

Sign Language Interpreter

Mr. R. Augustian Isaac¹, S. Sri Gayathri²

¹Assistant Professor, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

²B. Tech, Dept. of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

Abstract—Sign language, a language that involves hand gestures, postures and body movements, has been the first and only language of millions of people for ages. But in order for them to communicate with others, the other party is also required to know their language. In this rapidly changing world, the technological advancements pull everyone into the digital lifestyle. Thus, giving rise to the problem of communication through the internet, for the deaf and mute people, who are familiar only with the sign language they know. Here comes the necessity to build a sign language recognition system, through which the computer can be made to recognize and interpret sign language and translate it for the required task. This can be done using the concepts of machine learning and human computer interaction (HCI).

Key words—Human Computer Interaction, Hidden Markov Models, Dataset, Hand gestures, Camera capture, Artificial Neural Networks.

1. INTRODUCTION

There are more than 70 million people who are hearing or speech impaired. Sign languages help the deaf and mute community communicate using postures or body movements, eyes, eyebrows, hand gestures, that are used in different combinations to differentiate among lexical distinction, structure of grammar, etc. Thus, it's evident that visual observation is essential for recognizing sign language.

Recognizing sign language by human interpreters might not be a very challenging task as all that's required is for the interpreter to learn the particular sign language. But gone are the days where people could live without the influence of digitalization. With the exponential growth in technology, even the day-to-day tasks require the use of internet, thus making it a compulsion for people to stay aware and updated. But not everyone is able to deal with this, for instance, the deaf and mute community. And machine, unlike humans, cannot recognize sign language as easily as humans do. This is because a visual interface is needed for the computer to process and interpret it.

This project aims at solving the above mentioned problem by building a sign language interpreter. This interpreter enables a computer to recognize the signs used in American Sign Language (ASL) and convert the interpreted data to text for further use. Gesture recognition is done using camera capture, where the input given are the images of signs/gestures of the language. The processing is done using Human Computer Interaction (HCI). Human computer interaction is the discipline that mainly focuses on the interaction between people and computers. It is an up-and-

coming concept in the field of computer science. It paves way for many new inventions capable of making human lives easier to live.

2. EXISTING SOLUTION

Sensor based approach:

Sensor based approach has two types: glove sensors and built-in sensors (internal sensors built in the particular device). The existing system for sign language recognition involves the use of glove sensors. It is used for obtaining the input for recognizing the signs given by the user. These data gloves are also known as accelerometer gloves.

Sensor gloves refer to the gloves that have multiple sensors embedded in them. These track the location and motion of the fingers and the palm. They have been in use for many years now but for other purposes such as creation of virtual 3-dimensional environments. Its applications include gaming, commands to robots, etc. One big advantage of sensor gloves is that the sensing is not affected by external disturbances due to change in light, magnetic field or electric field.

The major disadvantage of the existing system is the complexity of using sensors for gesture recognition. Kinetic User Interface (KUI) is an interface that lets the user to interact with the computer by sensing the motion of the object under consideration. The user is expected to be wearing the gloves each time he wants to give an input. These gloves can be quite expensive and difficult to use. Thus, in this project, an alternative way is being suggested.

3. LITERATURE SURVEY

To develop a sign language recognition system, a number of methods have been used previously. Based on the literature survey conducted, the techniques used in the previously conducted experiments have been identified. Some of the methods found are as follows:

3.1 Support Vector Machine:

Support Vector Machine is a technique used for pattern recognition. It comes under supervised learning. This technique is known to deal with unknown data really well as it operates by dividing the feature space for each class. The supporting vector machine, or SVM, gets a set of input data and for each set, it predicts or determines which of the two classes can bring the output.

3.2 Human Computer Interaction:

Human Computer interaction, in short, HCI, is a concept of study that works on the interface between people and computers. It is an emerging topic and is really popular as its main aim is to make the computer understand humans more and more better. It overlaps with many topics such as behavioral sciences, media studies, computer science, etc. There are various terms with which HCI is referred as. Some of which are: HMI (Human-Machine Interaction), CHI (Computer-Human Interaction), MMI (Man-Machine Interaction).

3.3 Virtual Reality Modeling Language:

Virtual reality modeling language, also known as virtual reality markup language, is a programming language that is was developed in order to create 3-dimensional models and web-based models. It illustrates 3-dimensional objects and other objects that require a 3-dimensional structure. It is also considered to be similar to HTML (Hypertext Markup Language). The main use of this in sign language recognition system is its ability to define 3-dimensional credentials such as the coordinates and geometric values.

3.4 Bag of Features:

Bag of features is a model under Computer Vision that is used for image classification. Bag of features contains four main processes which are:

- 1) Extraction of features
- 2) Learning the "visual vocabulary"
- 3) Quantization of the features using the learnt visual vocabulary.
- 4) Representation of the images with the help of the frequency of "visual words".

But there are certain disadvantages in using this technique. These are the problems associated with internal parameters, camera position and illumination.

3.5 Particle Swarm Optimization:

Particle Swarm Optimization is a computational method used in the field of computational science, where a problem is optimized by iteratively trying to improve a candidate solution according to a given quality measure. It works by having a number of candidate solutions, known as dubbed particles, around which it moves in the search space, with simple mathematical formula applied on the particle's velocity, position, etc. This algorithm, combined with neural networks, produces better output than many others as many performance measures such as accuracy, precision, kappa statistic, F-measure where all calculated and the result was found to be accurate than many other models.

3.6 Genetic Algorithm:

Genetic algorithm is a sub-field of evolutionary computation. The mentioned parent field consists of algorithms used for global optimization. Evolutionary computational techniques are known to generate solutions that are highly optimized. Genetic algorithm can be combined with Particle Swarm Optimization, Evolutionary Computation and Neural Networks in building a sign language recognizing system.

From the previously-used techniques identified through the literature survey, it was found that there are certain disadvantages in the techniques used. Some of them are as follows:

- Hidden Markov Model is mainly intended for speech recognition.
- It is expected that the Graphical User Interface may shift to spoken human language interaction as it places severe limitations on HCI's complexity.
- The method using Artificial Neural Networks might not yield effective or accurate results.
- The accuracy of an algorithm may vary greatly with the number of instances in various classes.

4. PROPOSED SOLUTION

The major reason to develop this system is to make communication through the internet easier for the deaf and mute community. Thus, the sensor-based approach might not be very feasible due to its complexity and difficulty to use it. Compared to this, a video-based approach might prove to be more effective and easier to use. This project uses static hand gesture recognition as the input mode.

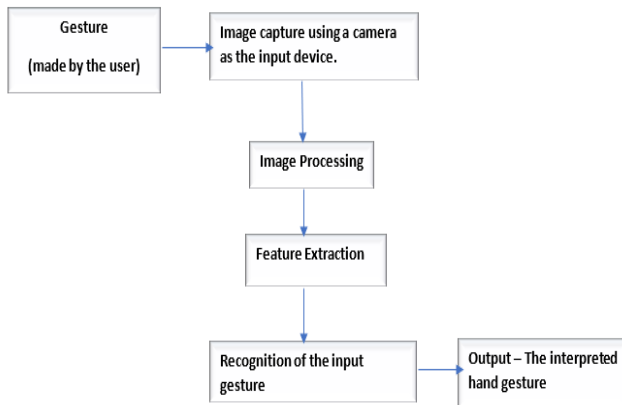
In this approach, camera capture is used to capture a sequence of images, that is, the signs made by the user. The image is then compared with the dataset pre-built and the sign is identified and the output is given accordingly. The feature extraction is done by knowing the position of the hand and its orientation. There are skin segmentation algorithms that are used widely in Computer Vision applications. In this project, there are different methods and algorithms used to extract the details of the hand such as:

- the area of the palm, which is obtained by finding the largest circle that can fit in it,
- detecting the convex hull
- The number of open fingers, which is measured by finding contours.

Machine learning, which is like the future of computer science, is a field that deals with providing computers the ability to act without being programmed explicitly. This project is entirely based on machine learning. The programming language used is Python as python is a language that comes with features to facilitate visualization and data analysis.

5. SYSTEM ARCHITECTURE

The system architecture of this project depicts the flow of control. The hardware device required is the input device to obtain the image of the gesture. In this case, a camera or a webcam is used as the input device. The architecture diagram is as follows:



The gesture made by the user is received through the camera. The image is then processed for extraction of its features. The collected data is then used for further calculations. It is compared with the existing data set and recognized. The interpreted gesture is then given as the output and then used for the required purpose.

6. MODULE IDENTIFICATION

The working of this approach is carried out with the help of certain modules which are as follows:

6.1 Data Set:

In supervised type of machine learning this is the very first step for image recognition. This is because supervised learning generally refers to making the machine learn a function that would map the given input to the output. In this step, a set of images for each letter in the sign language is fed to a database. The number of images may vary from 100 to 200, with different angles of each particular gesture included. The input obtained is then compared with the given images in the dataset to identify the gesture made. The reason for the number of images in the dataset is to get the output with a good amount of accuracy and also to avoid ambiguity, which has high chances of occurring in sign language as one gesture might be similar to another one. Thus, the data set can be considered as the fundamental need in supervised type of machine learning.

6.2 Image detection:

This is the step that comes right after camera capture. Image detection refers to detecting the image that is obtained and, in this case, it is found out if the obtained image is that of a hand or not. A binary classifier is to be trained beforehand to check the same. A binary classifier has the task of classifying

sets into two groups, depending on the criteria the sets meet. It checks for one or more qualities that a particular set should possess. It is according to that factor that a binary classifier decided to which group the set should be sent to.

6.3 Feature Extraction:

Feature extraction refers to extracting the details from the image captured. In a sign language interpreter, the image captured is a gesture made by a hand. Therefore, the features extracted from such images include the size of the palm, the number of fingers open, etc. These features are then used to recognize the gesture using certain algorithms.

6.4 Image Recognition:

Image recognition is the most crucial procedure of this project. The acquired image is converted to its vector form. The model used is SVM, support vector machine. Support Vector Machine is used to analyze data and classify them. Support Vector Machine comes under supervised machine learning. SVM represents the examples as points in space that are mapped so that they are separated according to the category they come under.

6.5 Output:

All of these modules contribute to the successful working of sign language interpreting system.

The flow of execution takes place in the following manner: The camera gets the input gesture image from the user, the detection process takes place to check if it is a hand or not using certain algorithms, image recognition is the next step where the image acquired from the user is compared with the images in the dataset, to interpret the shown gesture. Image recognition is done using a model known as SVM or Support Vector Machine. The next step is the output where the recognized symbol is converted to text form as the output.

7. CONCLUSIONS

This paper describes an approach for recognizing the American Sign Language (ASL), that is, to provide an interface for the deaf and mute population to communicate through the internet. This system is based on machine learning. This system takes into consideration the various drawbacks in the existing system and also their advantages for better working of this system. It is built to make communication over the internet easier for the deaf and mute community. But this also has certain limitations. The interpreter uses only static gesture recognition and not dynamic. Dynamic gesture recognition refers to identifying gestures in a continuous sequence in the form of videos. A system implementing dynamic gesture recognition is much more complex and is yet to be built.

ACKNOWLEDGEMENT

This project was supported by Prof. Mr. R. Augustian Isaac, thus I would like to express special thanks to him, for giving me the opportunity to work on this project. During the research on this project, many new terms came to be known for which I am thankful.

REFERENCES

- [1] Efthimiou, E., Fotinea, S.E., Vogler, C., Hanke, T., Glauert, J., Bowden, R., Braffort, A., Collet, C., Maragos, P. and Segouat, J., 2009, July. Sign language recognition, generation, and modelling: a research effort with applications in deaf communication. In International Conference on Universal Access in Human-Computer Interaction (pp. 21-30). Springer, Berlin, Heidelberg.
- [2] Gupta, P., Agrawal, A.K. and Fatima, S., Sign Language Problem And Solutions For Deaf And Dumb People.
- [3] Parton, B.S., 2005. Sign language recognition and translation: A multidisciplinary approach from the field of artificial intelligence. *Journal of deaf studies and deaf education*, 11(1), pp.94-101.
- [4] Nagarajan, S. and Subashini, T.S., 2013. Static hand gesture recognition for sign language alphabets using edge-oriented histogram and multi class SVM. *International Journal of Computer Applications*, 82(4).
- [5] Uddin, J., Arko, F.N., Tabassum, N., Trisha, T.R. and Ahmed, F., 2017, December. Bangla sign language interpretation using bag of features and Support Vector Machine. In *Electrical Information and Communication Technology (EICT), 2017 3rd International Conference on* (pp. 1-4). IEEE.
- [6] Hore, S., Chatterjee, S., Santhi, V., Dey, N., Ashour, A.S., Balas, V.E. and Shi, F., 2017. Indian sign language recognition using optimized neural networks. In *Information Technology and Intelligent Transportation Systems* (pp. 553-563). Springer, Cham.