



Available online at <http://www.advancedscientificjournal.com>
<http://www.krishmapublication.com>

IJMASRI, Vol. 3, issue 3, pp.952-961, March.-2023

<https://doi.org/10.53633/ijmasri>

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130

IBI IMPACTFACTOR 1.5

DOI: 10.53633/IJMASRI

RESEARCH ARTICLE

INDIAN SIGN LANGUAGE RECOGNITION SYSTEM

Prayas Chaudhary¹, Bhavika Jain², Sunil Maggu³ and Bhaskar Kapoor⁴

^{1,2,3,4} Department of Information Technology, Maharaja Agrasen Institute of Technology affiliated to Guru Gobind Singh Indraprastha University, Rohini, Delhi

Abstract

Sign language plays key role when it comes to addressing the social concerns of deaf-dumb community. Gesture Recognition is important to build reliable interface between user and computer. Effective research in computer vision and machine learning is conducted on the visual aspect of human-computer interaction. This research uses machine learning models to categorise Indian sign language. ASL has been the subject of extensive investigation, and there is sufficient data to analyse it. India is a diverse country, with many different regions and customs, giving rise to a wide range of communication languages. As a result, there aren't many standard data sets, and those that exist contain noise and variations. In contrast to ASL, ISL makes signals with both hands rather than just one. It results in the blockage of features, which is a significant roadblock for the lack of advancement in this area. With the aid of an ISL data set, this initiative hopes to advance the study of this area. We developed a database of sign language for alphabets and numbers. Segmentation is carried out using both background removal and skin colour. Following the extraction of SURF (Speeded Up Robust Features) features from the images, histograms are created to map the indications with the appropriate labels. Later, using picture pre-processing and the Bag of Words model, the features will be recovered from the segmented data that was obtained. These features will be fed to supervised models for classification in the last stage. The system's implementation will be a modest step toward removing the social communication barrier that exists between deaf-mute individuals and people who does not understand sign language.

Keywords: Sign language recognition; Gesture Analysis; Human Computer Interaction; Bag of Visual words(BoW); Image processing; Support Vector Machine; K-Means Clustering;

Introduction

A gesture is a movement that could be both static and dynamic. It is a type of nonverbal expression where physical movements are used to express meaning. When it comes to sharing or expressing knowledge or emotions, or when bringing individuals closer together for better understanding, communication is crucial. It is more difficult for those with disabilities, such as the deaf and dumb, to communicate using natural language. In order to communicate with one another and with the rest of the world, they use sign language. Normal individuals, however, have trouble understanding sign language since they typically lack any prior training or expertise with it. Deaf and mute people utilise sign language, which consists of visual movements and signals, to communicate. Every motion in this carefully thought-out code has a designated significance. These symbols are used for common expressions like greetings and sentences in addition to being utilised for alphabets and numbers. There are 143 different sign languages that are currently used around the globe, with the majority being American Sign Language (ASL), French Sign Language, British Sign Language, Japanese Sign Language, and Indian Sign Language (ISL).

Similar to how every nation has its unique language; sign language is not a universal language and varies from nation to nation. Given that ASL is a language that people all around the world learn, there has already been a lot of research done on its recognition. ASL is simpler than ISL since it just requires one hand to represent gestures. ISL is more complicated than ASL and employs both hands to represent gestures. There isn't as much research and development in this area as a result of this. The simple objective of this initiative is to use Indian Sign Language to build a bridge of social connection and communication between able-bodied persons and those with disabilities.

Since our research simply focuses on the alphabet and numbers in ISL, it may be expanded to include expressions and phrases that will help impaired and able-bodied people communicate and understand one other better. This project could potentially mark the turning point where technology meets humanity and helps in aiding those who suffer from hearing loss since we are living in the century where India is evolving rapidly in terms of digital and technical advancements.

Literature survey

S.No	Reference	Objective	Techniques used	Advantages	Result(s)
1.	(Sharmila Gaikward <i>et al.</i> , 2019)	The foundation of machine control and vision cooperation are hand gestures. The ASL system is the main focus of this essay. The device will offer a sign language recognition interface that makes it simple to interact with hearing-	<ul style="list-style-type: none"> • OpenCV • SIFT & SURF • CNN 	The technology not only recognises American sign language but also converts text to voice for the appropriate user.	It is the goal of the research to discern between English's several alphabets. Additionally, all of the English alphabets and digits are recognised in the project.

		impaired persons.			
2.	(Sahana <i>et al.</i> , 2020)	The ability to recognise gestures is crucial for building a solid user interface between a user and a computer. This research proposes a technique for hand gesture identification based on multiscale density data.	<ul style="list-style-type: none"> • OpenCV • Multilayer Perceptron • Support Vector Machine 	Predicts with an accuracy of 98.20%.	In this study, depth pictures of ASL numerals are taken into consideration. The resulting recognition rate is 98.20%, which is comparable to relevant state-of-the-art methods.
3.	(Adithya, and Rajesh 2020).	This study suggests a method for detecting hand gestures that is based on an powerful deep convolutional neural network (CNN) architecture. Hand motions are the main vocabulary item in sign language.	<ul style="list-style-type: none"> • ReLu • CNN 	Even with slight interclass variations, the model can automatically determine the potential traits that distinguish the hand postures. NUS hand posture dataset accuracy: 94.7% Accuracy in the use of American spelling Figure: 99.96%	In this study, the deep learning tool's strength has been used to recognise hand positions from raw photos (RGB). The suggested CNN architecture reduces the computational overhead associated with hand posture identification with traditional algorithms by doing away with the necessity for hand detection and segmentation from the collected pictures. The NUS hand posture dataset and American fingerspelling A dataset, two publicly accessible datasets, were used to test the approach, and it produced superior recognition accuracy results.
4.	(Shubendu Apoorv 2020).	The goal of this work is to create a reliable communication interpretation tool that can convert	<ul style="list-style-type: none"> • CNN • HSV • Deep Learning 	RaspberryPi-supported lightweight portable variant. It is 98% accurate.	The communication gap between those who are deaf or mute and others who cannot comprehend sign language is closed. Its utilisation will rise as a result of being made

		Indian sign language into readable output. The task is accomplished using machine learning and image processing. Our project may be used to teach various gestures to automatic systems that are gesture-based as well as for everyday communication.			portable. We can operate the gadget more smoothly and in real-time thanks to the increased speed and precision.
5.	(Katoch <i>et al.</i> , 2022)	In this research, we offer a method that recognises Indian sign language alphabets (A-Z) and numbers (0-9) in real-time video stream using the Bag of Visual Words model and outputs predicted labels as both text and voice.	<ul style="list-style-type: none"> • SURF • SVM • CNN • OpenCV 	The 36 static ISL alphabets and numerics have all been successfully taught into the system. The model can convert visual data into text or voice with 99% accuracy rate and recognise double hand signals.	The research presents a unique method for categorising and identifying Indian sign language signals (A-Z) and (0-9) using the SVM and CNN. Our effort focuses on developing a better real-time recognition tool that will make the system portable. It is accomplished by creating unique data set, turning the system into a rotation-invariant system, and fixing the background dependence problem.
6.	(Arpita Halder and Akshit Tayade 2021).	This paper's main goal is to demonstrate a methodology for recognising sign language utilising the open source MediaPipe framework and machine learning techniques. It intends to employ a method to acquire a pipeline	<ul style="list-style-type: none"> • ML Pipeline • SVM • Palm Detection Model • MediaPipe framework 	The model is sturdy and economical since it requires less computational power and may be used with smart devices. provides accuracy of 99% SVM-based faster real-time detection	This system may be efficiently customised for any regional sign language dataset in order to gain maximum accuracy, according to training and testing using different sign language datasets (American, Indian, Italian, and Turkish).

		that can be used as a sign language recognition system that is quicker, easier, cheaper, portable, and simple to implement.			
7.	(Nallusamy <i>et al.</i> , 2022)	Research in computer vision and machine learning is effectively conducted in the visual touch of human-computer interaction. Systems for recognising sign language are given priority. It will be an excellent aid for those who have hearing loss in a system that can recognise signs in real-time. A vision-based system is preferred because it offers a straightforward and precise means of interfacing between a human and a computer. This paper aims to review sign language recognition methods and to determine the best method used by researchers.	<ul style="list-style-type: none"> • OpenCV • CNN • LSTM 	Based on dimensionality reduction and calculating cost-reducing feature extraction techniques.	This study presents a check on sign language recognition and examines several approaches that have been researched and examined for the same. Segmentation, used to separate the skin region from the backdrop during the recognition process, is crucial since it typically determines the degree of recognition accuracy. Deep neural networks (CNN, Inception model, and LSTM) outperform conventional classifiers like KNN and SVM, based on a study of multiple categorization approaches.
8.	(Bantupallian and Xie 2018).	Deep learning and computer vision techniques have been	<ul style="list-style-type: none"> • RNN(Recurrent Neural 	Any length of input may be handled by RNN. Any time series	The suggested method uses video sequences to extract temporal and geographical information.

		used to make progress in the fields of motion and gesture detection thanks to recent advancements in deep learning and computer vision. The work aims to develop a vision-based application that provides text translation of sign language, in order to ease communication between signers and non-signers.	Network) • CNN	predictor that uses an RNN model will benefit. The model size does not grow even if the input size rises. 99.4% accuracy.	
9	(Arora <i>et al.</i> , 2018)	American Sign Language is intended to be recognised by the proposed system, which then translate it to text. Helping the deaf and hearing impaired connect with others who cannot comprehend sign language is one of the goals, as is expanding the use of technology to address societal issues.	• Open CV • Bhattacharya Distance Metric	It is really simple to use and put into practise. The technologies employed are all open source. Compared to its competitors, OpenCV is quicker and consumes fewer system resources. It is portable and economical.	Even when the input is only partially recorded by the camera for a variety of reasons, the system nevertheless produces correct results. Therefore, we can say that the system provides adequate outcomes with a little amount of processing.

10.	(Kozyra <i>et al.</i> , 2022)	This study describes the design and implementation of an application that uses deep learning methods built on convolutional neural network architectures to detect signals in American Sign Language.	<ul style="list-style-type: none"> • Keras • CNN • ReLu • Flask framework 	The training set's efficacy was 99%.	The research aimed to develop a convolutional artificial neural network model that could identify sign language motions from a static image. After investigation, 24 of the 40 photos were recognised.
-----	-------------------------------	---	---	--------------------------------------	--

Proposed Methodology

The system used in this work can decipher motions in Indian Sign Language from still images. As illustrated in figure 2, the system consists of multiple processes, including image collecting, image pre-processing (segmentation), feature extraction, and classification. The classifying of the photos has been done using the Bag of Visual Words (BoW) approach. The concept of BoW was inspired by natural language processing (NLP). The BoW model approach in image processing is also known as "histogram-based representation of independent features." In order to represent any gesture using the BoW paradigm, a picture may therefore be thought of as a document. It's equally crucial to describe "words" in visuals. The following three phases are often used to do this: feature description, codebook development, and code analysis (visual words). All of the photographs can have their histograms scanned using these codebooks. The Support Vector Machine (SVM) model may also be used to classify photos.

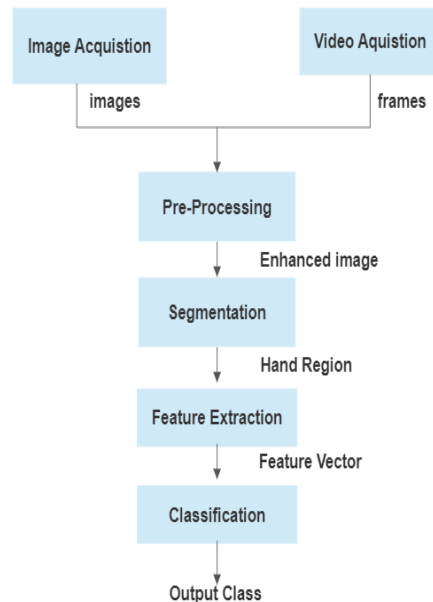


Figure 2. Overview of the system

Image Collection

Since there hasn't been enough study in this area, the right data set for ISL is currently unavailable. Therefore, we developed an ISL data set with 35 classes and 1200 photos each class.

Dataset: 35 classes total, 42000 photos in alphabetical and numeric order (A-Z). Webcam video is used to record each class's whole collection of photographs. The video's frames are all stored as individual images. Gestures were recorded against a dark background to reduce background noise.

All hand gestures fed to our model are given below in figure 1.



Figure 1. ISL gestures for numeric and alphabets

Image pre-processing

All photos are pre-processed during this stage in order to prepare them for feature extraction. Three procedures make up this phase: skin identification, edge detection, and picture segmentation (skin masking).

By converting the picture to HSV colour space, a skin mask is created from the raw images (figure 3a). All pixels between (0, 40, 30) and (43,255,254) that fall inside the (H, S, V) range are considered skin pixels. The skin mask can be used to segment skin (figure 3c). The Canny Edge approach (Adithya and Rajesh 2020) is used to detect the edges of a picture by identifying the existence of abrupt discontinuities in the image. (Figure 3d).

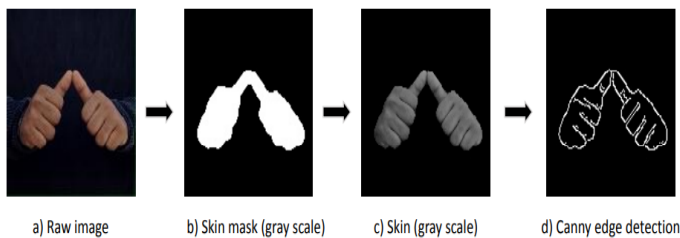


Figure 3. Image Pre-processing

Feature Extraction

The following step is feature extraction. Image feature identification, clustering, and code book production for the Bow model are the three phases in this phase. In the beginning, we discussed of using the Scale Invariant Feature Transform (SIFT) ((Nallusamy *et al.*, 2022)) technique to find important characteristics in the photos. However, we switched to SURF, an unique feature extraction approach that is resistant to scaling, rotation, variation, and perspective occlusion (Sahana *et al.*, 2020). In figure 4, you can examine SURF characteristics of a picture.

The last step is to group together all of these SURF elements that are comparable to create a visual lexicon. Finding all identical feature descriptions would probably be impossible for a person. In this phase, the K-Means clustering approach might be helpful. Unsupervised clustering algorithms like the K-means method are quite popular. It seeks to divide 'n' features into k-clusters and forecasts a new feature that belongs to a cluster based on the mean (centroid) of the clusters. K-Means clustering requires a lot of processing time and memory because we are working with enormous datasets of SURF features from 42000 photos. Therefore, we employed mini batch K-Means, a comparable algorithm to K-means that has benefits over K-means when it comes to processing speed and memory usage. After all SURF features have been trained to mini batch K means, the number of clusters (visual words) in the bag that include all the comparable features is equal to k. Since there are 35 classes to categorise, we utilised the k value of 280. Visual words for each image may be predicted using this technique.

The subsequent phase will use these anticipated visual words to compute histograms. The frequency of recurrence of each visual word associated with a picture in all visual words may be used to create the histogram.

Classification

The next stage is categorization when all histograms for the whole data set have been created. It's crucial to separate the data into training and testing sets before moving on to classification.

Therefore, we divided the total data in half (80:20), giving each class 960 photos for training and 240 for testing. After the data has been prepared, the training data will be fed to machine learning model. In order to train data and forecast it using a linear kernel, we employed a support vector machine. Similar to this, other models, including K-Nearest Neighbours, Convolution Neural Network, Logistic Regression classifier, etc. can also be utilised for predictions. Real-time gesture prediction is now possible thanks to the development of real-time gesture recognition.

applications by employing a broad and diverse selection of photos in the dataset.

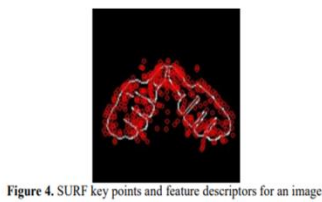


Figure 4. SURF key points and feature descriptors for an image.

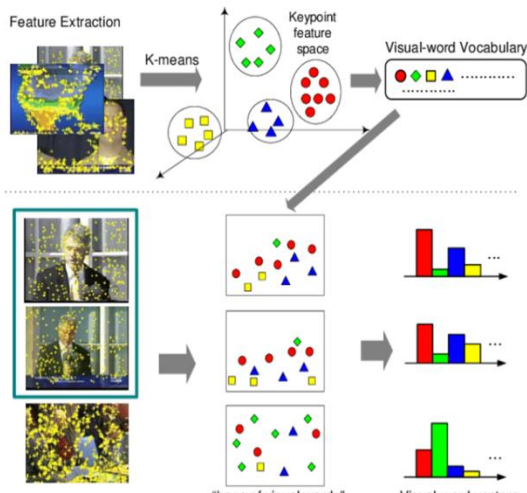


Figure 5. Bag of visual words representation. Above: Training for the K-means model using all features (SURF features) Below: Prediction of visual words and representing histograms for each image. [6]

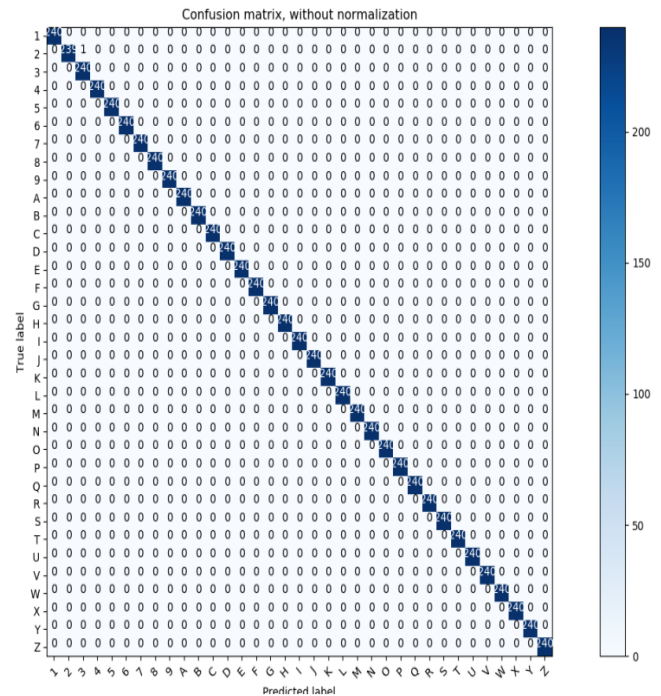


Figure 6. Confusion matrix using SVM model

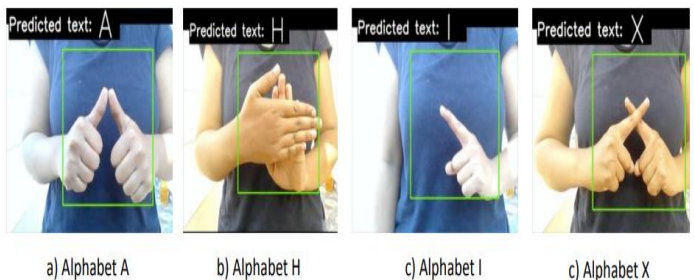


Figure 7. Real time recognition of gestures results

Result and Discussion

The model achieved 99% accuracy using Bow coupled with strong SURF feature descriptors. In picture 6, you can see the model's confusion matrix. Except for label 2, all labels were properly predicted by the SVM. Figure 7 displays real-time recognition prediction outcomes. A 99.98% accuracy, recall, and f1 score were also recorded. However, given that the data set contains many similarly styled photographs without changes in, for instance, light or skin tone, the model prediction may be slightly biased. So, this strategy can be more reliable for real-world

Conclusion

India's deaf and dumb population relies heavily on ISL for communication. This work provides a thorough implementation of the Bag of Words concept for recognition of the Indian sign language. Image collecting, image pre-processing, feature extraction (using K-means clustering, visual words collection), and classification have all been covered step-by-step in section 3. In section 4, the results of the Bow model were finally revealed. Real-time gesture detection has also advanced beyond the realm of static images. In future,

this project can be expanded to include basic ISL words and phrases, such as alphabets and numbers.

Reference

1. Sharmila Gaikward., Shetty, A. Akshaya Satam and Rathod, M. (2019). "Recognition of American Sign Language using Image Processing and Machine Learning", *IJCSMC*, Vol. 8, Issue. 3; March 2019, pg 352-357, ISSN 2320-088X.
2. Sahana, T., Soumi Paul. Subhadip Basu and Ayatullah Faruk, M. (2020). "Hand sign recognition from depth images with multi-scale density features for deaf mute persons", *Procedia Computer Science*, Volume 167, 2020, Pages 2043-2050, <https://doi.org/10.1016/j.procs.2020.03.243>.
3. Adithya, V and Rajesh, R. (2020). A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition, *Procedia Computer Science*, Volume 171, 2020, Pages 2353-2361, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.04.255>
4. Shubhendu Apoorv (2020). *IOP Conf. Ser.: Mater. Sci. Eng.* **872** 012026, doi = {10.1088/1757-899X/872/1/012026}.
5. Katoch, S., Varsha Singh. Uma Shanker Tiwary. (2022). Indian Sign Language recognition system using SURF with SVM and CNN, *Array*, Volume 14, 2022, 100141, ISSN 2590-0056, <https://doi.org/10.1016/j.array.2022.100141>.
6. Arpita Halder and Akshit Tayade. (2021). "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning", *International Journal of Research Publication and Reviews* Vol (2) Issue (5) (2021), Page 9-17, ISSN 2582-7421.
7. Nallusamy, C., Ari Haran, A. Arun, P.K. and Sabith, K.N. (2022). Sign language recognition. *International journal of health sciences*, Volume 7, Issue 6, ISSN: 2456-3315.
8. Bantupalli, K and Xie, Y. (2018). "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.
9. Arora., Sandhya. Roy. Ananya. (2018). Recognition of sign language using image processing. *International Journal of Business Intelligence and Data Mining*. 13. 163. 10.1504/IJBIDM.2018.088428.
10. Kozyra, K., Trzyniec, K. Popardowski, E and Stachurska, M. (2022). Application for Recognizing Sign Language Gestures Based on an Artificial Neural Network. *Sensors* 2022, 22, 9864. <https://doi.org/10.3390/s22249864>.
11. Shivashankara, S. and Srinath. (2017). A comparative study of various techniques and outcomes of recognizing american sign language: a review. *International Journal of Scientific Research Engineering & Technology (IJSRET)*, 6(9), pp.1013-1023.
12. Suharjito., Gunawan Herman. Thiracitta Narada and Nugroho Ariadi. (2018). The 1st 2018 INAPR International Conference (Jakarta, Indonesia) Sign Language Recognition Using Modified Convolutional Neural Network Model 978-1-5386-9422-0/18/\$31.00 ©2018 IEEE
