BEETHOVEN - SIGN LANGUAGE RECOGNITION AND TRANSLATION USING MACHINE LEARNING

Aadel Aboobacker

*Department of Computer Science,*

*College of Engineering Thalassery,*

Kannur, Kerala

[aadel999@gmail.com](file:///C:\Users\LENOVO\Downloads\aadel999@gmail.com)

Amal P K

*Department of Computer Science,*

*College of Engineering Thalassery,*

Kannur, Kerala

[pkamalprakash98@gmail.com](mailto:pkamalprakash98@gmail.com)

Abhijith N

*Department of Computer Science,*

*College of Engineering Thalassery,*

Kannur, Kerala

[abhijithnarakath.com](mailto:naivavinay@gmail.com)

Safdar Shahir

*Department of Computer Science,*

*College of Engineering Thalassery,*

Kannur, Kerala

[safdarshahir@gmail.com](mailto:yeduvr@gmail.com)

Nikhil Dharman

*Assistant Professor*

*Department of Computer Science,*

*College of Engineering Thalassery,*

Kannur, Kerala

[nikhildharmanmk@gmail.com](mailto:nikhildharmanmk@gmail.com)

***Abstract*** *-* ***Sign languages are full-fledged natural languages with their own grammar and lexicon. It is currently the most commonly used and effective communication language for hearing and speech impaired people. The foretold being a language which takes some time to absorb, most of us cannot remember or practice the signs as it is, also they do not need it other than communicating with the mentioned community.***

***Under this project, we are dedicated to find means to effectively deal with this communication issue using Machine Learning algorithms to train the path captured using mobile Camera. This dataset is thenceforth utilized to analyse the gestures and output the word. The output of our project includes three potential products including a mobile application, an application program interface and a sign language learning platform.***

***Keywords -*** ***Sign Language, Machine Learning, Motion History Image, Image processing***

# INTRODUCTION

Around 30 million people across the world have speech impairments and must rely on finger-spelling, dactylology or gestures, which poses a hurdle when seeking to speak with non-sign language users. But learning linguistic communication may be a tedious and time taking process. At this age of technology, it's quintessential to form these people feel a part of the society by helping them communicate smoothly. Our project aims at identifying and translating these gestures to assist them communicate.

Gesture recognition and Sign Language recognition has been a well-researched topic for American Sign Language (ASL), but few research works are published regarding Indian Sign Language (ISL). But rather than using high-end technology like gloves or Kinect, we aim to resolve this problem using state of the art computer vision and machine learning algorithms.

ISL is the predominant communication in South Asia, employed by a minimum of several hundred thousand deaf signers (2003). like many sign languages, it's difficult to estimate numbers with any certainty, because the Census of India doesn't list sign languages and most studies have focused on the northern and denser areas- smart cities and not the rural areas where difficulty is the most. The Indian deaf population of 1.1 million is 98 per cent illiterate.

Although they try to provide hearing aids and other necessary technologies, those are largely dysfunctional in an impoverished society. As of 1986, only 2 per cent of deaf children attended school. Deaf schools within the region are overwhelmingly oralist in their approach. Unlike American communication (ASL) and sign languages of European countries, ISL is in rudimentary stage of its development. The Deaf communities of India are still struggling for ISL to comprehend the status of communication as a minority language. Indian Sign Language has not yet been promoted well in schools or colleges although it has been widely accepted and used by the majority in the community.

NCERT in March 2006 launched a category III text includes a chapter on communication, emphasising the particular fact that sign languages should be considered a language of communication like any other languages. Our aim is to fulfil these words and make this "yet another language" accessible to common people who can't afford their time into learning it. The figure below shows the notations of English Alphabet in single handed sign language.

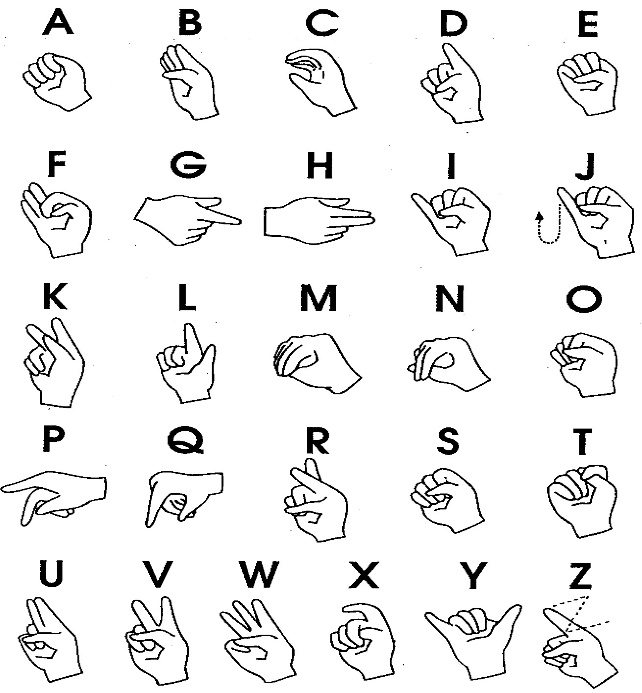


Fig1. ISL Notations for English Alphabet

Technology being more sophisticated day by day, there emerges different methods of interaction and communication with the computers which includes keyboard, mouse and thereafter joysticks, track pads, electromechanical gloves, and more. Aside from these methods, gesture recognition has also been used and this will be considered as a more natural mode of interaction since it mimics normal conversation with an individual's being. Therefore, it provides an honest interface for human computer interaction (HCI).

Gestures are often identified as a meaningful body motion which might involve hands, head, body to convey a meaningful information or to interact with the environment. These gestures are often 2 static, dynamic or both. In dynamic gesture recognition, it's required to spot both spatial and temporal movements, and this paper proposes a strategy for dynamic gesture recognition. In overall, gesture recognition is often divided into two broad areas and people are vision-based methods and methods require special hardware like gloves, armbands and body kits.

The second type is relatively difficult for the user since it's required to carry required hardware when the user must use the system. These kinds of systems could be required for specialized applications like Unmanned Arial Vehicle (UAV) control, but not for a general someone. On the opposite hand, vision-based systems use techniques like image processing, pattern recognition, clustering, image extraction, image segmentation and object detection. Using vision-based approaches provides a non-obstructive interface for the user, and thus created a lift during this area.

1. MACHINE LEARNING

As an application of Artificial Intelligence, Machine Learning provides systems the ability to automatically learn, adapt and improve from experience, upon a set of tasks, without being explicitly programmed.

Machine learning emphasizes on the buildout of computer programs that can access data and process it for themselves. The Learning procedure begins with observations, monitoring, surveying or data input, such as instances, direct events, or certain instructions, in order to search for patterns in data and make better conclusions in the future based on the instances that we provide. The primary aim is to allow the computers learn on their own without human intercession or aid and regulate processes suitably.

1. DATA COLLECTION

The data required for training includes 26 letters of English lexicon, 0-9 digits, 45+ gestures. Signs without motion and with motion are collected separately. Sample data was collected with the help of our classmates and family members.

One of the important questions raised here was why shouldn’t we use sensor gloves for data collection. While designing the glove there are some circuits and some sensors are fixed on glove which detects every single motion made by hand or by even fingers. These instrumented gloves design and sensors used in it can be differ according to use or by companies made it. That’s why it is not necessary every glove work same as other. In this glove there are some sensor whose working is Light based. In light based a tube is used which is flexible so that it can cooperate with hand movement in placed at one side and on the other side photocell is used. As we move or make movement by finger the light that strike on photocell changed, by this finger flexion could be measured. While using Instrumented gloved there are many difficulties faced by user like user have to always wear that glove while collecting the hand gestures. It is very difficult to give same posture same again and again because angle of hands and finger could be varying each time which make input gestures two different gestures.

Due to the above-mentioned limitations in glove-based data collection, Computer Vision Based (19) came into existence. The number of cameras and where to place them are the main two challenges in computer vision-based gesture recognition system. As we have to place cameras such that each and every movement of hand could be register.

Main problem that encounter in this type of approach that hand movement is not visible by cameras and they also fail to record hand movement so this is primary concern in this method. For Data Collection, we would like to visit a school/community for speech impaired in order to collect flawless data .We shall aim at around 15 students of different skin tones and make around 6-11 seconds of video for every alphabet per person using a 30fps camera summing up to 1 minute video for every alphabet.

While making the video the camera shall be tilted accordingly to obtain images in different positions within the frame. We would ask the students to show the most commonly used static representation for every alphabet (in case of multiple representations). We shall try to make the background as light as possible to make image segmentation easier. Later these videos will be converted into frames evaluating to 1000-1500 images per alphabet.

1. DESIGN AND METHODOLGY

Fig.1. System Architecture

## *Front-End*

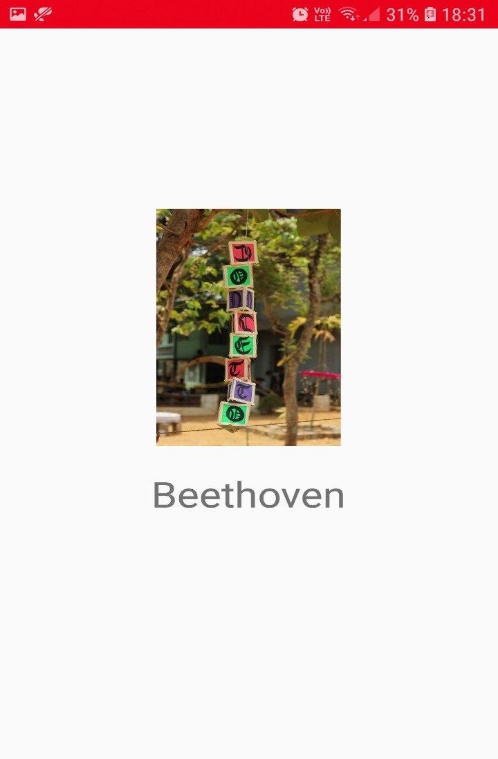


Fig2: Splash Screen of the android application

Figure 2 shows the front-end splash screen of the android application. The application was coded and built using Android Studio and Java. The code is uploaded onto our project GitHub repository.

Android Studio is the perfect option for front end designing. It is the official IDE of Google’s Android platform built on JetBrains’ IntelliJ IDEA software, designed specifically for android development and deployment. It is free and open platform as it is built on Linux.

Our application consists of a video capture and a text view that translates the sign that is captured using the video camera. Figure 3 shows the inner application view of the android application.

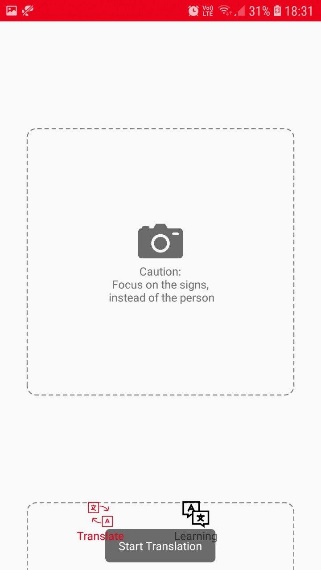
 

Fig 3: Android Application before granting camera permissions and after.

1. *Overview/Working*

Figure 4 shows the system overview of Beethoven. It is designed in such a way that when a user captures a sign language, live, the translated text shall appear at the text view and even an audio. We could also add a normal language translator so that the English language can be translated and can be understood by any user. When user points the camera to the visual (sign language), the camera captures it and the captured video is converted into a number of frames. These frames go through a handful of pre-processing steps.

When compared with the trained weights, if the system recognizes any frames, it returns the label. The gestures such as ‘thank you’, ‘Bye’ etc. are classified using Motion History Imaging.

This text could be enhanced to speech or translated to another language. We would add provisions to these and also to a teaching platform where people can learn sign languages and the people of our emphasising community can learn things in sign language.

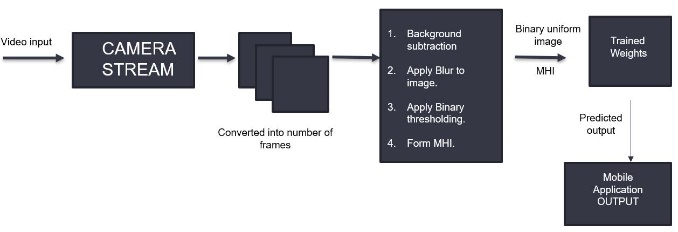


Fig. 4.System Overview

1. *System Features*
2. *Real Time Translation:* The live captured video is converted into sign language. The captured video is not stored and hence it won’t be eating up much memory
3. *Personal side language addition:* The user can also train himself the side langauges (which are locally recognised signatures) and can be locally stored for use.
4. *Text storage:* Stores only the text generated and not the image captured
5. *System Methodology*

Following methodology represents how computer vision based techniques, as well as CNN based classification techniques, have been used to develop a real-time dynamic gesture recognition system. Implementations of this approach have been done using python programming language and to ease the life of image processing, OpenCV (Open Source Computer Vision Library), which was initially developed by Intel has been used. For the machine learning tasks and to retain efficiency in large matrix manipulation tasks, python scikit-learn and NumPy libraries have been used. Images are captured using the inbuilt camera in MacBook Pro laptop in 29 frames per second with an original resolution of 1280 x 720 and later reduced to ease computation. Prior to creating motion history image from captured frames, each frame is preprocessed using multiple image processing steps. Following are those steps and rationale for them.



Fig 5: After applying Binary Thresholding

### *Background Subtraction:* It is crucial to identify moving objects of the scene since moving gestures are going to be identified. Z. Zivkovic suggested a method for this and it is based on Gaussian Mixture-based Background/Foreground Segmentation where it analyzes each pixel and he has developed an effective adaptive algorithm using Gaussian mixture probability density.Zivkovic proposed another improved version in 2006 and these two papers are based on the algorithm which has been used eliminate the background features. This algorithm is implemented in OpenCV library.

### *Apply Blur:* Blurring is done to eliminate noise.A median blur has been applied with a kernel size of 11 to smoth the image, whose background is recently subtracted.

### *Binary Thresholding:* After blurring there might be areas which doesn’t belong to either of black or white and in order to convert those areas to white and to get a consistent binary image, binary thresholding is applied. In addition, inversion is applied on the colors to get a black hand on a white background. But with the different illumination conditions, determining this value can be difficult. Therefore, Otsu’s binarization has been used and it will calculate a suitable threshold value by analyzing image histogram.

1. Motion History Images

The motion history image (MHI) is a stable image pattern or template that helps in understanding the location of motion and trail or the path as it progresses. The temporal motion information is caved into a single image template where intensity is a function of newest motion. Thus, the MHI pixel intensity is a function of the motion history at that area at a time, where brighter values correspond to a new motion. Using MHI, unstable or dynamic portions of a video sequence can be imprinted into a single image, from where one can anticipate the motion flow as well as the dynamic portions of the video action.

MHI detects the shift in an object's state based on the time stamps of pixels over a set of images corresponding to the sequence of motion. MHI gauges a single static and bidimensional outline that integrates both the spatial location as well as the temporal history of motion in the object and thus the spatial and temporal resolutions gets saved. The bidimensional MHI image can be built according to the algorithm that each pixel in the MHI is defined with its own timestamp and will cease if it does not present significant change following a certain time lapse, known as the lifetime of the MHI. In the MHI, the newest movement appears as bright in the gray scale.

# CONCLUSION

In this paper, we are proposing a machine learning based real-time sign language and dynamic gesture recognition system. Proposed system generates motion history images based on the structural deviation in the captured frame along with still lexicon images and this provides a notion of motion occurred in spatially and temporally. Detection of a gesture occurred or not identified using similarity measure and when motion history image diverts below a given threshold, this particular motion history image is classified using a neural network along with the English alphabet. The neural network is used to determine the probabilities for each category and if the category with the highest probability surpasses probability of 0.8, it is identified as a correct gesture. The future scope includes an application program interface for various video conference software and a peer to peer sign language learning and validating platform.

1. REFERENCES

1. Anil Kumar, D., S. S, P. Kishore, K. Eepuri, and T. Maddala (2018). S3drgf: Spatial

3d relational geometric features for 3d sign language representation and recognition.

IEEE Signal Processing Letters, 26, 1–1.

2. El-Bendary, N., H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu

(2010). Arslat: Arabic sign language alphabets translator, 590–595.

3. Khan, R. Z. and N. Ibraheem (2012). Hand gesture recognition: A literature review.

International Journal of Artificial Intelligence Applications (IJAIA), 3, 161–174.

4. Lu, W., Z. Tong, and J. Chu (2016). Dynamic hand gesture recognition with leap

motion controller. IEEE Signal Processing Letters, 23(9), 1188–1192. ISSN 1558-

2361.

5. Melnyk, M., V. Shadrova, and B. Karwatsky (2014). Towards computer assisted

international sign language recognition system: a systematic survey. Int. J. Comput.

Appl, 89(17), 44–51.

6. Naguri, C. R. and R. C. Bunescu (2017). Recognition of dynamic hand gestures from

3d motion data using lstm and cnn architectures, 1130–1133. ISSN null.

7. Potter, L. E., J. Araullo, and L. Carter, The leap motion controller: a view on sign

language. In Proceedings of the 25th Australian computer-human interaction conference:

augmentation, application, innovation, collaboration. ACM, 2013.