A Conversion from Sign Language to Text & Speech using Machine Learning

Prof. Manish Assudani [1] , Syed Ehteshamuddin [2], Himanshu Hatwar [3] ,Maaz Khan [4] , Mohammad Adnan [5]

[1] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[2] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[3] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[3] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India
[3] *Department of Computer Science & Engineering*
Anjuman College Of Engineering & Technology
Nagpur, India

***Abstract*— Language is the medium through which people communicate, which is a common need for humans. Most of the people are able to speak and listen, and they communicate using various languages. The use of signals by people with hearing impairments is common. Those who have hearing loss in America prefer to communicate using American Sign Language (ASL). It becomes essential to use PC-based ASL translators to make conversation between hearing-impaired and hearing people easier and less complicated. This project aims to develop a vision-based application that provides sign language to text translation, facilitating conversation between signers and non-signers. Video sequences are taken in consideration for the suggested model to extract material and contiguous features. Then, to identify spatial characteristics, we employ Inception, a CNN (Convolutional Neural Network). Then, to learn on temporal features, we employ an RNN (Recurrent Neural Network). The American sign language dataset was utilized.**

Keywords—Communication, Hand gestures, Sign language, ANN, CNN

I. INTRODUCTION

Language serves as the medium for communication, which provides rudimentary human needs. Large amount of people speaks and listens too, and they frequently converse with each other in a variety of languages, including Swedish, English and many more. People with hearing loss use signs to communicate. The preferred language of the deaf in America is American sign language (ASL) . Yet, connecting with deaf persons might be challenging for non-signers who do not know sign language. To help deaf and hearing people communicate, ASL interpreters are necessary. Without an interpreter, communication between the hearing and the deaf may be compromised or impossible. A system that can translate ASL into text and speech in this case will be very beneficial. To bridge the gap between signers and non-signers, we are working on creating a model that can translate ASL into both text and speech in English. The primary aim is to assist a sizable number of hearing-impaired people and unite them into society. The system consists of 5 modules dataset, capture, learning, detection and interface. The detection module, learning module, and capture module are the key modules that builds the system. The detection module is for word detection and the other is for alphabet detection. The capture module contributes fresh data to dataset, while the learning module is for interactive ASL learning.

II. OBJECTIVES

Image processing and data Glove are two well-known methods used in sign language research. There have been efforts made around the globe to help the deaf community communicate with non-signers. Maintaining the Integrity of the Specifications Nicholas Born[2] suggests a research-based creation of a data glove-based sign language recognition system. This method uses a sensor glove with an accelerometer and flex sensor to detect the hand gesture. The objective of this system are mentioned below.

1. Introduce an automated method that will reduce the cost of sign language interpretating without the need of specialized sensors.
2. To provide an interactive learning for ASL.
3. To improve the overall performance of the system.

III. LITERATURE SURVEY

Image processing and data gloves are two rudimentary modus operandi. The fundamental disadvantage of the data gloves is the requirement for specialized hardware, like as gloves, a jump motion sensor, or a depth sensor (such as Microsoft Kinect), which is pricy and frequently unavailable in real-life situations. The fact that these sensors require a closed environment—for example, the Kinect sensor cannot function correctly in sunlight—is another issue. There have been efforts made to bridge the communication gap between signers and non-signers. Many of the acknowledged solutions are as follows.

- Nicholas Born has proposed a research-based creation of a data glove-based sign language recognition system .This method consist of a sensor glove made of a flex sensor and an accelerometer which detects the presence of a hand [2].
- Another approach was made by Rini Akmeliawatil, Melanie PO-Leen Ooi, et al,". which proposed Real Time Malaysian Sign Language Translation using Color Segmentation(compares the color feature of each pixel with the color features of surrounding pixels) and Neural Network[1]. Instrumentation and Measurement Technology Conference Warsaw, Poland.IEEE.1-6, 2007.
- The K-NEAREAST NEIGHBOUR algorithm is one of the simple, non-parametric and supervised machine learning algorithms. which makes use of proximity(distance measure) to classify and predict the group of data points. A KNN Research Article Classification Technique Based on Shared Nearest Neighbor was published in the June 2010 proceedings of the NTCIR-8 Workshop Conference in Tokyo, Japan by Yun-lei Cai et al [6].
- The support vector machine is one of the supervised machine learning algorithms. The goal of this algorithm is to determine the best decision boundaries in n-spaces. This algorithm uses extreme points called as vectors which are close to the decision boundaries. Avinash Navlani, proposed the approach to "Support Vector Machines with Scikit-learn", available at https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python, 12 July 2018 [4].
- The other work is proposed by Zhe Cao, Yaser Sheikh, et al, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv:1611.08050v2 [cs.CV], Apr 2017 [3]. The method learns to associate body parts with people in the image using nonparametric representations called Part Affinity Fields (PAFs).
- The biological neural networks that shape the structure of the human brain are where the phrase "artificial neural network" originates. Artificial neural networks also feature neurons that are interconnected to one another in different levels of the networks, much like the human brain, which has neurons that are interconnected to one another. Nodes are the name for these neurons. International Journal on Current and Innovative Trends in Computing and Communication, Volume: 2 Issue: 1, Sonali B. Maind et al., "Research Paper on Basic of Artificial Neural Network," 2014 [7].
- TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, Preliminary White Paper, Mart n Abadi, Ashish Agarwal, et al., November 9, 2015. TensorFlow is an implementation for machine learning algorithms as well as an interface for defining such methods [8].

IV. PROPOSED SYSTEM

Based on the calculated human skeletal important points from independent library OpenPose, we suggest an ASL recognition method. An open source toolkit called OpenPose can estimate 130 key points in total, of which 18 key points are from the body, 21 key points are from each hand, and 70 key points are from the face. There isn't currently a publicly accessible dataset for ASL, particularly one that can be implemented using OpenPose. The sign language dataset is more challenging to compile than other datasets because it requires experienced signers to capture high-quality sign language movies, which are difficult to come by. We have therefore made the decision to create our own dataset of ASL words and alphabets. For the purpose of recognizing signs, various machine learning algorithms are applied and their performance is evaluated. Moreover, libraries for machine learning like TensorFlow, Scikit-Learn, and Keras are used. Because it exclusively identifies the human body, our recognition system is resilient of various backgrounds. Given that the extracted key points may have low degree of variation, the system based on human key point identification performs effectively in spite of the signers. The standard deviation and mean are also included to standardize the feature collection. Last but not least, when the range of dataset is insufficient, high-level of features are required to use.

V. Modules

1. Dataset

The term dataset is defined as the group of data which stores particular information. The dataset of ASL(American sign language) contains collection of all the English alphabets from "A-Z". The American Sign Language dataset is an object detection dataset of each ASL letter.

2. Capture

Algorithms like support vector machine(SVM), K-nearest neighbor(KNN), artificial neural networks(ANN), convolutional neural networks(CNN) are the algorithms which will work on the dataset and will convert the hand sign language into text and speech.

3. Detection

In the background, there is already a sign detecting algorithm. The system will process the picture with the aid of various machine learning algorithms and look for hand motions.

4. Interface

Interactive and user friendly interface is provided for capturing, detecting and learning ASL language and thus, the system will convert hand sign language into text and speech.

VI. Algorithms

1. SVM

The support vector machine is one of the supervised machine learning algorithms. Which aims to find the best decision boundaries in given n-spaces. It uses extreme points known as support points or vectors. These extreme points affect the position of the decision boundary. Hava Siegelmann and Vladimir Vapnik invented the support vector clustering algorithm, which uses support vector statistics from the support vector machines algorithm to classify unlabeled data. These data sets call for unsupervised learning methodologies that look for natural groupings in the data and then apply new data to these groups.

2. KNN

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbor. KNN distinguish between the new cases/dataset with the available dataset and put the new case into the appropriate category of the available dataset. A new data point is classified using the K-NN algorithm based on similarity after all the existing data has been stored. This means that utilizing the K- NN algorithm, new data may be easily sorted into a category that fits it. For regression and classification problem KNN can be used. It falls under the group of non-parametric algorithm which means it in independent on the elementary data. It is also termed as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data. The KNN method simply saves the information during the training phase, and when it receives new data, it categorizes data into a category that is similar to the new data.

3. RNN

A recurrent neural network (RNN) is a kind of artificial neural network that uses sequential data or time series data. These techniques are commonly used for ordinal or temporal situations, such as language translation, natural language processing(NLP), speech recognition, and image captioning; they are incorporated into popular applications such as Siri, voice search, and Google Translate. RNNs are a type of artificial neural network whose nodes are connected in a way that creates a directed graph along a sequence. This enables it to display dynamic and temporal behavior for a time sequence. RNNs can process input sequences using their internal state (memory), in contrast to feedforward neural networks. Recurrent neural networks (RNNs) are a particular kind of neural network in which the results of one step are fed into the current step as input.

4. CNN

A convolutional neural network(ConvNet/CNN) is a Deep Learning method that can take in an input image, give various elements and objects in the image importance (learnable weights and biases), and be able to differentiate between them. Comparatively speaking, a ConvNet requires substantially less preprocessing than other classification techniques. ConvNets have the ability to learn these filters and properties, whereas in primitive techniques filters are hand-engineered.

5. Openpose

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning technique that can take an input image, assign different elements and objects in the image importance (learnable weights and biases), and be able to distinguish between them. In comparison to other classification methods, a ConvNet requires significantly less preprocessing. Unlike to earlier approaches, where filters must be hand-engineered, ConvNets are capable of learning these filters and their attributes.

A. *Abbreviations and Acronyms*

TABLE I. DEFINITIONS, ACRONYMS AND ABBREVIATIONS

Term/ abbreviation	Description
OPENPOSE	This is a convolutional neural network key point detector of human body.
PYTHON	Python is an interpreter, high-level, general-purpose programming language.
TensorFlow	TensorFlow is an interface for expressing machine learning algorithms, and an implementation for executing such algorithms.
Eel	Eel is a software library that is providing a simple desktop GUI like applications using front-end components of web applications and back-end in Python/JavaScript.
SVM	Support Vector Machine The support vector machine is one of the supervised machine learning algorithms. Which aims to find the best decision boundaries in given n-spaces. It uses extreme points known as support points or vectors. These extreme points affect the position of the decision boundary.
KNN	K-Nearest Neighbors KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).[5][6]
RNN	Recurrent Neural Network RNN is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit temporal dynamic behavior for a time sequence. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs.
ANN	Artificial neural networks ANN or Neural Networks are computational algorithms. It intended to simulate the behavior of biological systems composed of “neurons”. ANNs are computational models inspired by an animal's central nervous systems. It is capable of machine learning as well as pattern recognition.[7]
ASL	American Sign Language
SD	Standard Deviation SD is a quantity expressing by how much the entities of a group differ from the mean value for the group.
JSON	JavaScript Object Notation JSON is a lightweight format for stacking and transporting data.

B. *Figures and Tables*

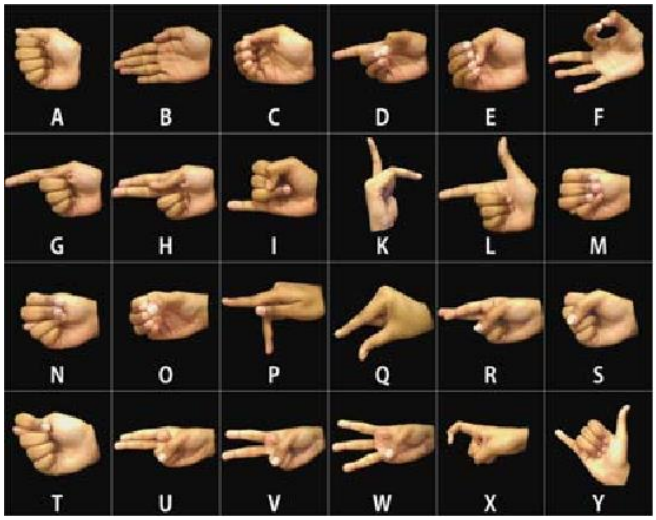

Fig. 1. American Sign Language (Alphabets)

Ref Fig 1. The majority sign language used by Deaf populations in the United State of America and most of Anglophone Canada is American Sign Language (ASL), a natural language. ASL is a comprehensive and well-structured visual language that uses both manual and non-manual characteristics to communicate. Dialects of ASL and creoles based on ASL are spoken throughout the globe, including much of West Africa and portions of Southeast Asia, besides North America.

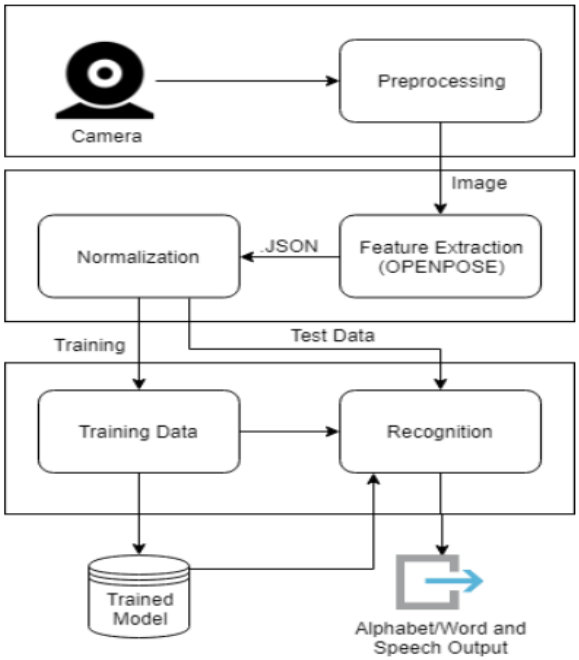

Fig 2. Product Perspective

Ref Fig 2. The suggested model is based on the approximated human skeletal key points from the independent library OpenPose. The OpenPose is an open-source toolkit that can predict 130 key points in total for real-time multi-person key point detection. The system uses skeletal tracking to capture images of a people making a specific sign using a webcam to specify what indication they are making in combination with machine learning. Our application's primary goal is to bridge gap between signer and non-signer. The user will make a sign in front of the webcam, which the system will identify and output as text and speech once it has done so.

C. Future Scope

A comprehensive system that covers all of these signals at this level is not feasible due to the hundreds of words that make up American Sign Language. All English alphabets, as well as some English words, can be converted into text and audio using the system which is provided. A structure for incorporating new terms into the system. The proposed method has demonstrated a very high rate of ASL recognition accuracy. But, there's still opportunity for development. The following recommendations are made for potential future improvements:

• The algorithm can be improved to recognize ASL signs that change over time.

• The ASL dataset can be expanded to include more vocabulary.

• ASL words can be utilized with Natural Language Processing (NLP) techniques to create meaningful sentences.

• The addition of the text and speech to ASL animation will enable two-way communication between signers and non-signers.

D. *Conclusion*

• A dataset module, a capture module, a learning module, a detection module, and an interface module make up the overall ASL translating system.

• Among these, the dataset module and the capture module have both been put into practice thus far. The dataset can be expanded by adding additional data using the capture module.
• An ASL dataset made up of the main spots on the hands of 9 distinct people as they signed 26 English alphabets.

REFERENCES

[1] Rini Akmeliawatil Melanie PO-Leen Ooi, et al, Real Time Malaysian Sign Language Translation using Color Segmentation and Neural Network. Instrumentation and Measurement Technology Conference Warsaw, Poland.IEEE.1-6, 2007.

[2] Nicholas Born, "Senior Project Sign Language Glove", ELECTRICAL ENGINEERING DEPARTMENT.California Polytechnic State University 1-49, 2010.

[3] Zhe Cao, Yaser Sheikh, et al, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv:1611.08050v2 [cs.CV], Apr 2017.

[4] Avinash Navlani," Support Vector Machines with Scikit-learn", available at https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python, 12 July 2018.

[5] Avinash Navlani," KNN Classification using Scikit-learn", available at https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikitlearn, 2 Aug 2018.

[6] Yun-lei Cai et al "A KNN Research Paper Classification Method Based on Shared Nearest Neighbor", Proceedings of NTCIR-8 Workshop Meeting, Tokyo, Japan, June 2010.

[7] Ms. Sonali. B. Maind et al," Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 1, 2014.

[8] Mart´ın Abadi, Ashish Agarwal, et al," TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems", Preliminary White Paper, November 9, 2015.

[9] M. Wasim, Abdulbasit Shaikh, A.A Siddiqui. Communicator for Hearing-Impaired Persons using Pakistan Sign Language (PSL). The International Journal of Advanced Computer Science and Applications, Vol. 9, No. 5, 2018.

[10] Kausar, S., Javed, M. Y., & Sohail, S. Recognition of gestures in Pakistani sign language using fuzzy classifier. In proceeding of the 8th conference on Signal processing, computational geometry and artificial vision (pp. 101-105). World Scientific and Engineering Academy and Society (WSEAS), August, 2008.

[11] Ali, S. A., "Detection of Urdu Sign Language using Harr Algorithms", International Journal of Inventive Engineering and Sciences (IJIES) ISSN: 2319–9598, Volume-1, Issue-6, May 2013.