International Journal of Computer Sciences and Engineering
Vol.**11**, Special Issue.**1**, pp.**155-160**, November **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Holistic Approach of Indian Sign Language Prediction Software with Emotion Detection

Dipankar Mazumder[1]*, Upamita Das[2], Hillal Kumar Roy[3], Nilava Sarkar[4], Abhishek Kumar Singh[5]

[1,2,3,4,5]JIS University, Kolkata, India

Corresponding Author: dipankarmazumder9831@gmail.com

Abstract: A real-time AI software solution for a holistic approach to recognizing Indian Sign Language (ISL) where elements of ISL such as hand shape, facial expression, orientation, movement etc. are analyzed, recognized, and converted into written text. Sentencesare formed by analyzing each sign one by one and overlapping detections are ignored. It is a software solution that a user can run on their system without installing any dependencies. We also use emotion detection to understand what a person is trying to say as any human being will have emotions while they convey their message. The model is also trained with an ideal state whereif no signs are being shown, that is if there are no hand movements, no sign is predicted.

Keywords: Mediapipe, LSTM, CV2, Indian Sign Language, DeepFace, PyInstaller, Keras, CNN

1. Introduction

Over the years Communication between human beings is categorized by languages. There are so many disabled people over there who lost or never could communicate through human language. For those disabled people sign language is one of the most important ways to communicate. Currently, that is as of 2023, more than 1.5 billion people (nearly 20% of the global population, i.e., 8 billion) live with hearing loss. In a country like ours, not many of the population have access to education or even computers [1]. Making an easy-to-use and intuitive software for the Indian deaf is crucial. Despite common misconceptions, sign languages are natural languages, just like languages like English or Hindi, with their syntax, grammar, emotional sense in sentences etc. [3]. Sign language is essentially made up of 5 components. These 5 components can be represented with the mnemonics HOLME which stands for Hand shape, Orientation, Location, Movement and Expression. Thus, we need to analyse all these components to properly analyse what someone means by showing a sign [9].

1.1 The model and the use of Mediapipe

Solving Artificial Intelligence for cognitive problems is a very trending topic to achieve in this field. Thus, using AI was our first approach towards this problem, where all components of the sign language, i.e., HOLME, needs to be analysed to give an output that is influenced by all the components. A Long Short-Term Memory (LSTM) model has been used to predict the signs. LSTM models are generally used to work on data which is sequential because the networks can learn long-term dependencies, which is exactly what we need for our application [13]. LSTM is a recurrent neural network and since our data stream is a sequential one, which is a real-time video, i.e., a sequence of frames or images, LSTM was an apt choice.

Now choosing a model isn't enough because the model will of course need data to work with LSTM is good for video analysis, but we do not train the model with pixels of data. Here comes the use of Mediapipe in our software. MediaPipe is a powerful and versatile machine-learning framework that can be used to build complex and multimodal applied machine- learning pipelines. It is an open-source framework that provides a unified platform for developing machine learning models for various platforms such as Android, iOS, desktop, edge, cloud, web, and IoT platforms [4]. We use Mediapipe toextract landmarks of the human pose that is being shown to the camera sensor. Landmarks are the vector coordinates of different parts of your body if the camera input is the cartesianplane. Mediapipe is so accurate in its job that it can track simultaneously and semantically 33 poses, 21 per hand, and 468 facial landmarks. Fig. 1 shows Mediapipe detecting landmarks on the body in the video feed. We use this continuous stream of vector coordinates to train our model for different signs [2]. As discussed earlier, our main focus while making this software is analysing every component of sign language properly, a whole-body tracking system needed to be implemented and this is done using Mediapipe.

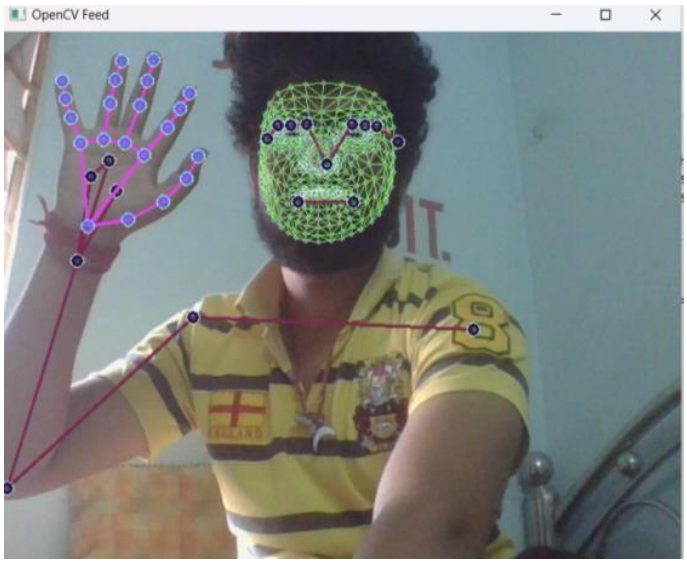

Figure 1. Mediapipe detecting landmarks

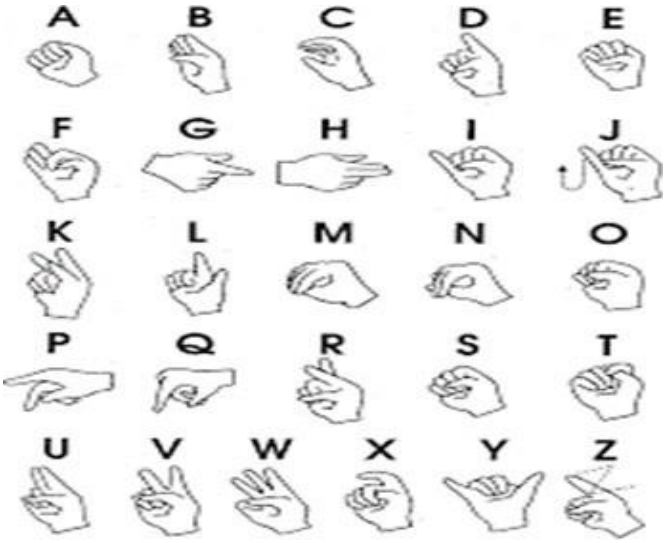

Figure 2. Handshapes for the classifier to identify

We used OpenCV and Numpy to collect data from the camera. OpenCV is an open-source computer vision and machine learning software library that provides a common infrastructure for computer vision applications and accelerates the use of machine perception in commercial products. For our purpose, we take the frame data using this library and feed it to the Mediapipe library which is compatible with the CV2 format [5].

1.2 Facial emotion recognition with DeepFace

Facial expression recognition is essential as its one of the components of sign language. This is detected in our software using the library called Deepface. Deepface for Python is a framework that allows you to perform various tasks related to face recognition and facial attribute analysis [3]. Deepface for Python is an open-source project that was inspired by the research done by Facebook on deep learning for face verification. It uses the HaarCascade to detect faces and is a technique for object detection that uses a cascade of simple features to rapidly identify objects in images. We detect what the person showing the sign's facial expression means such as angry, sad, happy, excited or even drowsy [7].

1.3 Alphabet classification

Apart from this holistic approach, there is also an implementation of alphabet recognition using the handshape. We use a classifier to detect handshapes and train the model using several handshapes that correspond to the respective alphabet. In Figure 2 the handshapes are depicted that the classifier can classify. This is required since many words in English may not have a particular sign. In sign languages, for such signs, we use the alphabet to make the deaf person understand what is the word. For example, someone's name will of course have no particular sign as that word doesn't even exist in a dictionary [6]. For such cases, this classification technique is useful.

1.4 The use of PyInstaller

Finally, this model was trained and saved in our hard drives locally which made running the software in other systems impossible. Running the predictor in this form required extensive technical knowledge of coding as dependencies needed to be handled. All of this needed to be packaged into a single executable file. This was done using the help of a program called PyInstaller. This software is a standalone and easy-to-use solution for sign language prediction [11]. The user just needs to click the EXE file and without any hassle, the prediction window opens up.

2. Related Work

American Sign Language (ASL) recognition based on Hough transform and neural networks by Munib Q., Habeeb M., Takruri B. and Al-Malik H. A. In the proposed paper it is aimed to develop a system that can translate gestures of the alphabet and signs into American Sign Language. Hough transform and neural networks are used to build the system for recognizing signs. This system was tested by 300 samples of sign images and for each sign, 15 individual images were taken.[1]

Implementation of Indian Sign Language in Educational Settings by Zeshan U., Vasishta M. M. and Sethna M. This is the first article or comprehensive effort on Indian sign language at a national level. The importance of Indian Sign Language or its recognition to the world was at a basic level when this article was published. There are several programs like instruction for hearing Indian Sign Language for people, Training of sign language by professionals, and study material for deaf people's schooling were mentioned in this article.[2]

Sign Language in India: regional variation within the deaf population by Vasishta M., Woodward J. and Wilson K in this paper is focused that as per research on the similarity between Indian sign language and Pakistani sign language. this paper examines the relationship between all of the varieties of Indian and Pakistani sign language by analyzing comparative lexical data in Nepal by comparing these three countries. Special paperwork has been done in this paper where the vocabulary

list of sign languages to compare the cognates for the variety of sign languages in Kathmandu, Karachi, Bangalore, Delhi, Bombay, Calcutta.[3]

Design and Development of a Frame-Based MT System for English-to-ISL by Suryapriya A. K., Sumam S. and Idicula M. This paper represents the architecture and framework for speech-to-sign language machine translation systems in the railways and banking domain. This project wills to help the deaf and disabled people with the help of AI(Artificial Intelligence). This system implements a 3D animation from previously recorded motion data.[4]

Object-Based Key Frame Selection for Hand Gesture Recognition by Kshirsagar K. P. and Doye D. This article mainly represents an object-based key frame selection. For the hand gesture recognition Hausdorff distance, Forward Algorithm and Euclidean distance are used in this project. This project wills to use the hidden Markov model and nonlinear time alignment model with a keyframe selection facility and gesture trajectory feature for hand gesture recognition for better performance.[5]

Real-time Ukrainian sign language recognition system by Davydov M. V., Nikolski I. V. and Pasichnyk V. This article, focuses on computerized real-time Ukrainian sign language. The proposed systems use different approaches to detect and recognize the hand shape in motion. It used the fingertip's location and pseudo-2-dimensional image deformation model for hand movement recognition.[6]

Arabic sign language recognition in user-independent mode by Shanableh T. and Assaleh K. This article represents a method to recognize the isolated Arabic sign language gestures in a user-independent mode. In this project, signers wearing gloves plays a significant role to simplify recognizing the hand movements via color segmentations. To filter out another source of motion this system used a covered box.[7]

A hand gesture recognition system based on local linear embedding by Xiaolong T., Bian W., Weiwei Y. and Chongqing Liu. This paper aims to implement a real-time vision system under the visual interaction surrounding hand movement recognition. The most important part of this system is a feature extraction process with the help of a local linear embedding method.[8]

A unified tensor framework for face recognition by Rana, S, Liu, W., Lazarescu, M and Venkatesh, S. This paper approaches a new optimization framework that combines some of the methods that are tensor based for facial recognition based on a common mathematical approach. [9]

A Sign Language Recognition Based on Tensor by Wang S., Zhang D., Jia C., Zhang N., Zhou C. and Zhang L. In this paper tensor subspace analysis is used to model a hand gesture which is multi-viewed for recognizing all the alphabets. Two experiments were conducted, one was on grey-scale images and one was on binary images. As a result, it was shown that the system well performed in multi-view.[10].

3. Theory

Any AI model has attributes that define how it will be trained. For our LSTM model, we have tried several combinations of values of attributes and have kept that gave the best accuracy. The learning rate of the model is kept at 0.001 and the number of epochs trained is 350. The batch size of the model is 32[12]. The model structure is shown in Figure 3. The accuracy that the model has is 99%.

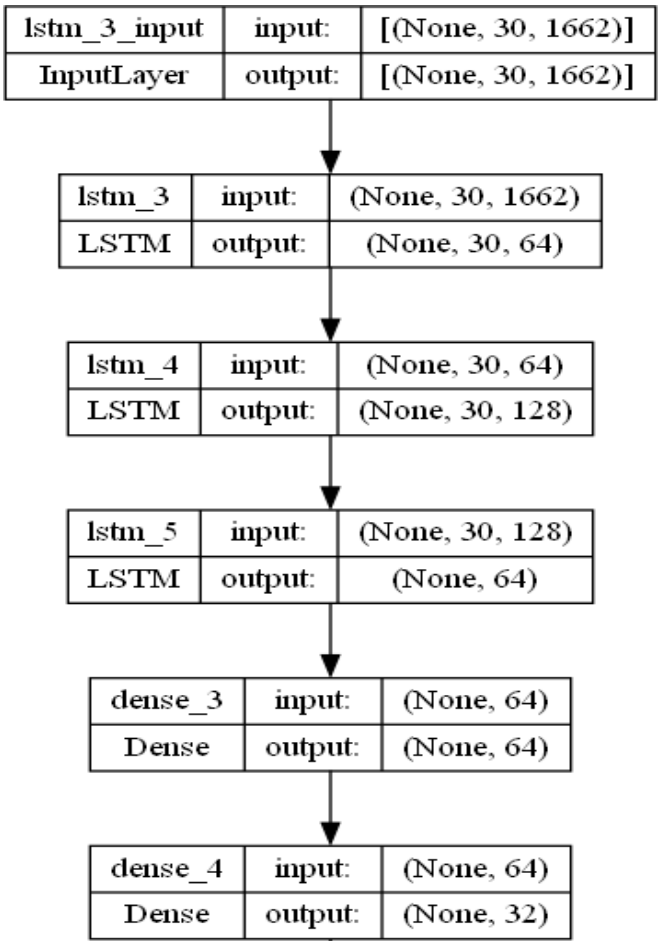

Figure 3. The LSTM model used

3.1 Description of the LSTM model

In the figure, we can see that the input shape to the model is 30 by 1662. Here, 1662 is the number of pixels each frame of the video contains and 30 represents the number of frames that the model inputs for output. The output layer has 32 possible outputs and one of which is an 'ideal' state. This ideal state has been added so that when no signs are being shown to the model, it must predict a state where no hand movements are there [14]. While training, this state was trained where no hand movements were shown. The other 31 outputs are the test signs that we have trained the model for.

3.2 The dataset creation and collection

As for the dataset creation, we collected data for each sign which is worth 30 frames 50 times. Thus, the same sign was done 50 times repeatedly and saved to folders. For this reason, no preprocessing was necessary as the dataset is created by us

only. Once we created the dataset, using Scikit Learn, we divided it into training and testing datasets where 5% of it went towards the testing dataset. The training dataset was given to the model as input [15].
As for the classification of handshapes, a sequential model is used. It is a CNN model which is generic. The MNIST dataset is used to train the model.

4. Procedure

First, the Indian Signs for particular words are obtained via different sources such as the Internet. All 5 components of sign language are understood for the words. Now in the early stage, for better accuracy, we the creator of the software only performed signs of a few words for training [16].

4.1 Folder directory for data collection

Each sign was recorded 50 times. For better space management, we directly saved the Mediapipe vector coordinates for each frame. Thus, the folder management and order are shown in Figure 4.

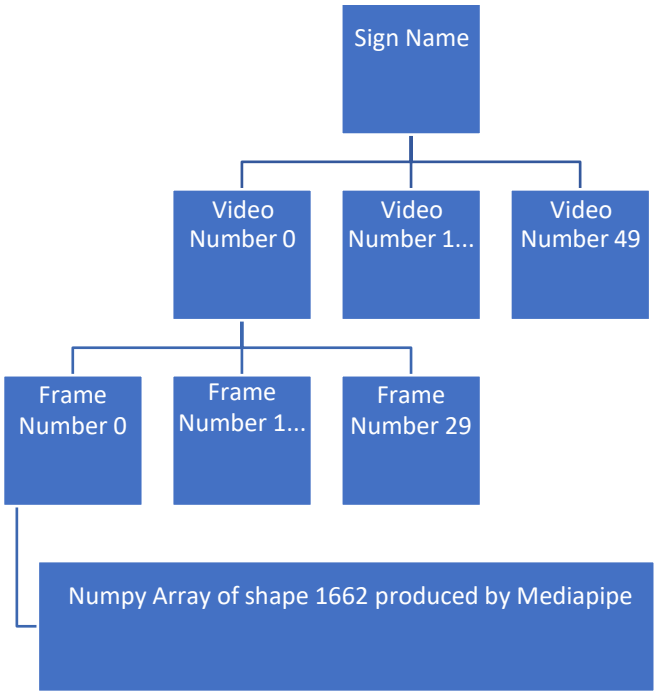

Figure 4. Folder directory for data collection

4.2 The working of the model

Once the data was collected, we used it to train the model using the specification provided in the Theory section. Once training was done, we exposed the model to real-time video and let it predict the signs we were showing. The model was predicting any of the 36 signs we trained it for. The percentage of output the model was predicting for a shown sign was displayed.

Also, the sentence that the user is trying to make, i.e., consecutive sign predictions, were shown on the screen. Any repetitive predictions were eliminated and filtered out, not shown to the user in the sentence formation. The flow diagram of working of our project is shown in Figure 5.

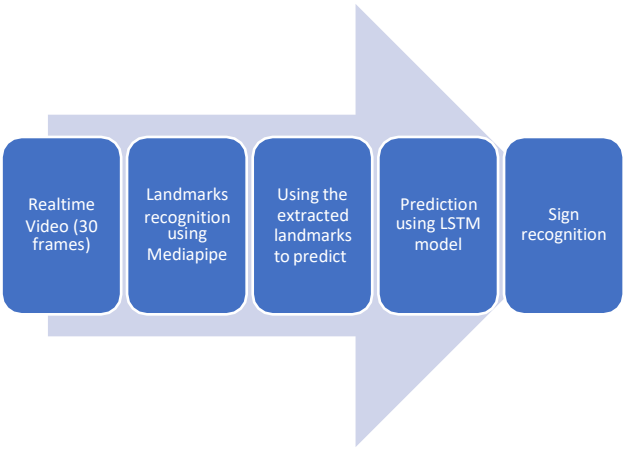

Figure 5. Flowchart of our model working

5. Results and Discussion

In this project, we have trained our model with 26 hand gestures or movements for 26 alphabets, 150 words with 150 hand and facial gestures with 30 frames from different views and angles for individual alphabets or words. More than 100 epochs got executed to increase the accuracy of the model. As a current state, our model can recognize both emotions and gestures using Mediapipe with 99% accuracy. At first for real-time recognition, the result of the model was not appropriate because in real-time hand gestures and facial expressions were not placed exactly at the center and aligned vertically. So, to overcome this shortcoming we trained the model by augmenting the dataset and after that, the accuracy of the model was up to the mark.

As we increased our training data frames, the accuracy got higher both during training and real-time implementation because the parameters of the training dataset got increased. As a finishing step of our project, we have successfully implemented a sign language recognition system in real-time. Some real-time testing instances of the proposed model are shown in the following figures.

Figure 5. Training of our proposed model

6. Conclusion and Future Scope

This paper aims to build software that makes things easy to communicate for disabled people in the world and vice versa. This software is capable to convert sign language into plain text with huge accuracy. To be honest, it's hard for a person to learn sign language, considering the large geographical area of the world sign language gets changed for different areas, so it will be the bridge between those deaf or mute people who face difficulties communicating with other people. Though there are human translators for this job our software will provide the best result with less error and more accuracy. It will make life better for deaf and mute people. Body tracking and facial emotion detection combinedly predict all components of sign language, thus understanding the meaning of a sign that has never been understood by computers before. The easy-to-use feature of this software is helpful for people who have little to no knowledge of computers.

For future work, we would like to do the followings mentioned-

- Upgradations: upgrade or modify this proposed sign language recognition software so that it would be able to detect the facial expression and emotions of humans and then it can convert those into text.
- Readability: Readability can be increased. The more readability, the easier it will be for the user to use.
- Better Picture Quality & More Dimension Add: The picture quality and no of frames can be increased so that it can cover the input from every angle.
- Working with More Sign Languages: For this project, we worked with Indian sign language so in future we can work with other sign languages as well.
- Increase in Dataset: More training data or a really large number of training datasets can be implemented and added.
- Conversion from static to dynamic: The proposed software is mainly focused on static signs/ manual signs/ alphabets/ numerals it can be shifted to continuous or dynamic signs and nonverbal types of communication in future.

Data Availability

The entire dataset is created by us and thus not available for public use. Hand gesture recognition uses the famous MNIST dataset that is easily available in Kaggle and even in Tensorflow.

Conflict of Interest

We do not have any conflict of interest during the making of this project.

Funding Source

None

Authors' Contributions

Author-1 Whole idea of the project, Implementation, Publication, and Paperwork.
Author-2 involved in the development of the software,
Author-3 involved in the development of the software,
Author-4 wrote the draft of the manuscript and research model training.
Author-5 Collecting datasets and publication.

Acknowledgements

We wish to express our sincere appreciation to all those who have contributed to this thesis and supported us in one way or the other during this amazing journey. We express our heartfelt thanks to Mr Dharmpal Singh, HOD of B.Tech CSE, JIS University, Kolkata, India, for giving us this opportunity to do this project. Special thanks to Ms Debmitra Ghosh and Mr Saumya Majumdar, our mentor, for helping us with our project.

References

[1]. Munib Q., Habeeb M., Takruri B. and Al-Malik H. "A. American Sign Language (ASL) recognition is based on Hough transform and neural networks", Expert Systems with Applications, Vol.**32**, pp.**24-37, 2007.**

[2]. Zeshan U., Vasishta M. M. and Sethna M, "Implementation of Indian Sign Language in Educational Settings", Asia Pacific Disability Rehabilitation Journal. Vol.**1**, pp.**16-40,2005.**

[3]. Vasishta M., Woodward J. and Wilson K, "Sign language in India: regional variation within the deaf population", Indian Journal of Applied Linguistics. Vol.**4**, Issue.**2**, pp.**66-74, 1978.**

[4]. Suryapriya A. K., Sumam S. and Idicula M, "Design and Development of a Frame-Based MT System for English- to-ISL", World Congress on Nature and Biologically Inspired Computing. pp.**1382-1387, 2009.**

[5]. Kshirsagar K. P. and Doye D, "Object-Based Key Frame Selection for Hand Gesture Recognition", Advances in Recent Technologies in Communication and Computing (ARTCom) International Conference on. pp.**181-185, 2010.**

[6]. Davydov M. V., Nikolski I. V. and Pasichnyk V. V, "Real-time Ukrainian sign language recognition system", Intelligent Computing and Intelligent Systems (ICIS), IEEE International Conference on , pp.**875-879, 2010.**

[7]. Shanableh T. and Assaleh K, "Arabic sign language recognition in user-independent mode", Intelligent and Advanced Systems ICIAS 2007 International Conferenceon. pp.**597-600, 2007.**

[8]. Xiaolong T., Bian W., Weiwei Y. and Chongqing Liu, "A hand gesture recognition system based on locally linear embedding", Journal of Visual Languages and amp; Computing, pp.**442-454, 2005.**

[9]. Rana, S, Liu, W., Lazarescu, M and Venkatesh, S, "A unified tensor framework for face recognition", Pattern Recognition, First edition, ELSEVIER Publisher, Australia, pp.**2850-2862, 2009.**

[10]. Wang S., Zhang D., Jia C., Zhang N., Zhou C. and Zhang L, "A Sign Language Recognition Based on Tensor. Multimedia and Information Technology (MMIT) Second International Conference on. Vol.2, pp.**192-195, 2009.**

AUTHORS PROFILE

Dipankar Mazumder is a graduate student in the Department of Computer Science and Engineering at JIS University. He has an interest in ML, AL and programming.

Hillal Kumar Roy is a graduate student in the Department of Computer Science and Engineering at JIS University.

Nilava Sarkar is a graduate student in the Department of Computer Science and Engineering at JIS University. They have a strong background in machine learning, deep learning, AI, and cloud computing. They have completed several courses in these fields and have hands-on experience in implementing and optimizing these techniques for various applications. Their current research focuses on developing machine-learning models for large-scale distributed systems and exploring the intersection of machine learning and cloudcomputing.

Upamita Das is an undergraduate student in the Department of Computer Science and Engineering at JIS University, West Bengal, India. She has published her papers in IOCER 2019 and CICBA 2023. She is an Assistant Engineer Intern at Ericsson Global India Private Limited. She has an interest in AI, ML, Cloud Computing and PC System design and optimization.

Abhishek Gupta is currently pursuing B.Tech in Computer Science and Engineering from JIS University, West Bengal, India. He is a member of ISTE, IETE. His research interests include Cyber-Physical Systems, Deep Learning, Robotics, Neutrosophic logic and cyber security.

