International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 07 | July 2022 www.irjet.net p-ISSN: 2395-0072

Literature Review on Indian Sign Language Recognition System

Sukanya Dessai[1], Siddhi Naik[2]

[1] Student, Department of Information and Technology and Engineering, Goa College of Engineering, India
[2] Assistant Professor, Department of Information and Technology and Engineering, Goa College of Engineering, India

---

Abstract - *Machines' ability to comprehend human actions and their meaning can be used in a wide range of situations. Sign language recognition is one area of research that has piqued curiosity. Deaf and dumb persons use sign language as a means of communication. They must rely on sign language translators to communicate with the rest of the world. The strategies employed in contemporary sign language recognition research are reviewed in this research. Data capture, preprocessing, segmentation, feature extraction, and classification are all stages of the methodologies that are examined. Each stage of the above-mentioned literature review is completed independently. Artificial neural networks, Support Vector Machines and Fuzzy Inference Systems are some of the most common categorization approaches used for recognition.*

Key Words: **Hand Gesture, Pattern, Sign Language, Recognition, Classification**

1. INTRODUCTION

The hearing-impaired community uses sign language as a means of communication. The Rights of Persons with Disabilities (RPwD) Act 2016, which took effect on April 19, 2017, recognizes sign language as a form of communication that is particularly effective for communicating with people who are deaf or hard of hearing [9]. It consists of a variety of hand formations, orientations, hand movements, and facial expressions that are used to transmit messages. Every gesture has a specific message. Sign language is not universal. Different sign languages, such as the American Sign Language (ASL), British Sign Language (BSL), Japanese Sign Language, and so on, can be found all over the world.

It is tough and expensive to find experienced and competent Interpreters on a regular basis. Non-deaf people, on the other hand, never try to learn sign language in order to communicate with deaf people. This leads to the isolation of deaf people. The distance between normal people and the deaf community can be bridged if a computer can be programmed to translate sign language into text format. It can be used to generate voice, or text, allowing the deaf to gain independence.

The organization of the paper is as follows: Section II gives information about the previous work done on sign languages. Section III focuses on the sign language approaches. Section IV describes the overview of the sign language recognition system. Section V contains discussion and conclusion.

2. RELATED WORKS

Starner and Pentland[10] provided one of the early studies on sign language recognition. They demonstrated a real-time hidden Markov model-based system that detected sentence-level American Sign Language (ASL) movements with the help of a webcam. They described two experiments: the first uses a desk-mounted camera to view the user, while the second uses a camera embedded in the user's cap.

The authors in paper [3] presented a system for the automatic translation of the gestures of the manual alphabets in Arabic Sign Language. This system made use of images of the gesture as input which were then processed and converted into a set of features that comprised of some length measures which indicated the fingertip's position. The subtractive clustering algorithm and the least-squares estimator were used for classification. The system achieved an accuracy of 95.55%. In [4] Nadia R. Albelwi and Yasser M. Alginahi proposed a real-time Arabic Sign Language system where a video camera was used to capture real-time video as an input to the system. The authors used a Haar-like algorithm to track the hand in the video frames and applied preprocessing techniques like skin detection and size normalization to extract the region of interest. To obtain the feature vectors, Fourier Transformation is applied to the resultant images which are transformed into the frequency domain. The classification is performed using the k-Nearest Neighbor (KNN) algorithm and the system achieves an accuracy of 90.55%.

In [5] Balakrishnan, G., P. S. Rajam, et al., proposed a system that converts a set of 32 combinations of the binary number, which represents the UP and DOWN positions of the five fingers into decimal. The binary numbers are first converted into a decimal form by using the binary-decimal conversion algorithm and then the decimal numbers are converted to their corresponding Tamil letters. Static images of the gesture were used as the input to the system where a canny-edge detection algorithm was applied to extract the edges of the palm and Euclidean Distance was applied to identify the position of the fingers. The system achieved an accuracy of 98.75%.

The authors in paper [6] proposed a system employing b-spline approximation to develop a novel vision-based recognition system for Indian sign language alphabets and digits. By using the Maximum Curvature Points (MCPs) as Control points, their technique approximates the extracted boundary from the region of interest to a B-Spline curve. The B-spline curve is then smoothed iteratively, resulting in the extraction of Key Maximum Curvature Points (KMCPs), which are the major contributors to the gesture shape. As a result, the spatial coordinates of the KMCPs in the 8 Octant Regions of the 2D Space that are given for classification yield a translation and scale-invariant feature vector. The accuracy of numbers was 93.2 percent, and the accuracy of alphabets was 91.83 percent.

In paper [7], the authors proposed a system that can recognize and convert ISL gestures from a video feed into English voice and text. They did this by segmenting the shapes in the video stream using several image processing techniques such as edge detection, wavelet transform, and picture fusion. Ellipsoidal Fourier descriptors were used to extract shape features, while PCA was utilized to optimize and reduce the feature set. The fuzzy inference system was used to train the system, which resulted in a 91% accuracy. In paper [8], the authors suggested a method for automatically recognizing Indian sign language gestures. The proposed method employs digital image processing techniques and uses the YCbCr color space for hand detection, with the input image being transformed beforehand. Distance transformation, Projection of distance transformation coefficients, Fourier descriptors, and feature vectors are some of the techniques used to extract the features. An artificial neural network was used to classify the data, and the recognition rate was 91.11 percent.

3. SIGN RECOGNITION SYSTEM APPROACHES

Sign language recognition system is classified into two categories: Sensor-based approach and Vision-based approach. Sensor-based gloves are employed in the first approach. Different sensors are used to collect data from the gesture. After then, the data is examined, and conclusions are drawn based on the findings. The problem of this approach is that the sensor glove must be carried with you at all times, i.e., the individual must wear the sensor as well as the glove for the system to work, and sensor-based gloves are fairly

Table -1: Comparison of Various Sign Language Recognition Systems

Ref No.	Input	Description	Feature Vector	Classification	Recognition Rate
[3]	Images	Processed and converted the images into a set of features that comprised of some length measures which indicated the fingertip's position.	A priori model of skin color	subtractive clustering algorithm and the least-squares estimator	95.55%
[4]	Videos	Haar-like algorithm to track the hand in the video frames and preprocessing techniques like skin detection and size normalization to extract the region of interest	Fourier Transformation	k-Nearest Neighbor (KNN)	90.55%
[6]	Images	Used the Maximum Curvature Points (MCPs) as Control points, to approximate the extracted boundary from the region of interest to a B-Spline curve.	A novel method based on B-spline approximation	Support Vector Machine (SVM)	Numbers-93.2% Alphabets-91.83%
[7]	Video	Segmented the shapes in the video stream using several image processing techniques such as edge detection, wavelet transform, and picture fusion.	Elliptical Fourier Descriptors	Fuzzy Inference System	91%
[8]	Images	Employs digital image processing techniques and uses the YCbCr color space for hand detection,	Distance transformation, Projection of distance transformation coefficients, Fourier descriptors, and feature vectors	Artificial Neural Network	91.11%

expensive. Cameras are utilized to capture images of the gesture as an input for the recognition system in the vision-based approach. To detect the gesture, this system employs a variety of image processing algorithms. The vision-based sign recognition systems are further categorized into appearance-based and 3D model-based [1].

Images or videos are utilized as inputs in appearance-based recognition, and features are derived from them. These characteristics are then matched to photos that have already been saved. The hand's 3D descriptors are used in the 3D model-based method. This method entails looking for the kinematic parameters that correspond to a 2D projection of a 3D model of a hand and an edge-based image of a hand [2].

4. OVERVIEW OF THE INDIAN SIGN LANGUAGE RECOGNITION SYSTEM

The goal of a sign language recognition system is to accurately recreate speech or text based on a given sign. Table 1 shows the comparison of various Sign Language recognition systems while Table 2 describes the positives and negatives of the various sign language systems. The important steps involved in the sign language recognition system include preprocessing, feature extraction, and classification.

4.1 Pre-processing

To extract the region of interest from the training images, a pre-processing step is performed (ROI). If only hand motions are taken into account, the ROI can be hands, or face and hands if facial gestures are also taken into account. Filtering, picture enhancement, image resizing, segmentation, and morphological filtering are common pre-processing steps. Any of the generally used methods for filtering and image enhancement can be employed. Otsu's thresholding, background subtraction, skin color-based segmentation, and motion-based segmentation are some of the algorithms utilized for segmentation. The test photos or videos are also pre-processed to extract the region of interest during the testing phase.

4.2 Feature Extraction

The feature vectors obtained from this step are the inputs to the classifier, so feature extraction is one of the most important steps in sign language recognition. The feature extraction approaches should be able to discover shapes accurately regardless of changes in lighting levels, orientation, and size of the object in a video/image. Wavelet decomposition, Haar wavelets, Haar-like features, orientation histogram, scale invariant feature transform, Fourier descriptors, and other methods can be used to produce the features. The classifier is then trained using the feature vector acquired using any of the feature extraction methods.

4.3 Classification

In order to classify the input signals into distinct classes, a classifier is required in sign language recognition. During the training phase, the feature vector obtained from the training database is used to train the classifier. When a test input is presented, the trained classifier recognizes the sign's class and displays or plays the appropriate text or sound. Images or videos can be used as test inputs.

Machine learning techniques for classification are classified into two types: supervised and unsupervised. Supervised machine learning is a method of teaching a computer to detect patterns in input data that can subsequently be used to predict future data. Supervised machine learning applies a collection of known training data to labeled training data to infer a function. Unsupervised machine learning is used to make inferences from datasets with no tagged response. Because no labeled response is offered to the classifier, there is no reward or punishment weightage to which classes the data is meant to go.

Hidden Markov Models (HMM), Artificial Neural Networks (ANN), Multiclass Support Vector Machines (SVM), Fuzzy Systems, and K Nearest Neighbor (KNN) are some of the most often used classifiers. The recognition rate is used to evaluate the classifier's performance.

i. Convolutional Neural Network

A multilayered feedforward neural network is referred to as a convolutional neural network (CNN). It's one of the most often used approaches for categorizing photographs. There are numerous convolution layers, a pooling layer, and an activation function in it. The fully connected layer is at the very end. The input to the CNN is a vector representation of the image, and the predicted label is the output. The completely linked layer has the same number of neurons as classes. It also has a loss function, which determines the algorithm's effectiveness.

ii. K-Nearest Neighbor

The k nearest neighbor algorithm is a supervised learning system that uses feature space to classify objects. The nearest-neighbor method is used by the KNN method to classify each row of data in the sample into one of the training groups. A set of features obtained during the training phase utilizing a variety of training photos makes up the category. The fundamental goal of classification is to discover the best matching features vector for the new vector among the reference features. To get the KNN in a dimensional space, it leverages feature vectors created during the training phase. A majority vote of the vector's neighbors classifies it. Neighbors are chosen from a group of items for which the correct classification has been determined.

Table -2: Positives and Negatives of Various Sign Language Recognition Systems

Ref No.	Positives	Negatives
[4]	Experimental results of the system showed that the system is capable of recognizing Arabic alphabets effectively with no need for any additional hardware such as gloves or sensors.	Due to the similarity in the representation of ArSL, large number of Fourier Descriptors were required to uniquely identify the shape.
[6]	The proposed gesture recognition system can handle different types of alphabets and number signs in a common vision-based platform. The system is suitable for complex ISL static signs. The proposed method is more advantageous when compared with the existing methods as it uses B-Spline curve to approximate the boundary which gives better recognition for the complex shapes in ISL.	The proposed gesture recognizer cannot be considered as a complete sign language recognizer, as for complete recognition of sign language, information about other body parts i.e., head, arm, facial expression etc. are essential.
[7]	Shape features of hand gestures are extracted using elliptical Fourier descriptions which to the highest degree reduces the feature vectors of an image.	Although the system obtained a recognition rate of 96%, it does not work in real-time.
[8]	The proposed algorithm has a very high accuracy and low computational complexity compared to the existing methods.	Although the algorithm gives 91% accuracy, there are 16 wrongly classified gestures.

iii. Support Vector Machine

The SVM, a common supervised learning algorithm, aids pattern recognition. Because it separates the feature space for each class, the SVM is particularly suited to dealing with unknown data. However, because it separates the feature space for each class, it is not well suited to grouping sample data. It was established in the first place to handle binary decision problems. The basic SVM takes a set of input data and predicts which of two output classes each input will produce. As a result, it's referred to as a non-probabilistic binary linear classifier. Multi-class problems are divided into two classes, each of which can be solved using a large number of SVMs. By implicitly mapping their inputs, SVMs can efficiently perform non-linear classification.

5. CONCLUSIONS

In this paper, we give a brief summary of various methods and techniques which are provided by various authors for the recognition of hand gestures. The ultimate goal of the hand gesture recognition system is to identify the language of physically impaired people as well as to build an efficient human-computer interaction system. Hand gesture is an active area of research in computer vision and must surpass current performance in terms of robustness and speed to achieve interactivity and usability. More focus must be given to extracting features that would distinguish each sign irrespective of source, color, and lighting conditions.

REFERENCES

[1] Noor Adnan Ibraheem, RafiqulZaman Khan, 2012, Survey on Various Gesture Recognition Technologies and Techniques, International Journal of Computer Applications, Volume 50, No. 7

[2] G. R. S. Murthy & R. S. Jadon, 2009. A Review of Vision Based Hand Gestures Recognition, International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405-410.

[3] Al-Jarrah, A. Halawani, "Recognition of gestures in Arabic sign language using neuro-fuzzy systems," The Journal of Artificial Intelligence 133 (2001) 117–138..

[4] Nadia R. Albelwi, Yasser M. Alginahi, "Real-Time Arabic Sign Language (ArSL) Recognition" International Conference on Communications and Information Technology 2012.

[5] Rajam, Subha & Ganesan, Balakrishnan. (2011). Real time Indian Sign Language Recognition System to aid deaf-dumb people. International Conference on Communication Technology Proceedings, ICCT. 10.1109/ICCT.2011.6157974.

[6] Geetha M., Manjusha U. C. 2013,A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, International Journal of Computer Science and Engineering (IJCSE)

[7] P.V.V Kishore, P. Rajesh Kumar, E. Kiran Kumar & S.R.C.Kishore, 2011, Video Audio Interface for

Recognizing Gestures of Indian Sign Language, International Journal of Image Processing (IJIP), Volume 5, Issue 4, 2011 pp. 479-503

[8] Adithya, V., Vinod, P. R., Gopalakrishnan, U., and Ieee (2013). "Artificial Neural Network Based Method for Indian Sign Language Recognition."

[9] http://timesofindia.indiatimes.com/home/education/news/realising-the-importance-of-indian-sign-language-dictionary/articleshow/86221166.cms

[10] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden markov models", Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.