International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 03 | Mar 2022 www.irjet.net p-ISSN: 2395-0072

Dynamic Hand Gesture Recognition for Indian Sign Language: A Review

Bhoomi Lodaya[1], Dr. Narendra Patel[2], Dr. Hemant Vasava[3]

[1]P.G, Student, Department of Computer Engineering, Birla Vishvakarma Mahavidyalaya, Ananad, Gujarat, India
[2,3]Proffesors, Department of Computer Engineering, Birla Vishvakarma Mahavidyalaya, Ananad, Gujarat, India

---

Abstract - *The hand gesture recognition technique in use during human–computer interaction frequently. The hand gesture recognition system is designed such that it does not require any special hardware other than a webcam. Human-Computer Interaction (HCI) is the study of interaction between users and computers. Communication provides interaction among people to exchange feelings and ideas. The deaf community suffers a lot when trying to interact with the community. Sign language is the way through which people communicate with each other. In order to provide interaction with normal people, there is a system that can convert the sign language to an understandable form. The purpose of this work is to provide a real-time system that can convert Indian Sign Language (ISL) to text. Most of the work is based on handcrafted features. Some of the papers used one unified architecture by combining Convolutional Neural Network (CNN) with a Long Short-Term Memory (LSTM) network in order to recognise dynamic hand gestures for Indian Sign Language (ISL). This review paper discuss various approaches used to recognise dynamic hand gestures for Indian Sign Language.*

Key Words: **Hand Gesture Recognition, Indian Sing Language (ISL), Deep Learning, Convolution Neural Network (CNN), Long Short Term Memory (LSTM)**

1. INTRODUCTION

Recognition of sign language is a modern research area in the field of Human-Centered Computing (HCC). Its goal is to recognize gestures in sign language and translate them into text or voice. Sign language consists of gestures having a variety of positions, orientations, and movements of the palm, arm, body, and face region. All the alphabets, numbers, and words of our language are assigned such gestures. There are many sign languages all over the world, like Indian Sign Language (ISL), American Sign Language (ASL), Japanese Sign Language (JSL), and British Sign Language (BSL) etc. In all these sign languages, hand gestures play an important role as they cover the majority of all gestures. Hand gestures can be modelled using two approaches: physical sensor-based gesture modelling and computer vision-based gesture modelling. Physical sensor based hand gesture modelling uses physical sensors such as flex sensors, accelerometers, and gyroscopes for accumulating the data. The devices used in conjunction with physical sensors are single-board computers such as Raspberry Pi and Arduinos for processing the data. These devices are not easy to configure and calibrate. Furthermore, because devices are fragile, they must be handled with extreme caution. Vision-based technique, on the other hand, uses a camera and an algorithm to model hand gestures. It mitigates the human vision system's ability to accumulate and interpret the gestures. Proposals of various fast algorithms, the introduction of multiple smaller parallel computational units such as GPUs, and the availability of large datasets such as Kaggle have paved the path for computer vision approaches to be used efficiently for modelling and recognizing hand gestures.

The extraction of relevant features capable of classifying images into our desired classes is the starting point for the majority of image processing problems. The main problem with choosing a handcrafted feature is that whenever new classes are added, the system has to choose other methods. As a solution to this problem, in the 1990's, the LeNet architecture was introduced, which runs on the convolutional architecture.

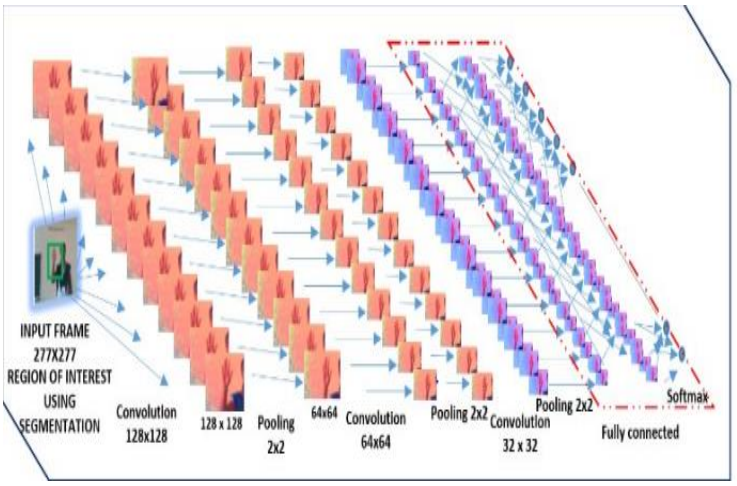

Fig.1. Convolutional neural networks (CNN) Layers.[13]

Now, computer vision-based applications used for processing hand gestures can be run on personal computers too, thereby attracting more research to be done in this field. Many approaches have been proposed for the classification of static and dynamic hand gestures. Among them, convolutional neural networks (CNN) have emerged as one of the promising techniques in the field of pattern recognition and image classification. CNN is equivalent to the traditional artificial neural networks. CNNs allow us to use

image-specific features embedded in the network, making it suitable for video Dynamic Hand Gesture Recognition for Indian Sign Language and digital images. It provides focused operations, along with a reduction in the trainable parameters as compared to ANNs. Image data needs a large amount of computational resources if processed by ANNs. Operations such as pooling and convolution introduce generalisation and reduce dimensionality in the network. Recent advances and the introduction of different dense CNN architectures result in an increase in the efficiency and accuracy of dynamic hand gesture recognition systems. The architecture has been tested for three different ISL datasets, and it accurately and efficiently classifies dynamic gestures with the help of an integrated model that combines Convolutional Neural Network (CNN) with Long-Short Term Memory (LSTM) having recurrent layers.

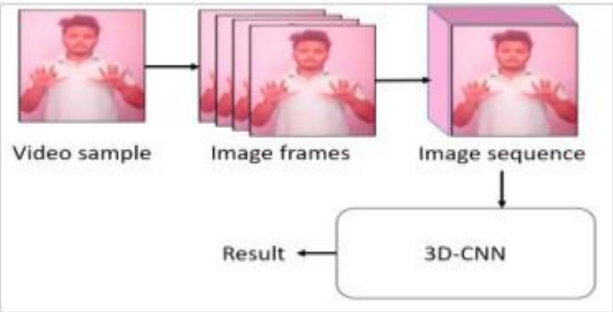

Fig. 2. Flow of image data in the system.[5]

2. Review of Dynamic Hand Gesture Recognition

Computer Vision has become one of the trending technologies used in most AI-based systems such as robots, cars, markets, etc. The system has more impact on image classification problems and object detection. The sign language system can be implanted using this method. There are many other methods used in the earlier systems.

Neel Kamal Bhagat [4] proposed a real-time model for ISL gesture recognition based on the incoming image data from the Kinect. Effective real-time background subtraction was done using depth perception techniques. Computer vision techniques were used to achieve a one-to-one mapping between the depth and the RGB pixels. Custom datasets were generated and different models were used for training. The depth + segmented static model achieves an accuracy of 98.81% and the dynamic model achieves 99.08% on the training set. In artificial data synthesis helped achieve real-time implementation of all 36 static-based gestures. Effective adaptability to ASL was also attained through transfer learning. The model trained on the dynamic dataset showed high variance, leaving scope for further research in this area to achieve real-time performance.

Dushyant Kumar Singh [5] proposed model is designed with a series of convolution 3D operations for convoluting multiple frame blocks in a single process, followed by Maxpooling 3D. The result obtained is then flattened and fed to a Multi-Layer Perceptron (MLP) having a softmax layer applied at the end to get the probability for each class as the result that shows the corresponding gesture representation. A deep neural network that would model and recognise the hand gestures of standard Indian Sign Language. In this authors are collected the dataset from different groups of age and complex backgrounds. The base 3DCNN architecture is used for analysing the modelling exercise for these dynamic gestures. Experimentation outcomes justify the model performance with a good accuracy of 88.24% of values achieved. Every person does not commonly known hand gestures in society. Moreover, each country's having its own set of symbols is a challenge. Standardization at a global level is not there, which leads to intricacy in communicating outside the country. Nevertheless, we have modelled the gestures with the expectation of having a large dictionary that would convert one country's hand gestures to another.

Pratik Likhar [6] proposed a method to implement real-time Indian Sign Language gesture recognition using two methods. First using a depth+RGB based Microsoft Kinect camera, and then using a normal RGB camera. For depth+RGB based techniques, the hand segmentation was done using depth perception techniques. For a normal RGB camera, a semantic segmentation approach was adopted. The usage of semantic segmentation completely removes the necessity of using a depth-based camera and segmentation. The U-Net model achieved an IOU score of 0.9920 and an F1 score of 0.9957 using just the RGB camera. For the depth+RGB trained models, techniques like having different lighting conditions and data augmentation helped achieve the generalisation in the case of static gestures. For dynamic gestures, the procurement of data was done at various fps values so that the model could learn the temporal features.

Sajanraj T D [13] proposed a method to implement Indian Sign Language recognition in the real time system that has developed for numeral signs from 0-9. The system has been trained using the 3000 static symbols of RGB images captured using the normal camera. The system has used 100 images for each symbol for testing. The model was created after the Deep Learning system was successfully implemented using a Region-based Convolutional Neural Network. The first layer is the input layer, which accepts the region of interest in the given video frame. In the convolutional layer, our entire image is considered as a multidimensional array and convolution operations are applied using a convolution matrix or kernel. The pooling layer is the layer that reduces the dimension of the image. Depending upon the pool size, in each image a single pixel is selected from the selected mask. The pooling has been

implemented by the max-pooling layer. In order to enable the positive values, we have used an activation function called RELU. After all the convolution and pooling operations, the system will flatten the entire processed images into a liner array, which becomes the node of the next layer. Each layer is connected to the next layer with corresponding weights, which is known as a "fully connected layer" or "dense layer." The output of the dense layer is called the scores, and these scores are given to the classification layer. Here, added the softmax layer as a classification layer. The system has attained an accuracy of 99.56% for the same subject while testing, and the accuracy was reduced to 97.26% in the low light condition.

Sruthi C. J [7] proposed a vision-based deep learning architecture for signer-independent Indian sign language recognition systems. A Convolutional Neural Network (CNN) is used to create a model named Signet, which can recognise signs based on supervised learning on data. The whole process can be divided into CNN training and model testing. The first and foremost step in all vision-based SLR systems is to pre-process the images in the data set to obtain a noise-free hand region extraction. In this work, dataset containing binary hand region silhouettes of the signer images. This image is then processed with a skin colour segmentation algorithm [2], followed by the largest connected component algorithm for hand region segmentation. These images are used in this work to train and test the signet architecture. Images along with their class labels are given to the developed CNN architecture to learn the classification model. The learned classification model can be tested and then saved for recognising ISL static alphabets. The system was successfully trained on all 24 ISL static alphabets with a training accuracy of 99.93% and with testing and validation accuracy of 98.64%. The recognition accuracy obtained is better than most of the current state of art methods.

Table -1: Approaches used in Literature

Reference	Methodology	Classes/Dataset	Accuracy	Gestures
[4]	CNN+LSTM	(36,1080)	99.08 %	Alphabet and numbers
[5]	CNN	(20,120)	88.2%	Own created dataset
[6]	CNN + LSTM	(36,45000)	78.3%	Alphabet and numbers
[13]	CNN	(10,3000)	97.2%	Only Numbers
[7]	CNN	(24,4125)	98.6%	Alphabet and numbers

4. Research Gap

Sign language is essential in the lives of anyone who is deaf or hard of hearing, but it does not have the same status as other languages. Hand gestures are the most widely used medium of a sign language-based communication framework among various forms of gestures. Recognizing gestures under dynamic parameters such as lighting conditions, multiple hands in the background, left-handed person, right-handed person, finger size, and so on is an intriguing field in automatic Indian sign language recognition. This review is based on datasets of hand gestures, alphabets, digits, words, and sentences with various real-time conditions used by different researchers at national and international levels. The detailed summarization of various methodologies to automate Indian sign language is tabularized in Table 1. A review is utilised to find out research gaps in the existing system and give inspiration to develop an interpreter for Indian sign language.

5. Challenges

Sign language recognition problems are a broad research area that includes problems like finger spelling dynamic alphabets, dynamic words, co-articulation detection and elimination for sentence identification. However, dynamic hand gesture systems in the real world struggle with various open challenges, including different lighting conditions, processing time, detection of hand segments, and many more.

6. CONCLUSIONS

A deep learning techniques based video classification system has been used for the recognition of dynamic hand gestures of Indian Sign Language. The system is based on an integrated architecture which combines layers from both CNN and LSTM. CNN layers have converted input data into features vector using feature extraction. Which tracks the sign demonstrator's hand motions using techniques including Object Stabilization, feature extraction, and finally Hand extraction. In ISL, the system can accurately and in real-time recognise hand poses and motions. The model's performance is justified by the results of the experiments, which shows good accuracy value.

REFERENCES

[1] P. K. Athira, "Indian sign language recognition," Phd thesis, Dept. CSE., NITC., Calicut, India, 2017, Accessed on: May, 2017.[online] Available: http://192.168.240.208:8080/xmlui/handle/123456789/35892

[2] Bhuyan, Manas Kamal, et al. "A novel set of features for continuous hand gesture recognition." Journal on Multimodal User Interfaces 8.4 (2014): 333-343

[3] Rekha, J., J. Bhattacharya, and S. Majumder. "Shape, texture and local movement hand gesture features for indian sign language recognition." Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on. IEEE, 2011.

[4] Neel Kamal Bhagat,Vishnusai Y,Rathna G N, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning", IEEE – 2019

[5] Dushyant Kumar Singh, "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modelling", 5th International Conference on AI in Computational Linguistics – 2021

[6] Pratik Likhar,Neel Kamal Bhagat, Dr. Rathna G N, "Deep Learning Methods for Indian Sign Language Recognition", ICCE-IEEE-2020

[7] Sruthi C. J ,Lijiya A, "A Deep Learning based Indian Sign Language Recognition System", IEEE-2019

[8] R. Jachak, "American Sign Language Recognition | Towards Data Science," 26 April 2020. [Online]. Available: https://towardsdatascience.com/american-signlanguagerecognition-using-cnn-36910b86d651

[9] https://www.kaggle.com/roobansappani/hand-gesture-recognition

[10] Kartik Shenoy, Tejas Dastane, Varun Rao, Devendra Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition", IEEE 9th ICCCNT -2018.

[11] Snehal Madhukar Daware, Manisha Ravikumar Kowdiki, "Morphological Based Dynamic Hand Gesture Recognition for Indian Sign Language "IEEE ICIRCA-2018

[12] Ms. Anagha Dhote Prof. S. C. Badwaik, "Hand Tracking and Gesture Recognition" IEEE International Conference on Pervasive Computing (ICPC)—2015

[13] Sajanraj T D , Beena M V, "Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network " IEEE 2nd ICICCT –2018

[14] Muthu Mariappan , Dr Gomathi V , "Real-Time Recognition of Indian Sign Language " IEEE ICCIDS ---2

[15] One-Shot Learning Hand Gesture Recognition Based on Lightweight 3D Convolutional Neural Networks for Portable Applications on Mobile Systems. (2019). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/8835909

[16] Hand Gesture Recognition for Sign Language Using 3DCNN. (2020). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/document/9078786

[17] On possibility of crystal extraction and collimation at 0.1-1 GeV. (1999). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/795508

[18] Human activity recognition with smartphone sensors using deep learning neural networks. (2016, October 15). ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0957417416302056

[19] Dynamic Hand Gesture Recognition Based on Signals From Specialized Data Glove and Deep Learning Algorithms. (2021). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9424596

[20] Hakim, N.L.; Shih, T.K.; Arachchi, S.P.K; Aditya, W.; Chen, Y.C; Lin, C.Y. Dynamic Hand Gesture Recognition Using 3DCNN and LSTM with FSM Context-Aware Model, Sensors, Year: 2019, Volume: 19, Issue: 24, 5429.

[21] Hand Gesture Recognition and Implementation for Disables using CNN'S. (2019, April 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8697980