Literature Review on

Real - Time Sign Language Recognition System

**PROBLEM STATEMENT:**

Speech impaired people use hand signs and gestures to communicate. Normal people face difficulty in understanding their language. Hence there is a need of a system which recognizes the different signs, gestures and conveys the information to the normal people. To overcome this barrier, one can design a Real Time Sign Language Detection System using TensorFlow object detection API which aims to bridge the gap between the speech and hearing-impaired people and the normal people.

**INTRODUCTION:**

People with impaired speech and hearing uses Sign language as a form of communication. Disabled People use this sign language gestures as a tool of non-verbal communication to express their own emotions and thoughts to other common people. 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. Other form of services such as video remote human interpret using the high-speed Internet connection, has been introduced, thus these services provide an easy-to-use sign language interpret service, which can be used and benefited, yet have major limitations.

**ABSTRACT:**

Communication is defined as the act of sharing or exchanging information, ideas or feelings. To establish communication between two people, both of them are required to have knowledge and understanding of a common language. But in the case of deaf and dumb people, the means of communication are different. Deaf is the inability to hear and dumb is the inability to speak. They communicate using sign language among themselves and with normal people but normal people do not take seriously the importance of sign language. Not everyone possesses the knowledge and understanding of sign language which makes communication difficult between a normal person and a deaf and dumb person. To overcome this barrier, one can build a model based on machine learning. A model can be trained to recognize different gestures of sign language and translate them into English. This will help a lot of people in communicating and conversing with deaf and dumb people. The existing American Sign Language Recognition systems are designed using machine learning algorithms with single and double-handed gestures but they are not real-time. In this paper, we propose a method to create an American Sign Language dataset using a webcam and then using transfer learning, train a TensorFlow model to create a real-time Sign Language Recognition system. The system achieves a good level of accuracy even with a limited size dataset.

**KEYWORDS:**

Sign Language Recognition (SLR), Computer Vision, Machine Learning, American Sign Language.

**SUMMARY:**

**LITERATURE SURVEY :**

Sign languages are defined as an organized collection of hand gestures having specific meanings which are employed from the hearing-impaired people to communicate in everyday life. Being visual languages, they use the movements of hands, face, and body as communication mediums. There are over 300 different sign languages available all around the world. Though there are so many different sign languages, the percentage of population knowing any of them is low which makes it difficult for the specially-abled people to communicate freely with everyone. SLR provides a means to communicate in sign language without knowing it. It recognizes a gesture and translates it into a commonly spoken language like English.

SLR is a very vast topic for research where a lot of work has been done but still various things need to be addressed. The machine learning techniques allow the electronic systems to take decisions based on experience i.e., data. The classification algorithms need two datasets – training dataset and testing dataset. The training set provides experiences to the classifier and the model is tested using the testing set. Many authors have developed efficient data acquisition and classification methods. Based on data acquisition method, previous work can be categorized into two approaches: the direct measurement methods and the vision-based approaches. The direct measurement methods are based on motion data gloves, motion capturing systems, or sensors. The motion data extracted can supply accurate tracking of fingers, hands, and other body parts which leads to robust SLR methodologies development. The vision-based SLR approaches rely on the extraction of discriminative spatial and temporal from RGB images. Most of the vision-based methods initially try to track and extract the hand regions before their classification to gestures. Hand detection is achieved by semantic segmentation and skin colour detection as the skin colour is usually distinguishable easily. Though, because the other body parts like face and arms can be mistakenly recognized as hands, so, the recent hand detection methods also use the face detection and subtraction, and background subtraction to recognize only the moving parts in a scene. To attain accurate and robust hands tracking, particularly in cases of obstructions, authors employed filtering techniques, for example, Kalman and particle filters.

For data acquisition by either the direct measurement or the vision-based approaches, different devices need to be used. The primary device employed as input process in SLR system is camera. There are other devices available that are used for input such as Microsoft Kinect which provides colour video stream and depth video stream all together. The depth data helps in background segmentation. Apart from the devices, other methods used for acquiring data are accelerometer and sensory gloves. Another system that is used for data acquisition is Leap Motion Controller (LMC) – it is a touchless controller developed by technology company “Leap Motion” now called “Ultra leap” based in San Francisco. Approximately, it can operate around 200 frames per second and can detect and track the hands, fingers, and objects that look alike fingers. Most of the researchers collect their training dataset by recording it from their signer as finding a sign language dataset is a problem.

**TECHNOLOGIES USED:**

1)OpenCV

2)LabelImg

3)Intel i5 7th generation 2.70 GHz processor

4) 8 GB memory and webcam (HP True Vision HD camera with 0.31 MP and 640x480 resolution) with running Windows 10 operating system.

**SYSTEM ARCHITECTURE :**

The architecture of SLR system is fragmented into six consequent phases namely; image capturing, image pre-processing, region extraction, feature extraction, feature matching and patten recognition.

Image capturing phase deals with capturing ASL symbol image in RGB colour space with frame size 160x120 using web camera attached to the laptop. Single frame of image is captured in complex background along with mixed lighting condition and the distance of I meter rom web camera. Image pre-processing includes pre-processing operations such as colour conversion, noise removal and morphological operations. Skin detector is applied for extracting only skin pixels and ignoring non-skin pixels rom input RGB image. Gray threshold is applied on skin pixels with specific probability. Filtering of Gray image is done using Median Filter for preserving edges followed by Gaussian filter is applied on current image for smoothing of image. Morphological operation such as blur, eroaion and dialation are applied on filtered binary image. In region extraction, BLOB of size 80x60 is employed to extract Region of Interest (ROI) of hand in SLR system. EOR technique uses coefficient of EOR to form feature vector. The noise free image obtained from the previous sub-phase (i.e. morphological operations) acts as input in this phase. The input image contain sub-images as regions and these regions are labelled with the use of eight-pixel connectivity. We have calculated the area of each labelled region as area set and selected largest area rom area set. This selected largest region is set be true value (white pixel) and other remaining regions are set be false value (black pixel). At this stage, image contains only a biggest region, the region as biggest BLOB and this BLOB represents only hand in human image. Feature extraction phase is based on extraction of EOR coefficient and formation of feature vector using EOR coefficients for both running image and raining dataset images. EOR technique is based on EOH descriptor model that measures the similarity between EOH of running image and training dataset images of ASL alphabet. The CompareHist descriptor is used to evaluate the best suited pattern as ASL alphabet as

Ci = max CompareHist (R, Xj)

In this model, CompareRist(R, Xj) computes the distance between EOR feature vector of running image and training dataset images, a represents the ASL symbol database for ASL alphabets, R is a running image of ASL alphabet, Xi G=I, ... , N) denotes the training dataset image EOR for ASL alphabets. Feature matching phase matches feature vector of training dataset images and running image using CompareRist (Training Image, Running Image) unction using Sim-EOH algorithm .. Patten recognition is comprised of finding similarity between feature vectors of rammg dataset images and running image thereby using K-Cluster EOH-Match algorithm. Maximum similarity endow with appropriate ASL alphabet (i.e. text) as recognized pattern.

video

Capturing

Image

Pre-processing

Region

Extraction

Feature

Extraction

Feature

Matching

Pattern

Recognition

***System Architecture of SLR system***

**WORKING MODEL :**

**ADVANTAGES :**

* The output of the sign language will be displayed in the text form in real time. This makes the system much more efficient and makes communications easy for the disabled.
* It is portable.
* It doesn’t get damaged through use.
* In future work proposed system can be developed and implemented using Raspberry pi.

**DISADVANTAGES:**

* Sign language recognition system is difficult for children to understand.
* Sign language has no written component. Deaf people can only use it to communicate face to face.
* Sign language recognition system requires the use of hands to make gestures. This can be a problem for people who do not have full use of their hands.
* Sign language recognition system is difficult to be used commercially.

**CONCLUSION:**

Our project aims to make communication simpler between the special categories of people i.e., deaf and dumb people by introducing computer in communication path so that sign language can be automatically captured, recognized, translated to text and display it on the screen

Sign languages are kinds of visual languages that employ movements of hands, body, and facial expression as a means of communication. Sign languages are important for specially-abled people to have a means of communication. Through it, they can communicate and express and share their feelings with others. The drawback is that not everyone possesses the knowledge of sign languages which limits communication. This limitation can be overcome by the use of automated Sign Language Recognition systems which will be able to easily translate the sign language gestures into commonly spoken language. In this paper, it has been done by TensorFlow object detection API. The system has been trained on the American Sign Language alphabet dataset. The system detects sign language in real-time. For data acquisition, images have been captured by a webcam using Python and OpenCV which makes the cost cheaper. The developed system is showing an average confidence rate of 85.45%. Though the system has achieved a high average confidence rate, the dataset it has been trained on is small in size and limited. In the future, the dataset can be enlarged so that the system can recognize more gestures. The TensorFlow model that has been used can be interchanged with another model as well. The system can be implemented for different sign languages by changing the dataset.

**REFERENCES:**

* The following links and documents have been referred to as a part of our project and for better understanding
* <https://www.ijser.org/paper/Sign-Language-Recognition-System.html#:~:text=The%20output%20of%20the%20sign,displayed%20at%20the%20same%20time>.
* Research paper on
  + - Sign Language Recognition System using TensorFlow Object Detection API.
    - By : Sharvani Srivastava, Amisha Gangwar, Richa Mishra, Sudhakar Singh
* Research paper on
  + - A Robust Sign Language and Hand Gesture Recognition System using Convolution Neural Network. By : ANITS
* Aditya Das, Shantanu Gawde, Khyati Surat Wala, Dr. Dhananjay Kalbande (2018, February). Facial expression recognition from video sequences: temporal and static modelling. Computer Vision and Image Undertaking 91.
* Zafar Ahmed Ansari and Gaurav Harit, “Nearest Neighbour Classification of American Sign Language Gestures using Kinect Camera”, in Sadhana, Vol. 41, No. 2, February 2016, pp. 161-182.
* [Recognition of Isolated American Sign Language Gesture in Real Time, Anup Nandy, Jay Shankar Prasad, Soumik Mondal, Pavan Chakraborty, G. C. Nandi, Communications in Computer and Information Science book series (CCIS, volume 70).
* Continuous dynamic American Sign Language gesture recognition with invariant backgrounds by Kumud Tripathi, Neha Baranwal, G. C. Nandi at 2015 Conference on Advances in Computing, Communications and Informatics (ICACCI).
* Adam: A Method for Stochastic Optimization, Diederik P. Kingma, Jimmy Ba, Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
* S. Tamura and S. Kawasaki, “Recognition of Sign Language Motion Images”, In Pattern Recognition, volume 21, pages 343-353, 1988.