to correctly recognize a

sequence of created signs and understand their context. Sign languages area unit complete tongue with their phonemics, morphology, syntax and synchronic linguistics. a signal language could be a visual-gesture language that's developed to facilitate the otherwise abled persons by making visual gestures victimisation face, hand, body and arms.

The main approaches used in sign language recognition system can be classified as: vision based approach or device based approach. Vision based techniques are used mostly in case of SLR. These techniques make use of image features like color, shape, texture etc. and also require the pre processing of the input image. The input to the vision based systems can be an image or video which is either recorded from single or multiple cameras or captured in real time through imaging device which is connected to the system. In device based approaches, devices such as motion sensors, data gloves etc. are used. These devices measure the hand shape and motion. These approaches are inconvenient to signer because they require cumbersome devices to be worn by the user and also restrict their movement.

The main dimensions of research in SLR can be grouped as: isolated sign recognition and continuous sign recognition. The first type i.e. the isolated sign recognition is concerned with the recognition of single signs without continuation to the other sign. These signs can be either static or dynamic. No other sign is performed before or after the isolated signs in continuation. Thus, this type of sign is not affected by the preceding or succeeding sign. In continuous signing, a complete sentence is to be recognized consisting of various signs performed one after the other. The aim is to identify the different signs being performed continuously.

CHAPTER 2

LITERATURE SURVEY

Hand gesture recognition provides an intelligent, natural, and convenient way of human– computer interaction (HCI). Sign language recognition (SLR) and gesture-based control are two major applications for hand gesture recognition technologies . SLR aims to interpret sign languages automatically by a computer in order to help the deaf communicate with hearing society conveniently. Since sign language is a kind of highly structured and largely symbolic human gesture set, SLR also serves as a good basic for the development of general gesture- based HCI. In this chapter we are discussing work done in the area of hand gesture recognition and analyze the methods for recognition of hand gesture.

Sunitha K. A, Anitha Saraswathi.P, Aarthi.M, Jayapriya. K, Lingam Sunny, "Deaf Mute Communication Interpreter- A Review", International Journal of Applied Engineering Research ,Volume 11, pp 290-296 , 2016: This paper aims to cover the various prevailing methods of deaf-mute communication interpreter system. The two broad classification of the communication methodologies used by the deaf – mute people are - Wearable Communication Device and Online Learning System. Under Wearable communication method, there are Glove based system, Keypad method and Handicom Touch-screen. All the above mentioned three sub-divided methods make use of various sensors, accelerometer, a suitable micro-controller, a text to speech conversion module, a keypad and a touch-screen. The need for an external device to interpret the message between a deaf –mute and non-deaf-mute people can be overcome by the second method i.e online learning system. The Online Learning System has different methods. The five subdivided methods are- SLIM module, TESSA, Wi-See Technology, SWI_PELE System and Web-Sign Technology.

Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, " An Efficient Framework for Indian SignLanguage Recognition Using Wavelet Transform" Circuits and Systems, Volume 7, pp 1874- 1883, 2016: The proposed ISLR system is considered as a pattern recognition technique that has two important modules: feature extraction and classification. The joint use of Discrete Wavelet Transform (DWT) based feature extraction and nearest neighbour classifier is used to

recognize the sign language. The experimental results show that the proposed hand gesture recognition system achieves maximum 99.23% classification accuracy while using cosine distance classifier.

Mandeep Kaur Ahuja, Amardeep Singh, "Hand Gesture Recognition Using PCA", International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015: In this paper authors presented a scheme using a databasedriven hand gesture recognition based upon skin color model approach and thresholding approach along with an effective template matching with can be effectively used for human robotics applications and similar other applications.. Initially, hand region is segmented by applying skin color model in YCbCr color space. In the next stage thresholding is applied to separate foreground and background. Finally, template based matching technique is developed using Principal Component Analysis (PCA) for recognition.

Sagar P.More, Prof. Abdul Sattar, "Hand gesture recognition system for dumb people": Authors presented the static hand gesture recognition system using digital image processing. For hand gesture feature vector SIFT algorithm is used. The SIFT features have been computed at the edges which are invariant to scaling, rotation, addition of noise.

International Journal of Science and Research (IJSR): In this paper a method for automatic recognition of signs on the basis of shape based features is presented. For segmentation of hand region from the images, Otsu's thresholding algorithm is used, that chooses an optimal threshold to minimize the within-class variance of thresholded black and white pixels. Features of segmented hand region are calculated using Hu's invariant moments that are fed to Artificial Neural Network for classification. Performance of the system is evaluated on the basis of Accuracy ,Sensitivity and Specificity.

Chandandeep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering: Authors presented various method of hand gesture and sign language recognition proposed in the past by various researchers. For deaf and dumb people, Sign language is the only way of communication. With the help of sign language, these physical impaired people express their emotions and thoughts to other person.

Pratibha Pandey, Vinay Jain, “Hand Gesture Recognition for Sign Language Recognition: A Review”, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015: In this paper author proposed a system to aid communication of deaf and dumb people communication using Indian sign language (ISL) with normal people where hand gestures will be converted into appropriate text message. Main objective is to design an algorithm to convert dynamic gesture to text at real time. Finally after testing is done the system will be implemented on android platform and will be available as an application for smart phone and tablet pc.

Nakul Nagpal,Dr. Arun Mitra.,Dr. Pankaj Agrawal, “Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People”, International Journal on Recent and Innovation Trends in Computing and Communication ,Volume: 3 Issue: 5,pp- 147 – 149: Author proposed a real time vision based system for hand gesture recognition for human computer interaction in many applications. The system can recognize 35 different hand gestures given by Indian and American Sign Language or ISL and ASL at faster rate with virtuous accuracy. RGB-to-GRAY segmentation technique was used to minimize the chances of false detection. Authors proposed a method of improvised Scale Invariant Feature Transform (SIFT) and same was used to extract features. The system is model using MATLAB. To design and efficient user friendly hand gesture recognition system, a GUI model has been implemented.

Neelam K. Gilorkar, Manisha M. Ingle, “Real Time Detection And Recognition Of Indian And American Sign Language Using Sift”, International Journal of Electronics and Communication Engineering & Technology (IJECET), Volume 5, Issue 5, pp. 11-18 , May 2014: Paper presented the recent research and development of sign language based on manual communication and body language.
Sign language recognition system typically elaborate three steps pre processing, feature extraction and classification.Classification methods used for recognition are Neural Network(NN), Support Vector Machine(SVM), Hidden Markov Models(HMM), Scale Invariant Feature Transform(SIFT),etc.

Neelam K. Gilorkar, Manisha M. Ingle, "A Review on Feature Extraction for Indian and American Sign Language", International Journal of Computer Science and Information Technologies, Volume 5 (1) , pp314-318, 2014: Author presented application that helps the deaf and dumb person to communicate with the rest of the world using sign language. The key feature in this system is the real time gesture to text conversion. The processing steps include: gesture extraction, gesture matching and conversion to speech. Gesture extraction involves use of various image processing techniques such as histogram matching, bounding box computation, skin colour segmentation and region growing. Techniques applicable for Gesture matching include feature point matching and correlation based matching. The other features in the application include voicing out of text and text to gesture conversion.

Ashish Sethi, Hemanth ,Kuldeep Kumar,Bhaskara Rao ,Krishnan R, "Sign Pro-An Application Suite for Deaf and Dumb", IJCSET , Volume 2, Issue 5, pp- 12031206, May 2012: In this paper, off-line signature recognition & verification using neural network is proposed, where the signature is captured and presented to the user in an image format.

CHAPTER 3

PROBLEM STATEMENT, ISSUES AND OBJECTIVE

3.1 PROBLEM STATEMENT

Differently able people are usually deprived of normal communication with other people in the society. Also normal people find it difficult to understand and communicate with them. These people have to rely on an interpreter or on some sort of visual communication. An interpreter won't be always available and visual communication is mostly difficult to understand. Sign Language is the primary means of communication in the deaf and dumb community. As a normal person is unaware of the grammar or meaning of various gestures that are part of a sign language, it is primarily limited to their families and/or deaf and dumb community.

At this age of technology, it is quintessential to make these people feel part of the society by helping them communicate smoothly. Hence, an intelligent computer

system is required to be developed and be taught. Researchers have been attacking the problem for quite some time now and the results are showing some promise. Interesting technologies are being developed for speech recognition but no real commercial product for sign recognition is actually there in the current market.

3.2 Issues and Challenges in Indian Sign Language

Developing a system for deaf people to recognize sign language is a very challenging issue due to its cardinal social interest and its intrinsic complexity. The various issues and challenges related to Indian sign language are as follows:

Issue: Geographical variation in Indian Sign Language.

Challenge: Creating a standard data-set for ISL. Since the sign language within India differs in different regions, there exist no standard data-set for Indian sign language.
Issue: Segmentation

Challenges: The various challenges which for segmenting the face and hands are occlusion, presence of disturbances and cluttered background. There are possibilities of hand occluding the face which makes segmentation of hand typical. Various disturbances like illumination change, noise, and undesired movement also cause problem in segmentation. It is often difficult to segment the face and hands in cases where the background is cluttered.

Issue: Recognition Accuracy

Challenges: Variation in viewpoint. The major difficulties of sign language recognition are feature extraction and the recognition itself. Secondly, nature of information is multi-channel. This means that face, body, hand can be used in parallel to communicate a sign.

Issue: Tracking.

Challenge: Signing speed varies from person to person. Articulated objects (such as the hand) are more difficult to track than single rigid objects. Because gestures are highly variable from one person to another, and from one example to another within a

single person, it is essential to capture their essence -their invariant properties- and use this information to represent them.

Issue: Analysis and integration of manual and non-manual signals

Challenge: Usage of both static and dynamic hand gestures. The non-manual signs are typical to recognize. In certain cases, the signs use complex hand shape

Issue: Gesture Spotting

Challenge: The task in gesture spotting is to differentiate the meaningful gestures of the user from the unrelated ones. The garbage movements that come before and after a pure gesture should be removed. Succeeding signs are affected with preceding sign in continuous sign language recognition which is called co articulation problem.

3.3 Objective

The objective of this project work is to develop a robust automatic system to recognize signs from Indian Sign Language using vision based approach for one handed dynamic isolated signs. The proposed system will translate the video of the sign to text. The proposed system's performance will be tested on large vocabulary of sign from ISL with reduced training samples. The motive of this work is to provide a real time interface so that signers can be able to easily and quickly communicate with non-signers. In India, there is a need of developing an automatic sign language recognition system, which can accomplish the need of hearing impaired people. Unfortunately, till date not much research work has been reported on Indian Sign Language recognition. Moreover, our work has been done on various signers for testing. This had lead the proposed system to be user independent which does not restricts different users to use our proposed system.

CHAPTER 4

DESIGN AND IMPLEMENTATION

Design is the first step in the development phase for any engineering product or system. Design is the planning that lays the basis for the making of every object or system. This document consists of several phases.

4.1 Detailed Design

4.1.1 Methodology

The Methodology used in our project is Convolutional Neural Network. Convolutional neural network are the special type of feed forward artificial neural network in which the connectivity between the layers are inspired by the visual cortex. Basically the convolutional neural networks have 4 layers that is the convolutional layers, ReLU layer, pooling layer and the fully connected layer. In convolution layer after the computer reads an image in the form of pixels, then with the help of convolution layers we take a small patch of the images. These images or patches are called the features or the filters. By sending these rough feature matches is roughly the same position in the 2 images, convolutional layer gets a lot better at seeing similarities than whole image matching scenes. These filters are compared to the new input images if it matches then the image is classified correctly. Here line up the features and the image and then multiply each image, pixel by the corresponding feature pixel, Add the pixels up and divide the total number of pixels in the feature. We create a map and put the values of the filter at that corresponding place. Similarly, we will move the feature to every other position of the image and will see how the feature matches that area. Finally, we will get a matrix as an output.

4.1.2 Use Case diagram

A use case is an interaction between users and the system. It captures goal of users and the responsibility of the system to its user. A use case diagram is a group of actors, a set of cases enclosed by a system boundary, communication associations between the

actors and use case and generalization among the use case. An actor specifies the role played by the person or thing when interacting with the system.
An actor is someone one something outside the system that either acts on the system is a primary actor or is acted on by a system is a secondary actor. Actors represent the different roles that something outside has in its relationship with the system whose functional requirements are being specified.

A use case describes a sequence of actions between system and actors which is drawn as a horizontal ellipse. Associations between actors and use cases are indicated in use case diagrams by solid lines. An association exists whenever an actor is involved with an interaction described by a use case.

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor.

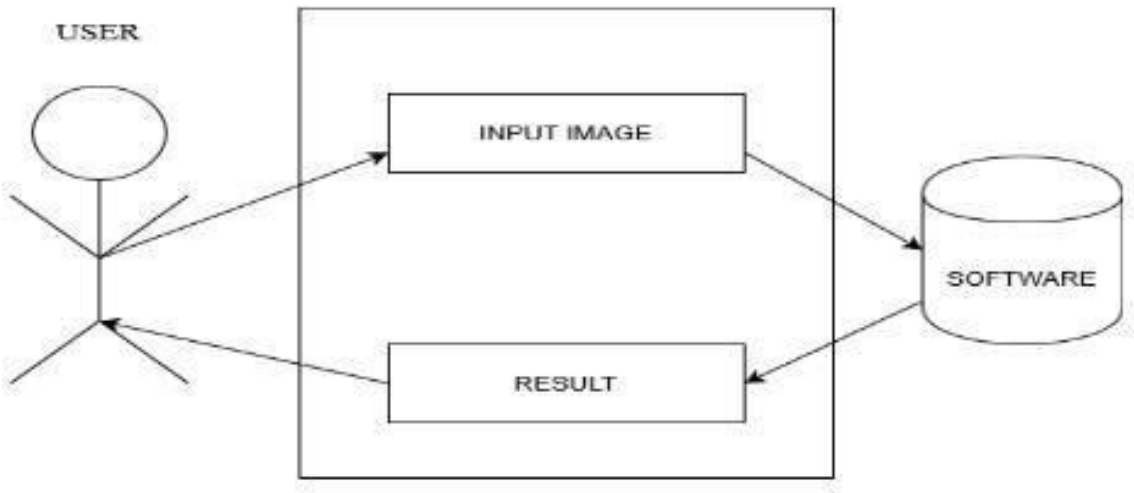

Figure 4.1: Use case Diagram

4.1.3 Flow Chart

A diagram that usually shows the process, workflow or represents a step by step process in solving a task or finding solution to a problem is said to be a flowchart. Flowcharts use boxes and arrows to show the steps taken to solve the problem. These boxes will be connected to each other by using the arrows in the order in which they will be evaluated. Flowcharts are very helpful when it comes to documenting or designing the program or a process. It is also used to analyze a particular problem and find a solution for it in steps.

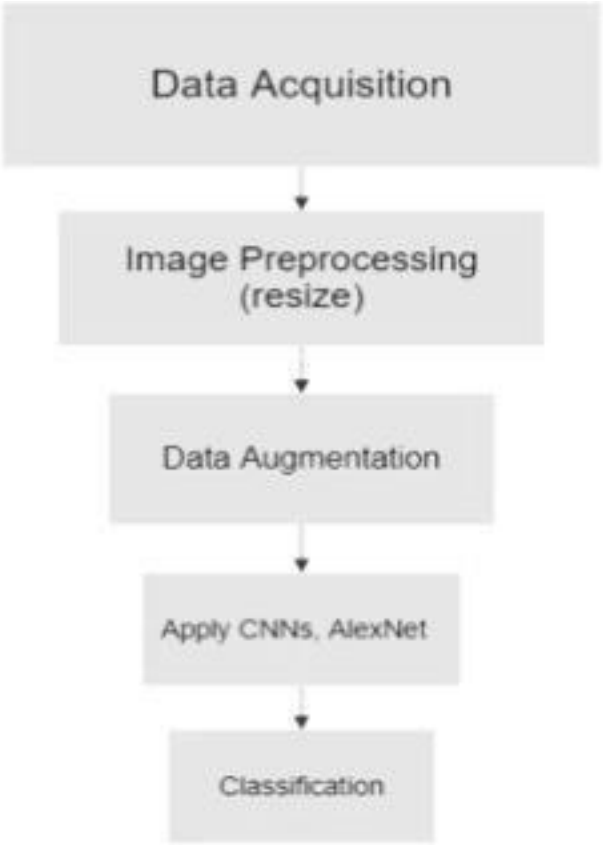

Figure 4.2: Flowchart

4.1.4 Sequence Diagram

A sequence diagram in Unified Modeling Language(UML) is a kind of interaction diagram that shows how processes operate with one another and in what order it is a construct of a Message Sequence Chart. A sequence diagram shows, as parallel vertical lines,different processes or objects that live simultaneously, and, as horizontal arrows. The messages exchanged between them, in the order in which they occur. This allows the specification of simple run-time scenarios in graphical manner in successful transmission. The following diagram shows the sequence of steps involved.

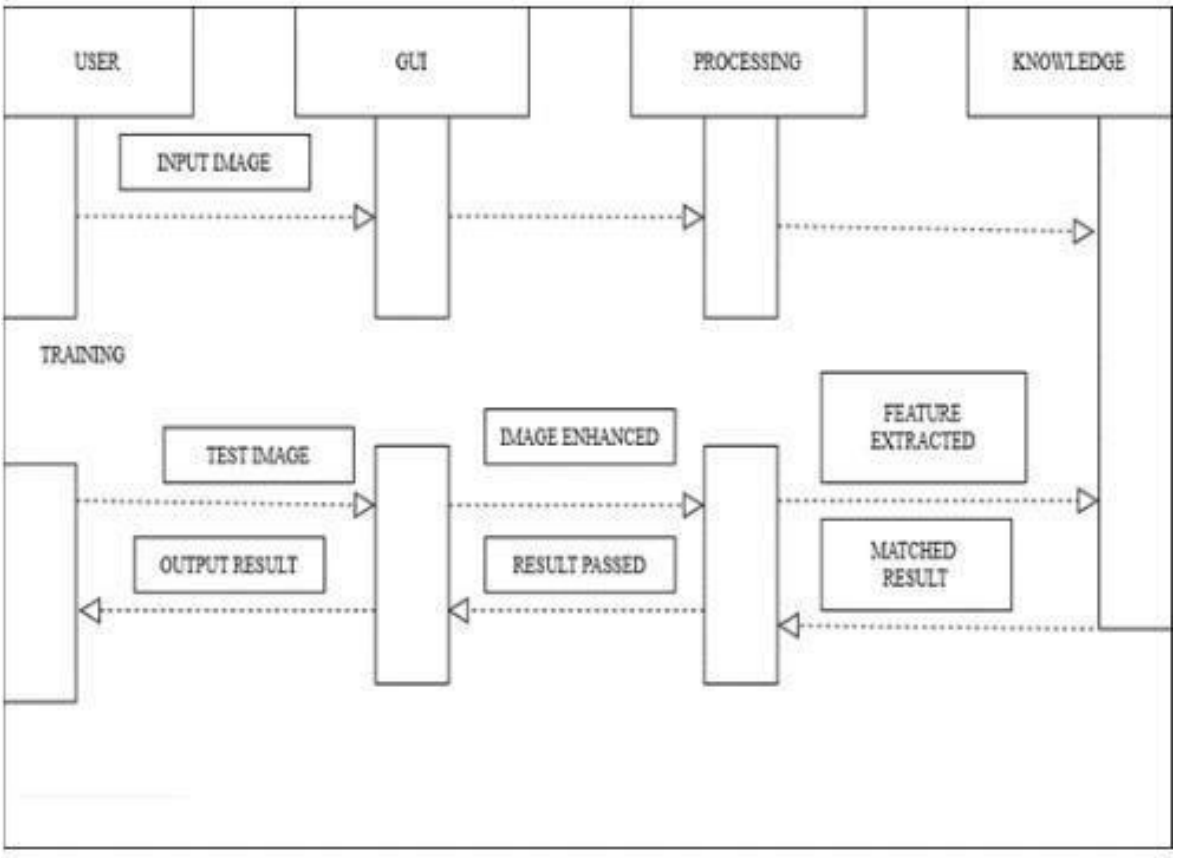

Figure 4.3: Sequence Diagram

4.2 Implementation

The first step of the proposed system is to collect data. Many researchers have used sensors or cameras to capture the hand movements. For our system, we make use of

the web camera to shoot the hand gestures. The images undergo a series of processing operations whereby the backgrounds are detected and eliminated using the colour extraction algorithm HSV(Hue, Saturation, Value). Segmentation is then performed to detect the region of the skin tone. Using the morphological operations, a mask is applied on the images and a series of dilation and erosion using elliptical kernel are executed. With open CV, the images obtained are amended to the same size so there is no difference between images of different gestures. Binary pixels are extracted from each frame, and Convolutional Neural Network is applied for training and classification. The model is then evaluated and the system would then be able to predict the alphabets.

4.2.1 System Architecture

A CNN model is used to extract features from the frames and to predict hand gestures. It is a multilayered feedforward neural network mostly used in image recognition. The architecture of CNN consists of some convolution layers, each comprising of a pooling layer, activation function, and batch normalisation which is optional. It also has a set of fully connected layers. As one of the images moves across the network, it gets reduced in size. This happens as a result of max pooling. The last layer gives us the prediction of the class probabilities.

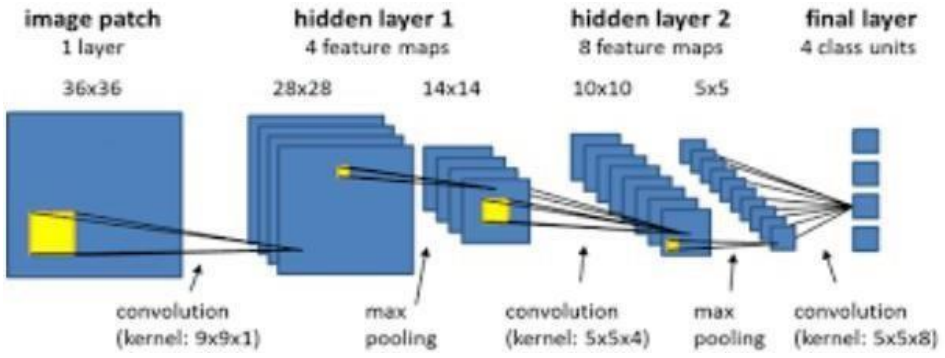

Figure 4.4: CNN Architecture

CHAPTER 5

RESULTS

The system in general requires only two things. Data and model. We create our datasets on our own. This power of creating datasets is given to the users because of pre-stated reasons.
The user chooses a sign then performs it, the system captures images of users hands, preprocess it and use it to train the model. Below figure shows how an image is taken cropped, augmented and made important for our model to be trained.

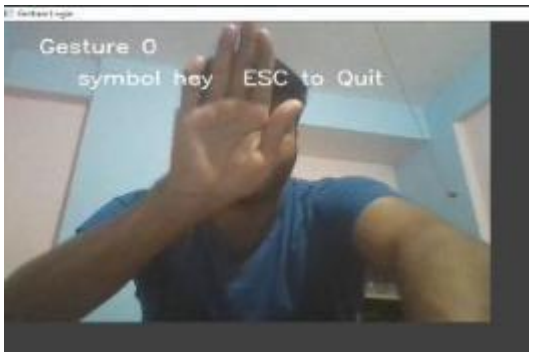

Fig 5.1: The image taken by the system


Figure 5.2: Augmented image ready to be sent to the machine

Once the Dataset is created the images are sent to our algorithm. As any of neural network based algorithm works here also we try to find out optimum weights for the kernel for which we can get highest accuracy. The image is passed through and back to a network of convolution layer and max pooling layers to reduce the loss. Also one

of the key ingredient in creating a accurate model is using a activation layer which suits the job.
For our task we used relu which stands for rectified linier unit. The reason of choosing this activation function is its speed, and lower complexity in implementation.

Below is the comparison of relu and other activation functions.

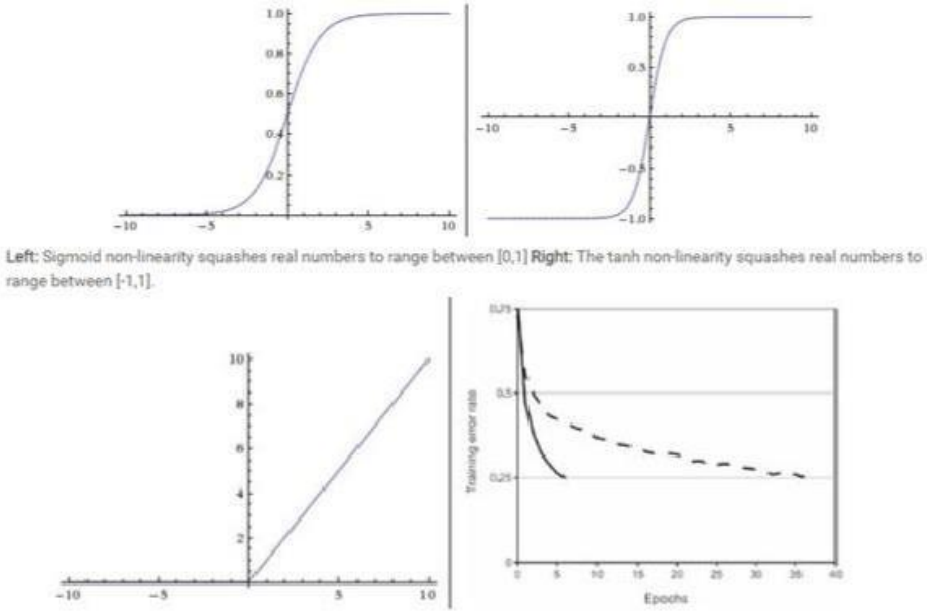

Figure 5.3: graphs of different activation functions

In our project we have created different modules as part of a big project. Below is the list of modules created.

Figure 5.4: signup page

This is the signup page for the users, here users can create a username password combination to create an account.

. Figure 5.5: login page

Here users can login using their user's name and password after which they can show signs to the system.

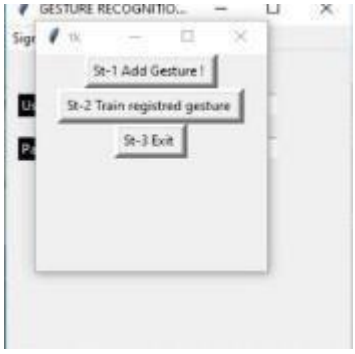

Figure 5.6: menu for registered users

After the user signs up this menu is given to the users. Here user can add gestures to them account and then train the model and then can exit.
The users can then login to their account to start using the model.

CONCLUSION

After testing the model, it was seen that the model gave a very nice accuracy (90%). The system works nicely in real time. The created system can be used by differently abled person to aid their communication. The system also does not take time to show the results i.e., the accuracy and latency of the system is very nice. Multiple improvements can be done for creating a better model. Also, more module can be created which can help us to make it more user friendly. Artificial intelligence can be used to make communication easier for differently abled person. These kind of systems are easy to install and users can easily create their own gestures.

References

1. O. Aran, I. Ari, L. Akarun, B. Sankur , A. Benoit , A. Caplier, P. Campr, A. H. Carrillo, and F. Xavier Fanard, "SignTutor: An Interactive System for Sign Language Tutoring," IEEE feature article, pp. 81-93, 2009.

2. T. Dasgupta, S. Shukla, S. Kumar, S. Diwakar, and A. Basu, "A Multilingual Multimedia Indian Sign Language Dictionary Tool," The 6‟Th Workshop on Asian Language Resources, pp. 57-64, 2008.

3. J. F. Lichtenauer, E. A. Hendriks, and M. J. T. Reinders, "Sign Language Recognition by Combining Statistical DTW and Independent Classification," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, pp. 2040- 2046, 2008.

4. U. Zeshan, "Indo-Pakistani Sign Language Grammar: A Typological Outline," Sign Language Studies, vol. 3, pp. 157-212, 2003.

5. H.D. Yang, S. Sclaroff, and S.W. Lee, "Sign Language Spotting with a Threshold Model Based on Conditional Random Fields," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, pp. 1264-1277, 2009.

6. P. Dreuw, D. Stein, T. Deselaers, D. Rybach, M. Zahedi, J. Bungeroth, and H. Ney, "Spoken Language Processing Techniques for Sign Language Recognition and Translation," Technology and Dissability , vol. 20, pp. 121- 133, 2008.

7. S.F.Wong and R. Cipolla, "Real-time interpretation of hand motions using a sparse bayesianclassifier on motion gradient orientation images," In: Proceedings of the British Machine Vision Conference, vol.1, Oxford, UK (2005), pp. 379-388, 2005.

8. M. Zahedi,D. Keysers and H. Ney, "„Appearance-based recognition of words in american sign language," In: Second Iberian Conference in Pattern Recognition and Image Analysis, vol.1, pp. 511-519, 2005.