ISSN 1975-4736 ©MITA2021

Real-Time Hand Tracking System for Indian Sign Language Gesture Recognition

Barathi Subramanian, Jeonghong Kim*
The School of Computer Science and Engineering
Kyungpook National University
Daegu South Korea.
Corresponding author e-mail: achu_samriti@yahoo.com
and jhk@knu.ac.kr

Bekhzod Olimov[2]
The School of Computer Science and Engineering
Kyungpook National University
Daegu South Korea.

***Abstract*— Real-Time Hand Tracking is much complicated and gives more variance in Human-Machine Interaction as it tracks the size of the hand, finger posture and other features. However, the existing hand tracking approaches for signing communication executed in various literature are more expensive. In this paper, we have implemented an automatic recognition of Real-time hand tracking system for the prediction of Indian Sign Language numbers in a cost-effective method. In this method, the input is taken from the webcam, and the name of the captured image is displayed on the screen with Frames per second count. It is executed by a platform for building hybrid Machine Learning resolutions called Mediapipe. Therefore, the proposed model recognized all the signs most accurately and achieved a high prediction accuracy of approximately 99%.**

Keywords*— Indian Sign Language (ISL), Mediapipe, Sign language recognition, Artificial Neural Network (ANN), Single Shot Detector (SSD).

I. INTRODUCTION

Sign language is a vision-based interactive language with complex linguistic rules. It is commonly used by hard of hearing community people as their mode of communication to exchange their thoughts, ideas and feelings by using different facial expressions, hand movements, body gestures and lip reading[1].Like spoken language, sign language has their unique linguistic structure, and it differs according to the geographic location[2].Although the sign language differs from one place to other, still it is viewed as a replacement of speech for hearing and visually impaired people everywhere around the world[3].Each country has their own sign language developed. Therefore, this makes it difficult for the normal hearing community people to understand gesture-based interactions[4]. According to the report of World Health Organization (WHO) around 500 million people around the world suffer from hearing loss[5]. Because of the high prevalence of deaf and dumb population, there is a need for eradicating communication challenge globally[6].

This paper in particular focuses on real time hand tracking system for Indian sign language gestures (ISL). This language is considered to be one of the most complex language when it is compared with other sign languages in the world[7]. In general, ISL includes single handed static numeral signing, word by word signing to form a complete sentence and letter by letter signing to express words for which no signs exist. The rest of the paper is organized in further 4 sections:2. Related Works, 3. Proposed Methodology,4. Experimental results and 5. Conclusion.

II. RELATED WORK

Previous researchers have emphasized their work on the prediction of sign language gestures in support of people having hearing impairments using the advancement of technologies with Artificial Intelligence algorithms. Though there has been a lot of research done for the recognition of sign language, still there exist limitations and improvements need to be done for the betterment of hard of hearing community[8]. Here we present a brief literature review on recent works in sign language recognition and address the problem in recognition of sign language gestures. In [7], authors have proposed a real time ISL hand gesture recognition system based on Microsoft Kinetic RGB-D camera as input images and have used deep learning techniques to achieve one-to-one mapping between the depth and the RGB pixels on training over 45,000 RGB images and depth images achieved a prediction accuracy of 98.81%. Despite good accuracy, the model emphasized the need of big dataset with more images to train and use of high pixel RGB camera.

In [9], authors have introduced an algorithm with Support Vector Machine (SVM) and using Microsoft Kinetic Xbox 360 RGB images, for the translation of Indian sign language gestures into English text and speech with the 100% prediction accuracy for the signs representing one numeric value and six ISL alphabets alone. Using expensive Leap motion controller [LMC][10], the authors have proposed a training method for American Sign Language (ASL) with LSTM recurrent neural network for handling sequence of input and yields an average accuracy rate of 91.08%. P.S Neethu et al.[11] introduces a deep convolutional neural network (CNN) classification approach with connected component analysis algorithm to segment the fingertips from the hand image to classify only eight different gestures into varieties of classes with 96.2% recognition rate. The performance is analyzed only in terms of sensitivity, accuracy and recognition rate.

III. PROSPOSED METHODOLOGY

The proposed architecture involves a series of processes to be performed with the input images in order to get the desired

gesture recognition output. There are 3 core processes implemented:

1. Data Wrangling
2. BlazePalm Detection
3. Hand Landmark Localization

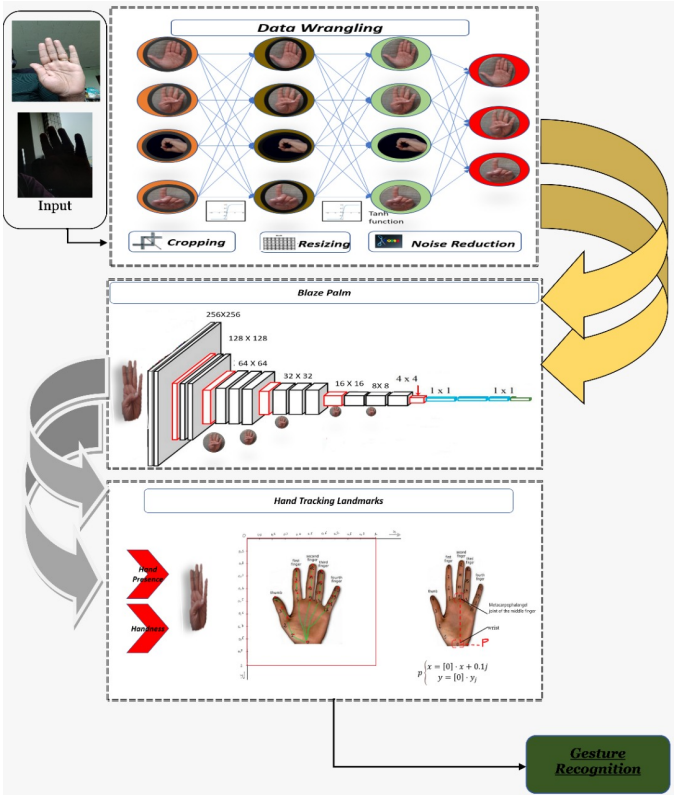


Figure 1.Proposed Architecture for Real-time Hand Tracking

A. *Data Wrangling:*

This process deals with the real-time images, captured with a web camera for Indian sign hand gestures for numerals. The input image data is prone to be noisier and may contain unrelated background data. Hence, we perform this data wrangling process through a sequence of sub processes such as cropping, resizing and noise reduction. Cropping is essential in order to bring the focus of hand gesture from the irrelevant background objects. Image resizing is not completely performed in this process as the Mediapipe platform perform this process once again in the blaze palm detection process for improving the computation time. We have trained the model with ANN for performing these processes in order to get clean and cropped image data. We have used hyperbolic tangent activation function in the ANN as it maps even the negative values as negative and enhances the classification accuracy of the images.

B. *BlazePalm Detection:*

This process is performed with the pre-processed image data from the data wrangling process. Using the single-shot detector named BlazePalm detector present in the Mediapipe, the initial hand locations are detected. However, detecting the hand locations is a tedious task as there are a variety of hand sizes possible in real-time environment. Once, the pre-processed image is completely resized at 256 x 256 pixels by the BlazePalm detector as per the processing requirement of the framework, we trained the palm detector to create the bounding boxes of rigid objects like palms and fists to focus on palm detection rather than hand detection so that the computing time complexity is reduced, and reliability of detection is improved. As palms are smaller objects in the image, non-maximum suppression algorithm employed in Mediapipe framework worked well in training the model.

C. *Hand Landmark Localization:*

After the palm detection process is completed, the process of hand landmark localization is executed. Below are the outputs:

1. Handedness classification as an output component helped to recognize whether left or right hand is posed in the image.

2. Hand flag output was used to detect the presence of hand in the image.

3. 21 landmark key points in 2-dimension space (x, y coordinates) which are labelled from 0-20 denoting the finger points in the hand which supports the gesture recognition.as shown in the proposed architecture.

4. Using these key points as x and y coordinates and applying the formula of Euclidean distance, the open or close state of each finger is identified which helps the model in detecting the numeral for the respective sign gesture displayed.

IV. EXPERIMENTS AND RESULTS

A. *Datasets:*

We created a real-time dataset of images having size 640 x 480 and 1920 x 1080 with a web camera and a smart phone camera respectively. To recognize every number, we captured 150 images in various angles, position and hand size under different lighting environments to train the model.

B. *Experimental Setup:*

We ran the simulation using Python 3.8.4 version on a desktop computer with RAM memory-32GB, Intel i7 processor at frequency 3.60Ghz on Windows 10 OS. The input image capturing was performed using a web camera with resolution of 720pixels/30fps and smart phone camera.

C. *Experimental Results:*

The model structure was well tuned by adjusting the hidden layers, neurons and weights in ANN so that the images were resized and cropped with lesser noise so that every action in Mediapipe such as palm detection, hand landmark localization was executed as expected and it resulted in very good

prediction accuracy of approximately 99%, We have hand tracking system tested and implemented for gesture recognition of ISL numbers. As a part of the future work, we are planning to implement for words recognition where the gestures of static sign images must be combined in a sequence by video as input. Figure 2. shows the Indian sign images versus the output gesture recognized image for the respective numerals.



Figure 2. (a) Input signs, (b) Output gesture recognized from the proposed methodology

Figure 3. gives a comparison of prediction for our proposed method with the state of art. On the x-label, we have the methods used in various research works and on the y- label, we have the prediction accuracy achieved.

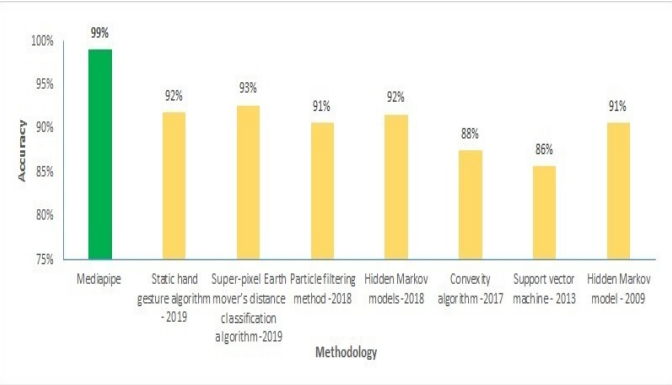


Figure 3.Plot of prediction accuracy achieved on comparison with state of art

V. CONCLUSION

In conclusion, we have implemented Mediapipe platform with ANN for designing a real-time hand tracking system to recognize the gestures for Indian Sign Language. One of the major features of the proposed architecture is cost efficiency along with good prediction accuracy. Our model was tested and proved to be suitable for application in real-time environment. Thereby, the model overcomes the challenge of communication gap between the normal hearing people and hearing impaired or hard hearing community. On comparing the prediction accuracy against the other existing methods used in literature works, we conclude that Mediapipe outperformed in this model. Therefore, our future research direction involves a designing and testing an efficient model for recognizing the words from the dynamic sign language gestures as video input with a focus of contributing to the communication challenge faced by all kind of people in the society.

Acknowledgment

This study was supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394).

References

[1] A. K. Sahoo, G. S. Mishra, and K. K. Ravulakollu, "Sign language recognition: State of the art," *ARPN J. Eng. Appl. Sci.*, vol. 9, no. 2, pp. 116–134, 2014.

[2] D. Li, C. R. Opazo, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 1448–1458, 2020, doi: 10.1109/WACV45572.2020.9093512.

[3] R. A. Kadhim and M. Khamees, "A real-time american sign language recognition system using convolutional neural network for real datasets," *TEM J.*, vol. 9, no. 3, pp. 937–943, 2020, doi: 10.18421/TEM93-14.

[4] R. Elakkiya and K. Selvamani, "Subunit sign modeling framework for continuous sign language recognition," *Comput. Electr. Eng.*, vol. 74, pp. 379–390, 2019, doi: 10.1016/j.compeleceng.2019.02.012.

[5] N. B. Ibrahim, H. H. Zayed, and M. M. Selim, "Advances, Challenges and Opportunities in Continuous Sign Language Recognition," *J. Eng. Appl. Sci.*, vol. 15, no. 5, pp. 1205–1227, 2019, doi: 10.36478/jeasci.2020.1205.1227.

[6] O. Koller, "Quantitative survey of the state of the art in sign language recognition," *arXiv*, 2020.

[7] N. K. Bhagat, Y. Vishnusai, and G. N. Rathna, "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning," *2019 Digit. Image Comput. Tech. Appl. DICTA 2019*, 2019, doi: 10.1109/DICTA47822.2019.8945850.

[8] V. Adithya, P. R. Vinod, and U. Gopalakrishnan, "Artificial neural network based method for Indian sign language recognition," *2013*

IEEE Conf. Inf. Commun. Technol. ICT 2013, no. Ict, pp. 1080–1085, 2013, doi: 10.1109/CICT.2013.6558259.

[9] T. Raghuveera, R. Deepthi, R. Mangalashri, and R. Akshaya, "A depth-based Indian Sign Language recognition using Microsoft Kinect," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 45, no. 1, pp. 1–13, 2020, doi: 10.1007/s12046-019-1250-6.

[10] D. Naglot, "Recognition using the Leap Motion Controller," *2016 Int. Conf. Inven. Comput. Technol.*, vol. 2, pp. 1–6.

[11] P. S. Neethu, R. Suguna, and D. Sathish, "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," *Soft Comput.*, vol. 24, no. 20, pp. 15239–15248, 2020, doi: 10.1007/s00500-020-04860-5.