Indian Sign Language Recognition

A Synopsis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

# Bachelor In Technology

in

# Information Technology

by

# Dinesh Kumar (35251203117)

# Vaibhav Khera (35651203117)

# Ayush Rastogi (40151203117)

Under Supervision of

[**Mrs.**](http://it.bvcoend.ac.in/site/home/index/249) **Shafali Dhall**

DEPARTMENT OF INFORMATION TECHNOLOGY

# BHARATI VIDYAPEETH’S COLLEGE OF ENGINEERING

# (AFFILIATED TO GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY, DELHI)

# Paschim Vihar, New Delhi

# September, 2019

Synopsis

1. **Introduction**

Deaf is a disability that impair their hearing and make them unable to hear , while mute is a disability that impair their speaking and make them unable to speak . Both are only disabled at their hearing and/or speaking, therefore can still do much other things. The only thing that separate them and the normal people is communication. And the only way for them to communicate is through sign language. Indian Sign Language (ISL) substantially facilitates communication in the deaf community. One of the solution to communicate with the deaf-mute people is by using the services of sign language interpreter. But the usage of sign language interpreter can be costly. Cheap solution is required so that the deaf-mute and normal people can communicate normally.

Therefore, researchers want to find a way for the deaf-mute people so that they can communicate easily with normal person. The breakthrough for this is the Sign Language Recognition System. The system aims to recognize the sign language, and translate it to the local language via text or speech. Early researches have known to be successful in Sign Language Recognition System by using data gloves. But, the high cost of the gloves and wearable character make it difficult to be commercialized. In order to diminish this obstacle and to enable dynamic communication, we present an ISL recognition system that uses Convolutional Neural Networks (CNN) in real time to translate user’s ISL signs into text. With the use of Neural Networks in image processing the input image is compared with set of images in the dataset the word corresponds to the matched image will gives the output. Thus human can easily interact with others.

1. **Motivation**

Communication is one of the basic requirement for survival in society. Deaf and dumb people communicate among themselves using sign language but normal people find it difficult to understand their language. The only thing that separate them and the normal people is communication. If there is a way for normal people and deaf-mute people to communicate, the deaf-mute people can easily live like a normal person. Sign language is boon for the deaf and dumb people. Sign language is the combination of different gesture, shape and movement of hand, body and facial expression

In order to diminish this obstacle and to enable dynamic communication, we present an ISL recognition system. Extensive work has been done on American sign language recognition but Indian sign language differs significantly from American sign language. ISL uses two hands for communicating(20 out of 26) whereas ASL uses single hand for communicating. Using both hands often leads to obscurity of features due to overlapping of hands. Our 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 of this project 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.

1. **Related work**

ASL recognition is not a new computer vision problem. Over the past two decades, researchers have worked on this project and developed sign recognition project via using electromechanically device such as a glove based system when this system developed research has been limited to small scale systems able of recognizing a minimum subset of a sign language.

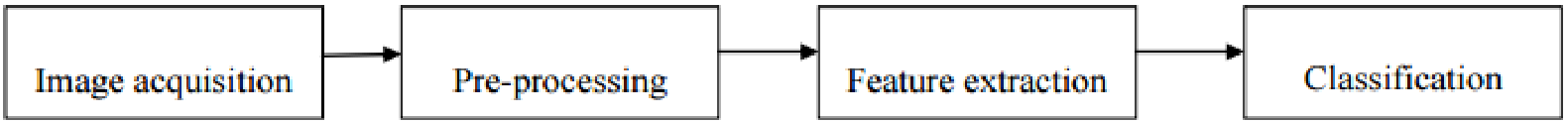
Researcher’s also proposed the use of a back propagation neural network for recognition of gestures from a set of segmented hand images. This system showed promise in the field of language invariant teleconferencing.

1. **Proposed Method**

## Vision­based

In vision based methods computer camera is the input device for observing the information of hands or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices. These systems tend to complement biological vision by describing artificial vision systems that are implemented in software and/or hardware. This poses a challenging problem as these systems need to be background invariant, lighting insensitive, person and camera independent to achieve real time performance. Moreover, such systems must be optimized to meet the requirements, including accuracy and robustness.

The vision based hand gesture recognition system is shown in fig.­­:



### Fig : Block Diagram of vision based recognition system

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far.

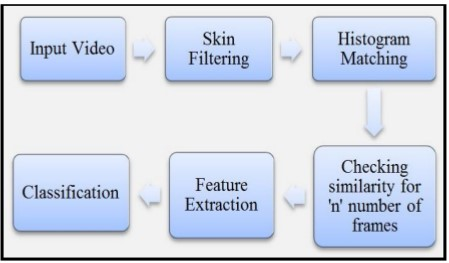
* One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters

corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification.

* Second one to capture the image using a camera then extract some feature and those features are used as input in a classification algorithm for classification.

## **Automatic Indian Sign Language Recognition for Continuous Video Sequence**

The proposed system comprises of four major modules: Data Acquisition, Pre­processing, Feature Extraction and Classification. Pre­processing stage involves Skin Filtering and histogram matching after which Eigen­vector based Feature Extraction and Eigen value weighted Euclidean distance based Classification Technique was used.



* **Recognition of isolated Indian Sign Language Gesture in Real Time**

This proposed method demonstrates the statistical techniques for recognition of ISL gestures in real time which comprises both the hands. A video database was created by the authors and utilized which contained several videos for large number of signs. Direction histogram is the feature used for classification due to its appeal for illumination and orientation invariance. Two different approaches utilized for recognition were Euclidean distance and K­nearest neighbor metrics.

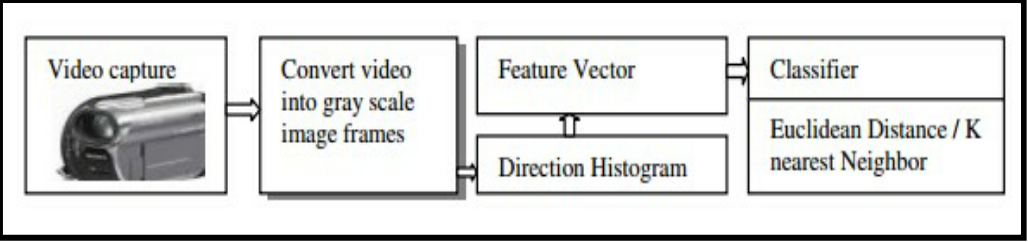


Fig: Methodology for real time ISL classification

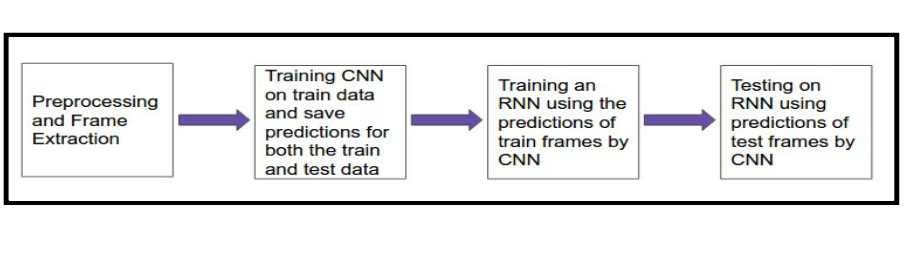
1. **Methodology**

We have used the approach to train the model on the temporal and the spatial features.

In this approach we extracted spatial features for individual frames using inception model (CNN) and temporal features using RNN(Recurrent Neural Network ). Each video (a sequence of frames) was then represented by a sequence of predictions made by CNN for each of the individual frames. This sequence of predictions was given as input to the RNN.

Methodology:

* First, we will extract the frames from the multiple video sequences of each gesture.
* After the first step, noise from the frames i.e. background, body parts other than hand are removed to extract more relevant features from the frame.
* Frames of the train data are given to the CNN model for training on the spatial features. We have used inception model for this purpose which is a deep neural net.
* Store the train and test frame predictions. We’ll use the model obtained in the above step for the prediction of frames.
* The predictions of the train data are now given to the RNN model for training on the temporal features.

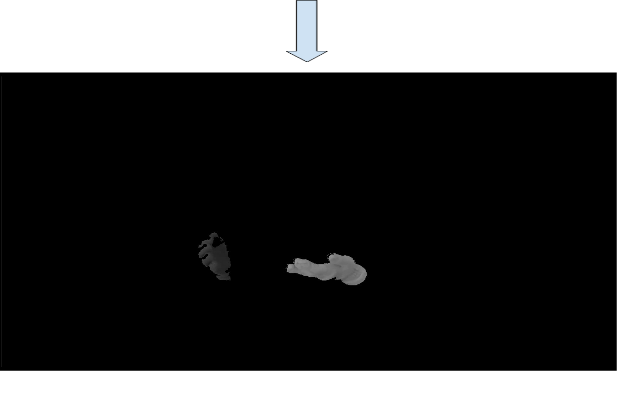


Each step of the methodology has been shown diagrammatically for better understanding of that steps.

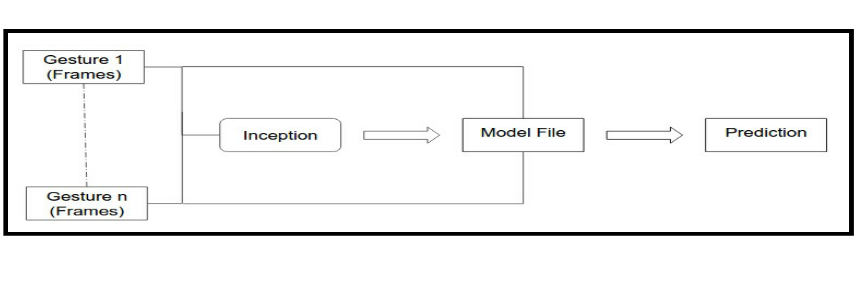
#### Frame Extraction and Background Removal :

Each video gesture video is broken down into a sequence of frames. Frames are then processed to remove all the noise from the image that is everything except hands.

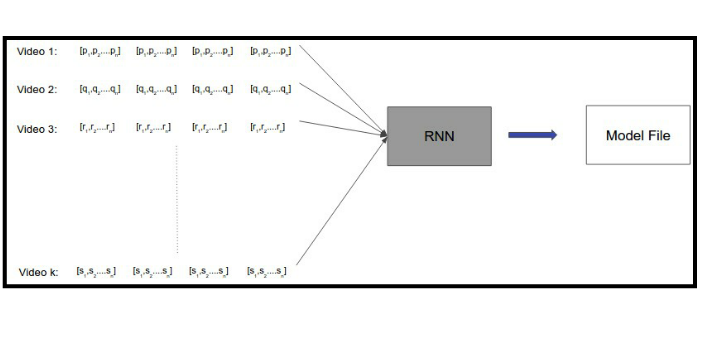
The final image consists of grey scale image of hands to avoid color specific learning of the model

**Background Removal**

#### Train CNN(Spatial Features) and Prediction :



#### Training RNN (Temporal Features) :



Videos are difficult to classify because they contain both the temporal as well as the spatial features. We have used two different models to classify on the spatial and temporal features. CNN was used to classify on the spatial features whereas RNN was used to classify on the temporal features.

1. **Plan of work**

First Month-

1. Firstly we will be collecting Dataset for the Indian Sign Language. As per now, there is no such dataset available for Indian Sign Language. All dataset is for ASL.
2. After Dataset is obtained, we will be extracting frames of each gesture from data and removal of background and other noises from it except hands(gesture). The final image consists of grey scale image of hands to avoid color specific learning of the model.

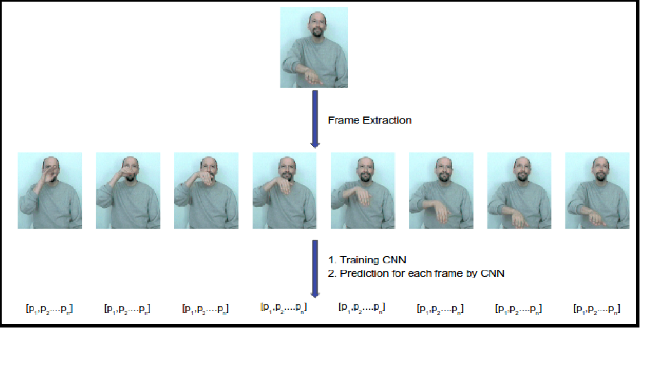
Second Month-

1. Then after we will be providing frames of the train data to the CNN model for training on the spatial features. We have used inception model for this purpose which is a deep neural net.
2. We’ll use the model obtained in the above step for the prediction of frames. The predictions of the train data are now given to the RNN model for training on the temporal features.
3. **Future Work**

Currently ,we will be working on extracting letters and recognizing them through our gestures. This Project can be further improvised and can be worked upon it so that it can recognize words through our gestures .

For Example-

The first row in the below illustration is the video of a gesture Elephant. The second row shows the set of frames extracted from it. The third row shows the sequence of predictions for each frame by CNN after training it.



We can also build an android application which will be used to convert sign language to text and vice-versa.

**References**

[1] Singha, Joyeeta, and Karen Das. "Automatic Indian Sign Language

Recognition for Continuous Video Sequence." ADBU Journal of Engineering Technology 2, no. 1 (2015).

[2] Tripathi, Kumud, and Neha Baranwal GC Nandi. "Continuous

Indian Sign Language Gesture Recognition and Sentence Formation." Procedia Computer Science 54 (2015): 523­531.

[3]https://medium.com/technologymadeeasy/the­best­explanation­of­convolutional­neural­networks­on­the­internet­fbb8b1ad5df8

[4] Zhang, Chenyang, Xiaodong Yang, and YingLi Tian. "Histogram of 3D facets: A characteristic descriptor for hand gesture recognition." In Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on, pp. 1­8. IEEE, 2013.

[5] Cooper, Helen, Eng­Jon Ong, Nicolas Pugeault, and Richard Bowden. "Sign language recognition using sub­units." Journal of

Machine Learning Research 13, no. Jul (2012): 2205­2231

[6] JOYEETA SINGH, K. D. Indian sign language recognition using eigen value weighted euclidean distance based classiﬁcation technique. International Journal of Advanced Computer Science and Applications 4, 2 (2013).

[7] NEHA V. TAVARI, P. A. V. D. Indian sign language recognition based on histograms of oriented gradient. International Journal of Computer Science and Information Technologies 5, 3 (2014), 3657–3660.