

**Sign language**

**Recognition (Hand**

**Gestures)**

**TY B.Tech. Mini Project Report**

**SUBMITTED BY**

**Shantanu Jumde [T1940]**

**Deven Bharati [T1940]**

**Saudagar Saif [T1940]**

**GUIDED BY**

**Prof. MS.**

**Shubhangi.P.Kale**

**SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY,**

**MIT ACADEMY OF ENGINEERING, ALANDI(D), PUNE-412105**

**MAHARASHTRA (INDIA)**

**MAY, 2020**



**Sign language Recognition (Hand Gestures)**

**TY B.Tech. Mini Project Report**

*submitted in partial fulfilment of the*

*requirements for the award of the degree*

*of*

**Bachelor of Technology**

*in*

**COMPUTER ENGINEERING**

**BY**

**SHANTANU JUMDE, DEVEN BHARATI, SAUDAGAR SAIF**

**SCHOOL OF COMPUTER ENGINEERING & TECHNOLOGY**

**MIT ACADEMY OF ENGINEERING, ALANDI(D), PUNE-412105**

**MAHARASHTRA (INDIA)**

**MAY, 2020**



**CERTIFICATE**

It is hereby certified that the work which is being presented in the TY B.Tech. Mini Project Report entitled **“*SIGN LANGUAGE RECOGNITION (HAND GESTURES)*”,** in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Computer Engineering** and submitted to the **School of Computer Engineering and Technology of MIT Academy of Engineering, Alandi(D), Pune, Affiliated to Savitribai Phule Pune University (SPPU), Pune** is an authentic record of work carried out during an Academic Year 2019-2020, under the supervision of **MS. Shubhangi.P.Kale,** **School of Computer Engineering and Technology.**

**Shantanu Jumde PRN No.** **0120170240 Exam Seat No. TY109452**

**Deven Bharati PRN No.** **0120170240 Exam Seat No. TY109452**

**Saudagar Saif PRN No.** **0120170240 Exam Seat No. TY109452**

**Date: 6 May 2020**

|  |  |
| --- | --- |
|  |  |
| Ms. Shubhangi.P.Kale | Mrs. Ranjana Badre |
| School of Computer Engineering and Technology, | School of Computer Engineering and Technology, |
| MIT Academy of Engineering, Alandi(D), Pune | MIT Academy of Engineering, Alandi(D), Pune |

**(STAMP/SEAL)**

|  |  |
| --- | --- |
| *Signature of Internal examiner/s* | *Signature of External examiner/s* |
| *Name………………………………* | *Name………………………………* |
| *Affiliation…………………………* | *Affiliation…………………………* |

**ACKNOWLEDGEMENT**

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals, colleagues, and mentors. We would like to give our sincere thanks to all of them.

We are highly indebted to MS. Shubhangi P. Kale for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

We would like to express my gratitude towards my group members & faculty staff of MIT Academy of Engineering for their kind co-operation and encouragement which help us in completion of this project. Our thanks and appreciations also go to our colleagues, teachers and guides who in developing the project and people who have willingly helped me out with their abilities.

We would be failing in our duty if we do not thank all the other staff and faculty members for their experienced advice and evergreen co-operation

1. Shantanu Jumde sign
2. Deven Bharati sign
3. Saudagar Saif sign

**ABSTRACT**

Sign language interpretation is an important aspect with the increase number of deaf people in India. This report present the application for computerized interpretation of Indian sign language in a text in English to support the conversation and It could be used as a translator for people that do not follow sign language, bypassing by this way the interference of an intermediate person and allow communication using their natural way of speaking.. The sign language image is acquire using inbuilt camera. The analysis of image is performed using our CNN model. We are able to represent one hand sign representation of Indian sign language. The outcomes are found to be extremely uniform, reproducible, with reasonably great precision and efficiency.

|  |  |  |
| --- | --- | --- |
| **LIST OF FIGURES** | | |
| **Fig. No.** | **Fig. Name** | **Page No.** |
| Fig 3.1 | Block diagram | 5 |
| Fig 3.2 | Sequence Diagram | 6 |
| Fig 3.3 | Block diagram | 6 |
| Fig 3.4.1 | Convolutional Layer | 7 |
| Fig 3.4.2 | Backward Propagation | 8 |
| Fig 4.1.1 | Flow Chart | 10 |
| Fig 4.2.1 | Accuracy of the model | 12 |
| Fig 4.2.2 | Model loss | 13 |
| Fig 4.2.3 | Model accuracy | 13 |
| Fig 4.2.4 | Implementation of Trained model (Gesture: Baby) | 14 |
| Fig 4.2.5 | Implementation of Trained model (Gesture: House) | 14 |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **LIST OF TABLES** | | |
| **Table. No.** | **Table. Name** | **Page No.** |
| Table 2.1 | Authors Citation Details | 3 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# CONTENTS

|  |  |  |  |
| --- | --- | --- | --- |
| Acknowledgements | | | Iv |
| Abstract | | | V |
| List of Figures | | | Vi |
| List of Tables | | | vii |
|  | Introduction | | 1 |
|  | 1.1 | Motivation for the project | 1 |
|  | 1.2 | Problem Statement | 2 |
|  | 1.3 | Objectives and Scope | 2 |
|  | Literature Survey | | 3 |
|  | System Design | | 5 |
|  | 3.1 | Block diagram/ Proposed System setup | 5 |
|  | 3.2 | Sequence Diagram | 6 |
|  | 3.3 | Activity Diagram | 6 |
|  | 3.4 | Related mathematical modelling | 9 |
|  | 3.5 | Hardware and Software Requirements | 9 |
|  | Implementation and Results | | 10 |
|  | 4.1 Algorithm and flowcharts | | 11 |
|  | 4.2 Results | | 11 |
|  | 4.3 Discussion | | 15 |
|  | Conclusion and Future scope | | 16 |
| References | | | 17 |

# INTRODUCTION

Sign Language is the language that uses actions and movements to convey the messages. These languages are primarily developed for the deaf and other verbally challenged people. These people use a simultaneous and precise combination of movement of hands, orientation of hands, hand shapes etc. Each nation has its different sign language. We will be focusing on Indian Sign language in this project.

There are many special features present in this language that distinguish it from other Sign Languages. Features like Number Signs, Family Relationships, use of space etc. are really crucial features of this language. Also, it does not have any temporal inflection.

In this project, we aim towards analyzing and recognizing various signs from a database of sign images. Database consists of variety of images with each image clicked in different light condition with different hand orientations. With such a distributed data set, we are able to train our system to good levels and thus obtain good results.

## Motivations

Deafness is third most common disability seen in human beings, currently, there are 366 Million peoples in the world are deaf which is 5% of the world’s population among which there are 34 Million are children, and it is predicted that over 2030 will be 630 Million and by 2050 over 900 million people which is one in every ten people will have disabling hearing loss. According to WHO (2018) data, the pervasiveness of listening impairment in India is nearly 6.3% which is approximately 63 million people undergoing a vital auditory loss. The predicted pervasiveness of adult-onset deafness in our country is about 7.6% and childhood onset deafness is 2%.

Sign language is the single way in which deaf people can interact with the world but as the speaking languages vary in the world the sign language to alters, Sign language differs from region to region, and most of the nation holds their sign language. Sign languages also defined grammar. For example, a nice built question need to be amid the right eyebrow position. While an individual is questioning associated with who, where, what, why, and when, then the eyebrows are put down. If the issue is concerning a yes/no position, eyebrows are kept as it is.

## Problem Statement

In a research it is found that about 1% of the Indian population is deaf and there are limited people who understand the sign language thus, there is no efficient way in which the deaf people can have a communication with other people who dose not know the sign language.

In this project, we are going to build the prediction model based on Convolutional neural network (CNN) which will predict the hand gestures of the sign language.

## Objectives and Scope

1. To ease the communication between the deaf people and the people who do not know the sign language.
2. To accurately predict the sign language.
3. To translate the result in speech (Audio).
4. To have easy to use user interface.

# LITERATURE SURVEY

# Table 2.1 Authors Citation Details

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No | Name of Author | Citation index | i10-index | H-index |
| 1 | Sharmila Gaikwad | 120 | 2 | 20 |
| 2 | Akanksha Shetty | 22 | 2 | 1 |
| 3 | Akshaya Satam | 2739 | 25 | 42 |
| 4 | Rafiqual Zanman | 909 | 13 | 15 |
| 5 | Noor adnan ibraheem | 730 | 7 | 7 |

2.1 Sharmila Gaikwad

RGB based approaches

Author(s) - Sharmila Gaikwad, Akanksha Shetty

Date Added to IEEE Xplore: 2002

Cited by - 400

Publisher: IEEE

Published in: Sixth International Conference on Computer Vision

They ought to use natural methods to create an education program for hand gestures. It reduces the count of features which in thus reduces learning, training, and action. They have used 15 different types and surface color-based segmentation and thus got around 94 percent success of recognition. In an ASL (American Sign Language) they have used 18 terms utilizing a camera sensor to capture the skin color. The two cameras were placed at the head level and the desk level. Determine the hand by extension about a seed point found using complexion intensity. By doing this they were able to achieve a 95.8 percent recognition rate

2.2 Akshaya Satam

Glove based approaches

Author(s) - Akshaya Satam

Date Added to IEEE Xplore: July 2000

Cited by - 1628

Publisher: IEEE

Published in: International Journal of Computer Vision

Herewith the help of an instrument that is hand glove is used to interact with the system. Here they have practice a sensor-enabled hand glove with combined microelectronics to sense movement, agility, and posture of the hands. Thus the data collected by the sensors in the glove is send wirelessly to the computer. In the initial phase the system supports six different gestures then the support for other gestures was also being added. In the other approach, they used the shape of the hand orientation as peculiarities to obtain class by a difference based heuristic. Thus, were able to add vocabulary of more than 250 words.

2.3 Rafiqual Zanman, Noor adnan ibraheem

3D based approaches

Author(s) - Rafiqual Zanman, Noor adnan ibraheem

Date Added to IEEE Xplore: Nov 2019

Cited by - 182

Publisher: Springer

Published in: International Journal of Computer Vision

They used a hybrid approach via a coupling depth map and RGB and using an inbuilt camera for observing hand motion. They determine the palm within the RGB image and acquire and showed in the depth image, and for other residing areas that meet a threshold value and recognize complexion employing a Gaussian-Bayesian Model-based complexion model. due to which they were able to recognize the hand in front of the camera. Once the hands are detected, they use K-means and softmax regression to find the matches from the database.

# SYSTEM DESIGN

|  |
| --- |
| 3.1 Block diagram |
| 3.2 Sequence Diagram |
| 3.3 Activity Diagram |
| 3.4 Related mathematical modelling  Convolutional Neural Network (CNN)- see the image and attempt to recognize different characteristics, patterns and points from the image. verified knowledge you gather, you'd say thing. Roughly you'll divide CNN into two part the convolutional layer which Extracts characteristics of the input and the fully connected dense layer utilizes image values from convolutional layer to offer output. like all other NN it uses forward and backward propagation technique for training.  The Convolutional Layer- It identifies image by comparing pixel value of the image.  Fig.3.4.1 Convolutional Layer  As you'll in above picture that pixel values of the image round the edges of number is different that the remaining portion. If we have a data image denoted as “X” and a filter denoted with “f” then the definition would be,  Z = X \* f  Dimension of output are going to be ((n-f+1) , (n-f+1))  Fully connected layer- The features obtained from Convolutional layer is send to completely connected layer and here the ultimate result's generated, it's nothing but traditional neural network. It can work with only 1D data so, the result obtained from the Convolutional layer which is in 2D is converted into 1D. Here it performs two actions against input data that is a linear transformation and non-linear transformation.  The equation for linear transformation is:  Z = WT.X + b  Here, X is that the input, W is weight, and b is called a bias and may be a constant. Here the matrix of (m, n) denotes m as features and n as number of neurons within the layer  Nonlinear transformation- because the failure of capturing the complex features lead us to the present, this is often done using activation function.  Backward Propagation-  It stands of backward propagation of error, it adjusts the weights and bias according its previous iteration. Proper values of weights and bias will drastically mitigate the errors rates.  Since the error isn't directly hooked into the load matrix, we are using recurrent backpropagation it is passed on to forward till we get determined value.backprop cnn  Fig. 3.4.2 Backward Propagation |
|  |
| 3.8 Hardware and Software Requirements  Hardware Requirements  • 4 GB Ram  • 1 GB Free Space  • Web Cam (5 MP preferable)  Software Installation Requirements  • Python 2.7.13  • OPENCV 2.4.8  • Keras 2.0.2  • Theano 0.9.0 |

# IMPLEMENTATION DETAILS

|  |
| --- |
| 4.1 Algorithm and flowcharts    Fig. 4.1.1 Flow Chart  Each noise in the image or matrix value is mitigated by ignoring other part of image other than hand, thresholding, filtering and user index.  CNN’s ability to recognized or able to interpret the features in images helps in recognizing the image.  Recognition is accomplished by implementing unique convolutions layers on the image by using filters of distinct weights. Unique features can be found by using introducing non-linearity and multiple filters. It is attended using pooling technique, where only the exceptional learning about the unique features are grouped. These techniques are performed in multiple layers.  We use pooling method to reduce dimensionality of the image while retaining the important aspects of the image. The convolutional layer extracts features from input image. After which we try to introduce the non linearity in the network using activation functions. A multilayer perceptron uses features from previous upper layers for classifying the input image into various classes based on training data. |
| 4.2 Results  In our dataset we realized with an accuracy rate of over 90%. Many approaches were tried to call at this regard and would establish a useful program to communicate within the near prospect.  Here the model can be tested based upon the results we get based upon the images we pass to the model. Performance of this model is based on how efficient is our model for that the appropriate dataset As there will always be fluctuation in either the person's hand posture or action trajectory, a more extensive database containing a larger variety of hand images for each sign is required. Note the performance of our method depends and can be improved on each step we have mentioned above. Like data collection can be better. We have successfully implemented the method mentioned and output has been as predicted at the start.    Fig. 4.2.1 Accuracy of the model    Fig.4.2.2 Model loss    Fig. 4.2.3 Model accuracy |
|  |
| Fig.4.2.4 Implementation of Trained model (Gesture: Baby)    Fig.4.2.5 Implementation of Trained model (Gesture: House) |
| 4.3 Discussion  In the literature survey we in RGB based approaches we observe that they have used the different colour based approaches thus to divide the skin color into different segments. Starner et al used only 40 words vocabulary to train the nuralnetwork. In the glove-based approaches they developed the sensor-based gloves then train the model on different orientations of the glove thus when the person wearing the glove make any gesture the sensors detects the motion and send the feedback to the computer. 3D based approaches uses multiple camera sensor where one of which is a normal color sensor while other is TOF depth sensor the data thus collected from multiple sources is send to the trained model thus get more accuracy as compare to other models. |

# CONCLUSION & FUTURE SCOPE

Sign identification is a compact and challenging job. We created a model that could be readily used. Various methods were tried out in this respect and would establish a valuable program to join in the future. ISL is still unexplored as opposed to ASL and additional such sign language practices worldwide. Using other models it can be applied to generate complete sentences.

We were focused on passive gesture only but here we can also need to take in account the active gesture which the facial and body expression also.

**REFERENCES**

[1] Kenn H, Megen F. V. and Sugar R. 2007 A glove-based gesture interface for wearable computing applications. In: 4th Inter-national Forum on Applied Wearable Computing (IFAWC),2007, \pages 1–10

[2] Hernandez-Rebollar J. L, Lindeman R.W. and KyriakopoulosN 2002 A multi-class pattern recognition system for practical ﬁnger spelling translation. In: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ICMI’02, pages 185, Wsahington, DC, USA. IEEE Computer Society

[3] Liang R.H and Ouhyoung M 1998 A realtime continuous gesture recognition system for sign language. In:Proceedings of the Third IEEE international conference on automatic face and gesture recognition,1998, pages 558–567

[4] Saengsri S, Niennattrakul V and Ratanamahatana C 2012Tfrs: Thai ﬁnger-spelling sign language recognition system.In: Second International Conference on Digital InformationAndCommunication Technology and it’s Applications (DICTAP), 2012, pages 457–462

[5] Quek F K and Zhao M 1996 Inductive learning in hand pose recognition. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996, pages 78–83

[6] Starner T, Weaver J and Pentland A 1998 Real-time American sign language recognition using desk and wearable computerbased video. IEEE Trans. Pattern Anal. Mach. Intell. 20(12):1371–1375

[7] Luis-Pérez F E, Trujillo-Romero F and Martínez-Velazco W 2011 Control of a service robot using the mexican sign

[8] Singha J and Das K 2013 Indian sign language recognition using eigen value weighted euclidean distance based classiﬁcation technique. arXiv preprint arXiv:1303.0634

[9] Bhuyan M, Kar M K and Neog D R 2011 Hand pose identiﬁcation from monocular image for sign language recognition.In: 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pages 378–383

[10] Kenn H, Megen F V and Sugar R 2007 A glove-based gesture interface for wearable computing applications. In: 4th International Forum on Applied Wearable Computing (IFAWC),2007, pages 1–10

[11] Hernandez-Rebollar J L, Lindeman R W and Kyriakopoulos N 2002 A multi-class pattern recognition system for practical ﬁnger spelling translation. In: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, ICMI’02, pages 185–, Washington, DC, USA. IEEE Computer Society

[12] Liang R-H and Ouhyoung M 1998 A real-time continuous gesture recognition system for sign language. In: Proceedings of the Third IEEE international conference on automatic face and gesture recognition,1998, pages 558–567

[13] Saengsri S, Niennattrakul V and Ratanamahatana C 2012

Tfrs: Thai ﬁnger-spelling sign language recognition system. In: Second International Conference on Digital Information and Communication Technology and it’s Applications (DIC-TAP), 2012, pages 457–462

[14] Van den Bergh M and Van Gool L 2011 Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In:IEEE Workshop on Applications of Computer Vision (WACV),2011, pages 66–72

[15] Wang F and Zhang C 2007 Feature extraction by maximizing The average neighbourhood margin. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR’07,pages 1–8 language. In: Adv. Soft Comput., pages 419–430. Springer

[16] N.N. Aizenberg Dept. of Cybern., Uzhgorod Univ., Ukraine <https://ieeexplore.ieee.org/abstract/document/274330>