

# Sign Language Recognition using Machine Intelligence for Hearing Impaired Person

Mrs. G. Nalina Keerthana  
Associate Professor  
M.I.E.T. Engineering college  
Trichy, India

Roshan Shabiha. A  
Computer science and Engineering  
M.I.E.T. Engineering College  
Trichy, India

Sahana.T  
Computer Science and Engineering  
M.I.E.T. Engineering college  
Trichy, India

Sana Zaffira  
Computer Science and Engineering  
M.I.E.T. Engineering College  
Trichy, India

**Abstract**— People with impaired speech and hearing use Sign language as a form of communication. Disabled People use sign language gestures as a tool of non-verbal communication to express their own emotions and thoughts to other common people. Conversing with people having a hearing disability is a major challenge. To overcome these issues, systems that recognize different signs and convey the information to normal people are needed. But these common people find it difficult to understand their expression, thus trained sign language expertise are needed during medical and legal appointment, educational and training session. Over the past few years, there has been an increase in demand for these services. To address this problem, we can implement artificial intelligence technology to analyse the users hand with finger detection. In this proposed system we can design the vision based system in real time environments. And then using deep learning algorithm named as Convolutional neural network algorithm to classify the sign and provide the label about recognized sign .

**Keywords**— *Sign Language Recognition, Convolutional Neural Network, Image Processing, Segmentation.*

## I. INTRODUCTION

Machine learning is an application of artificial intelligence (AI). It enables the system to automatically learn and improve from experience without being programmed explicitly. Machine learning aims on the development of computer programs that can access data and use it to learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning algorithms are often categorized as supervised or unsupervised. Supervised algorithms require a datascientist or data analyst with machine learning skills to provide both input and desired output, in addition to furnishing feedback about the accuracy of predictions during algorithm training. Once training is complete, the algorithm will apply what was learned to new data. Unsupervised algorithms do not need to be trained with desired outcome data. Instead, it uses an iterative approach called deep learning to review data and arrive at

conclusions. These algorithms are also called neural networks are used for more complex processing tasks than supervised learning systems, including image recognition, speech-to-text and natural language generation. These neural networks work by combing through millions of examples of training data and automatically identifying often subtle correlations between many variables. Once trained, the algorithm can use its bank of associations to interpret new data. These algorithms have only become feasible in the age of big data.

Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly

In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Semi-supervised machine learning algorithms fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn't require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant

characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

## II. EASE OF USE LITERATURE REVIEW

### A. *Sign Language Recognition Using Multiple Kernel Learning: A Case Study Of Pakistan Sign Language*

Sign language is the way of communication and interaction for deaf people all around the world. This kind of communication is accomplished over some hand gestures, facial expressions, or movement of arm/body. The sign language recognition system aims to enable the deaf community to communicate with normal society appropriately. It is a highly structured symbolic set that provides the human computer interaction. Sign language is very beneficial as a communication tool, and every day millions of deaf people around the world use sign language to communicate and express their ideas. This facilitation and assistance to deaf persons enable and encourage them to be a healthy part of society and integrate them into society. The dataset is obtained from the sign language videos. At a later step, four vision-based features are extracted. The extracted features are individually classified using Multiple kernel learning (MKL) in support vector machine (SVM). A voting scheme is adopted for the final recognition of PSL. The performance of the proposed technique is measured in terms of accuracy, precision, recall, and F-score. The simulation results are promising as compared with existing approaches.

### B. *Asl-3dcnn: American Sign Language Recognition Technique Using 3-D Convolutional Neural Networks*

Hand gestures are used as a way for people to express thoughts and feelings, it serves to reinforce information delivered in our daily conversation. Sign language is a structured form of hand gestures involving visual motions and signs, which are used as a communication system. For the deaf and speech-impaired community, sign language serves as useful tools for daily interaction. Sign language involves the use of different parts of body namely fingers, hand, arm, head, body and facial expression to deliver information. However, sign language is not common among the hearing community, and fewer are able to understand it. This poses a genuine communication barrier between the deaf community and the rest of the society, as

a problem yet to be fully solved until this day. There are growing numbers of emerging technology such as EMG, LMC, and Kinect which capture gesture information more readily. The common pre-processing method used are Median and Gaussian filter as well as downsizing of images prior to subsequent stages. Skin color segmentation is one of the most commonly used segmentation method. Color space which are generally more robust towards illumination condition are CIE Lab, YCbCr and HSV. More recent research utilizes combination of several others spatial features and modelling approaches to improve segmentation performance.

### C. *Comprehensive Study On Deep Learning-Based Methods For Sign Language Recognition*

Spoken languages make use of the “vocal - auditory” channel, as they are articulated with the mouth and perceived with the ear. All writing systems also derive from, or are representations of, spoken languages. Sign languages are different as they make use of the “corporal - visual” channel, produced with the body and perceived with the eyes. SLs are not international and they are widely used by the communities of the Deaf. They are natural languages since they are developed spontaneously wherever the Deaf have the opportunity to congregate and communicate mutually. A comparative experimental assessment of computer vision based methods for sign language recognition is conducted. By implementing the most recent deep neural network methods in this field, a thorough evaluation on multiple publicly available datasets is performed. The aim of the present study is to provide insights on sign language recognition, focusing on mapping non-segmented video streams to glosses. For this task, two new sequence training criteria, known from the fields of speech and scene text recognition, are introduced. Furthermore, a plethora of pre-training schemes is thoroughly discussed. Finally, a new RGB+D dataset for the Greek sign language is created. Sign Language Recognition (SLR) can be defined as the task of inferring glosses performed by a signer from video captures. Even though there is a significant amount of work in the field of SLR, a lack of a complete experimental study is profound.

### D. *Independent Sign Language Recognition With 3d Body, Hands, And Face Reconstruction*

In this paper we investigated the extraction of 3D body pose, face and hand features for the task of Sign Language Recognition. These features are compared, to Openpose key-points, the most famous method for extracting 2D skeleton parameters and features from raw RGB frames and their optical flow that are fed in a state-of-the-art Deep Learning architecture used in action and sign language recognition. The experiments revealed the superiority of SMPL-X features due to the detailed and qualitative features extraction in the three aforementioned regions of interest. Moreover, SMPL-X to point out the significance of combining all these three regions for optimal results in SLR are exploited. Future work on 3D

body, face and hands extraction for SLR includes further experiments in different independent datasets with more signers and varying environment. Furthermore, Applying SMPL-X in continuous SLR will give further prominence to this method, where facial expressions and body structure are even more crucial. Finally, applying SMPL-X in different action recognition tasks is an interesting experiment to examine the universality of SMPL-X success. In this work, SMPL-X, a contemporary parametric model that enables joint extraction of 3D body shape, face and hands information from a single image. This holistic 3D reconstruction for SLR, demonstrating that it leads to higher accuracy than recognition from raw RGB images and their optical flow fed into the state-of-the-art I3D-type network for 3D action recognition and from 2D Openpose skeletons fed into a RNN.

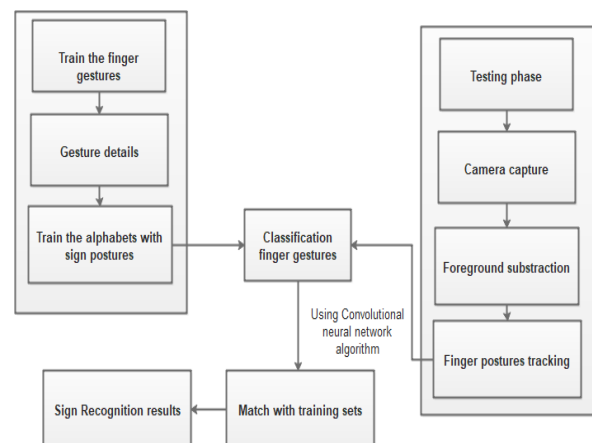
#### E. *Explicit Quaternion Krawtchouk Moment Invariants For Finger-Spelling Sign Language Recognition*

Sign language recognition (SLR) is an important aspect in human computer interaction applications, used by deaf and hearing impaired. In fact, SLR are gestural languages which uses signs for communication without speaking. This task is a challenging problem, mainly because of the nature of the hidden computer vision problem, such as the visual analogy of specific signs and the complex articulations presented by the hand. Typically, there are three components to constitute a sign gesture: manual features which are gestures made with hands, nonmanual features such as facial expressions, body posture, and finger-spelling when words are spelt into alphabet. In this context, the importance of fingerspelling can be noticed when a concept lacks a specific sign, such as names, technical terms, or foreign words. Sign recognition is a difficult task due to the complexity of its composition which uses signs of different levels, words, facial expression, body posture and finger-spelling to convey meaning. With the development of recent technologies, such as Kinect sensor, new opportunities have emerged in the field of human computer interaction and sign language, allowing to capture both RGB and Depth (RGB-D) information. In the regard to feature extraction, the traditional methods process the RGB and Depth images independently. A robust static fingerspelling sign language recognition system adopting the Quaternion algebra that provide a more robust and holistical representation, based on fusing RGB images and Depth information simultaneously is proposed in this system. A new sets of Quaternion Krawtchouk moments(QKMs) and Explicit Quaternion Krawtchouk 19 Moment Invariants (EQKMIs) is proposed for the first time. The proposed system is evaluated on three well-known fingerspelling datasets, demonstrate the performance of the novel method compared to other methods used in the literature, against geometrical distortion, noisy conditions and complex background, indicating that it could be highly effective for many other computer vision applications.

### III. EXISTING SYSTEM

The sign language is used widely by people who are deaf-dumb these are used as a medium for communication. A sign language is nothing but composed of various gestures formed by different shapes of hand, its movements, orientations as well as the facial expressions. There are around 466 million people worldwide with hearing loss and 34 million of these are children. Deaf people have very little or no hearing ability. They use sign language for communication. People use different sign languages in different parts of the world. Compared to spoken languages they are very less in number. In existing system, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in finger gesture detection. Existing project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people using Indian sign language. Effective extension to words and common expressions may not only make the deaf and dumb people communicate faster and easier with outer world, but also provide a boost in Developing autonomous systems for understanding and aiding them. The Indian Sign Language lags behind its American Counterpart as the research in this field is hampered by the lack of standard datasets.

### IV. SYSTEM ARCHITECTURE



### V. PROPOSED SYSTEM

Instead of acoustic sound patterns, Sign Language is a gesture-based language that uses hand movements, hand orientation, and facial expression. This form of language is not universal and has different patterns depending on the people. However, because most individuals aren't familiar with sign language, Deaf-mute persons are finding it more difficult to communicate without the aid of a translation of some sort. They believe they are being shunned. Between deafmute people and normal people, Sign Language Recognition has become a commonly accepted communication approach. Computer vision- based and sensor-based systems are two types of recognition models. The camera is utilized for input in computer visionbased gesture recognition, and image processing of input motions is done before recognition. Following that, several algorithms such as region of interest algorithm and Neural

Network approaches are used to recognize the processed gestures. The fundamental disadvantage of a vision-based sign language recognition system is that the picture collecting process is subject to numerous environmental concerns, such as camera placement, background conditions, and lightning sensitivity. However, it is more convenient and cost-effective than employing a camera and tracker to collect data. However, for greater accuracy, Neural Network methods like the Hidden Markov Model are combined with camera data.

### 5.1 CONVOLUTION NEURAL NETWORK

Artificial Neural Networks (ANN) can learn and therefore can be trained to recognize patterns, find solutions, forecast future events and classify data. CNN is well documented to be used for traffic related tasks. Neural Networks learning and behavior is dependent on the way its individual computing elements are connected and by the strengths of these connections or weights. These weights can be adjusted automatically by training the network according to a specified learning rule until it performs the desired task correctly. CNN is a supervised learning method i.e. a machine learning algorithm that uses known dataset also known as training dataset. These known parameters help CNN to make predictions. Input data along with their response values are the fundamental components of a training dataset. In order to have higher predictive power and the ability to generalize for several new datasets, the best way is to use larger training datasets. The fingers can be classified by using convolutional neural network algorithm. CNN is a common method of training artificial neural networks so as to minimize the objective function. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feedforward networks.

### 5.2 HAND IMAGE ACQUISITION

The hand gesture, during daily life, is a natural communication method mostly used only among people who have some difficulty in speaking or hearing. However, a human computer interaction system based on gestures has various application scenarios. In this module, input the hand images are acquired from real time camera. The inbuilt camera can be connected to the system. Gesture recognition has become a hot topic for decades. Nowadays two methods are used primarily to perform gesture recognition. One is based on professional, wearable electromagnetic devices, like special gloves. The other one utilizes computer vision. The former one is mainly used in the film industry. It performs well but is costly and unusable in some environment. The latter one involves image processing. However, the performance of gesture recognition directly based on the features extracted by image processing is relatively limited. Hand image captured from web camera. The purpose of Web camera is to capture the human generated hand gesture and store its image in memory. The package called python framework is used for storing image in memory.

### 5.3 BINARIZATION

Background subtraction is one of the major tasks in the field of computer vision and image processing whose aim is to detect changes in image sequences. Background subtraction is any technique which allows an image's foreground to be extracted for further processing. Many applications do not need to know everything about the evolution of movement in a video sequence, but only require the information of changes in the scene, because an image's regions of interest are objects human, car, text, etc. in its foreground. After the stage of image preprocessing object localization is required which may make use of this technique. Detecting foreground to separate these changes taking place in the foreground of the background. It is a set of techniques that typically analyze the video sequences in real time and are recorded with a stationary camera. All detection techniques are based on modeling the background of the image i.e. set the background and detect which changes occur. Defining the background can be very difficult when it contains shapes, shadows, and moving objects. In defining the background it is assumed that the stationary objects could vary in color and intensity over time. Scenarios where these techniques apply tend to be very diverse. There can be highly variable sequences, such as images with very different lighting, interiors, exteriors, quality, and noise. In addition to processing in real time, systems need to be able to adapt to these changes. The implement the techniques to extract the foreground from background image. Using Binarization approach to assign the values to background and foreground. Foreground pixels are identified in real time environments.

### 5.4 REGION OF FINGER DETECTION

Segmentation refers to the process of partitioning a digital image into multiple segments. In other words, grouping of pixels into different groups is known as Segmentation. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The division of an image into meaningful structures, image segmentation, is often an essential step in image analysis, object representation, visualization, and many other image processing tasks. But segmentation of a satellite image into differently textured regions or groups is a difficult problem. One does not know a priori what types of textures exist in a satellite image, how many textures there are, and what regions have certain textures. The monitoring task can be performed by unsupervised segmentation and supervised segmentation techniques. A region of interest is a subset of an image or a dataset identified for a particular purpose. In other words, region of interest can be defined as a portion of an image which is needed to be filtered or to be performed some other operations on.

## VI. CONCLUSION

The ability to look, listen, talk, and respond appropriately to events is one of the most valuable gifts a human being can have. However, some unfortunate people are denied this opportunity. People get to know one another through



sharing their ideas, thoughts, and experiences with others around them. There are several ways to accomplish this, the best of which is the gift of "Speech."

Everyone can very persuasively transfer their thoughts and comprehend each other through speech. Our initiative intends to close the gap by including a low-cost computer into the communication chain, allowing sign language to be captured, recognised, and translated into speech for the benefit of blind individuals. An image processing technique is employed in this paper to recognise the handmade movements. This application is used to present a modern integrated planned system for hear impaired people. The camera-based zone of interest can aid in the user's data collection. Each action will be significant in its own right.

#### REFERENCE

- [1] Bhanu.B and Kumar.A Deep Learning for Biometrics. Springer, 2017.
- [2] Boyd.S and Vandenberghe.L Convex Optimization. Stanford University, 2004.
- [3] Bubeck.S Convex optimization: Algorithms and complexity. Foundations and Trends in Machine Learning, 2015.
- [4] Chinese Academy of Sciences Institute of Automation. CASIA iris image database, Aug 2017.
- [5] Daugman.J How iris recognition works? IEEE Transactions on Circuits and Systems for VideoTechnology, 2004.
- [6] Daugman.J New methods in iris recognition. IEEE Transactions on Systems, Man and Cybernetics, 2007.
- [7] Daugman.J Information theory and the iriscode. IEEE Transactions on Information Forensics and Security, Feb 2016.
- [8] Daugman.J andDowning.C Searching for doppelgangers: assessing the universality of the iriscode impostors distribution. IET Biometrics, 2016.
- [9] Franceschi.L, Frasconi.P, Salzo.S, Grazzi.R, and Pontil.M Bilevel programming for hyperparameter optimization and metalearning. In International Conference on Machine Learning (ICML), 2018.
- [10] Gangwar.Aand Joshi.ADeepirisnet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition. In IEEE International Conference on Image Processing (ICIP),Sep 2016.