Hand Landmark Based Approach to Enhance Sinhala Sign Language Gesture Recognition

H.G.C. Wijayagrahi[a, *], S.S. Vitharana[b], S.C. Gama Achchige[c]

[a] *Business Analytics Centre, National Institute of Business Management, Colombo 00600, Sri Lanka*
[b] *Department of Mechanical and Electrical Engineering, National Institute of Technology, Tokuyama College, Shunan 745-8585, Japan*
[c] *Faculty of Physics and Earth Sciences, University of Leipzig, Germany*
**Email correspondence: wijayagrahi1997@gmail.com*

Abstract

Sign language serves as a significant mode of communication for individuals with hearing impairments, facilitating their interaction and integration within society. In the context of Sri Lanka, where Sri Lankan Sign Language is prevalent, effective communication between deaf individuals and the hearing population becomes predominant. This research paper proposes a novel approach to Sri Lankan Sign Language recognition by leveraging hand landmarks in combination with a Feedforward Neural Network. Most of the approaches for gesture recognition are based on object detection which leads to requirement of large training datasets and classic image processing techniques to classify the image data is often unreliable with gesture recognition. In our approach we minimize the data set by extracting 21 hand landmarks coordinates by using a robust pretrained 'Media-Pipe Hands Model', the model has proven performance and is reliable in accurately tracking the multiple hands and making landmark predictions. Thorough experimentation and evaluation are conducted to establish the effectiveness of the proposed approach performance is compared to other state-of-the-art sign language recognition techniques, demonstrating competitive outcomes. The research analyzes and discusses various factors like network architecture, training data size, and hand landmark representation, ensuring the approach's validity and efficiency. The study's outcomes are applicable to real-world scenarios, including communication aids, educational tools, and human-computer interaction systems for the deaf community in Sri Lanka. By highlighting the integration of hand landmark data and advanced machine learning techniques, the research extends its relevance beyond Sri Lankan Sign Language to the broader domain of sign language recognition. Ultimately, this research contributes to inclusivity, accessibility, and effective communication for the hearing-impaired population.

Keywords: **Sinhala Sign Language, Gesture Recognition, Machine Learning, Prediction**

Introduction

Sign languages are complex visual-gestural systems used by deaf communities around the world as their primary means of communication [1]. There are around 300 million people [2] and 1 to 3 per 1000 children [3] suffering permanent hearing loss affects in the world. However, there are challenges in bridging the communication gap between sign language users and those who do not understand it. In educational settings, sign language detection can aid teachers in understanding needs of students better, thereby improving inclusivity. Moreover, it can enhance accessibility in healthcare by facilitating smooth communication between healthcare professionals and deaf people [4]. This is where sign language detection techniques play a crucial role.

Sri Lankan Sign Language is unique and has its own grammar, syntax, and vocabulary, reflecting the rich linguistic diversity. Sri Lankan Sign Language shares similarities with British Sign Language due to historical influences during British colonial rule in Sri Lanka [5]. However, it also has distinct features that distinguish it as a separate sign language.

The performance of existing sign language recognition approaches is limited by the accuracy and scale of training data. The sign language recognition systems could become more robust and accurate with better data and proper machine learning techniques. [6]. This research aims to improve the reliability of Sinhala Sign Language gesture recognition by using a state of the art 'Media-Pipe Model' to obtain the hand landmarks which is used to train a feed forward neural network using TensorFlow to make the predictions, the model could be expanded and deployed in multiple platforms as required.

Methodology

Our methodology begins with the capture of hand images through the OpenCV library, followed by the utilization of Google Media-Pipe to detect 21 key landmarks on the hand [7]. The data then undergoes a series of preprocessing steps that build the following system architecture.

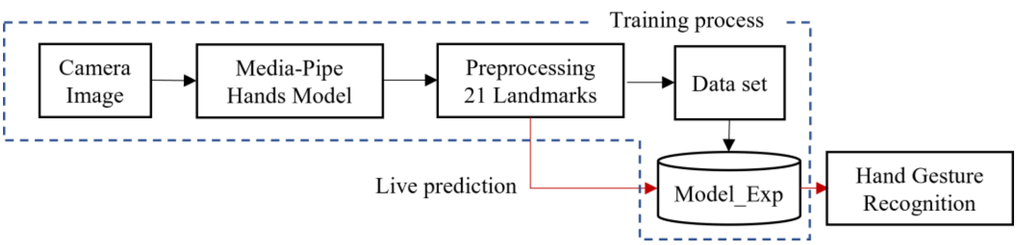

Figure : 1 System Architecture.

Preprocessing of landmarks starts with the conversion of landmark coordinates into relative coordinates with respect to a reference point which is the wrist of the user. These relative coordinates are subsequently flattened into a one-dimensional array and normalized to the maximum value.

Data Set Preparation

In the data collection phase of our study, we focused on acquiring a comprehensive dataset for sign language recognition. Specifically, we concentrated on six static sign characters (අ, ආ, ඇ, ඉ, උ, ඵ) each of which were denoted to an English letter as follows.

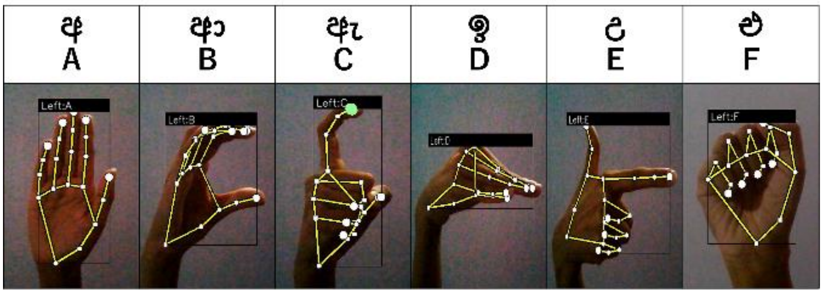

Figure : 2. Static Gesture Set.

These characters were carefully selected to represent a range of hand configurations and gestures commonly used in sign language communication. During data collection, we employed the Google Media-Pipe framework to detect and track these static sign characters, capturing their x and y coordinates, or landmarks, at a rate ranging from 15 to 25 frames per second (fps) which is considerable responsive for majority of the applications, where a single frame is captured for each character gesture, providing sufficient detail for visualization. For each of the six sign characters, we used a substantial dataset consisting of 3500 arrays. This preprocessed data forms the basis for training a feed-forward neural network, which we refer to as 'Model_Exp,' using our dataset for about 1400 steps. The architecture of the feed forward neural network is as follows.

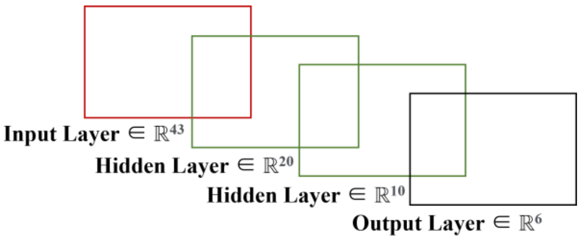

Figure : 3. Neural Network Architecture.

The input layer has 43 neurons in total with the first neuron accepting the ID of the gesture and the rest of the 42 neurons accepts the normalized x and y coordinated preprocessed from each of the 21 landmarks results from the ‘Media-Pipe Hands Model’.

In the prediction phase, real-time hand images from a webcam are processed to detect the same set of landmarks, and the data undergoes identical preprocessing steps. The resulting preprocessed data is then input into 'Model_Exp' for the prediction of sign language gestures. Finally, these predictions are visually represented using OpenCV, facilitating a clear and intuitive means of conveying sign language gestures to non-signers.

Results and Discussion

The evaluation of the model was logged as follows in Tabel 1 for each tested gesture.

Table 1. Model Evaluation.

	Precision	Recall	F1-Score	Support
අ (A)	0.68	0.99	0.81	1577
ආ (B)	0.99	0.54	0.70	1730
ඇ (C)	0.71	0.99	0.82	1935
ඉ (D)	0.94	0.97	0.95	914
උ (E)	0.97	0.96	0.97	900
එ (F)	1.00	0.15	0.27	869
Accuracy			0.79	7925
Macro avg	0.88	0.77	0.75	7925
Weighted avg	0.85	0.79	0.76	7925

According to the Tabel 1 the accuracy of the model was evaluated as 79.0% which is proves the reliability of this implementation. A similar study [8] conducted on Sinhala Sign Language gesture recognition based on neural networks is extracted to the following table.

Table 2. Reference Gesture Evaluation Data.

Letter	Accuracy
අ	100%
ව	100%
ඉ	80%
එ	80%
ඊ	70%
ආ	70%
උ	70%
ය	60%

The overall average accuracy when calculated from the Tabel 2 data yields 78.75 %, in comparison the approach presented by our study is similarly reliable for Sinhala Sign Language gesture recognition with plenty of room for performance tuning by adjusting the data set with diverse images and improving the architecture of the neural network.

Conclusion

This research introduces an advanced hand gesture recognition approach that combines machine learning techniques and hand landmark data. These techniques have potential for diverse applications, offering advantages over traditional image processing by using a concise landmark coordinate dataset.

The potential for transfer learning expands gesture recognition capabilities. Factors such as network architecture, data size, and landmark representation were established to enhance the prediction performance.

Our research's practical implications are wide-reaching, benefiting communication aids, educational tools, and human-computer interaction for Sri Lanka's deaf community. Additionally, the integration of hand landmark data and advanced machine learning widens its relevance beyond sign language recognition. This fosters inclusivity, accessibility, and effective communication for the hearing-impaired.

This proposed gesture recognition system holds potential for broader technological applications. The Media-Pipe framework's robust and high accuracy opens avenues for collaborative development and multi-body part gesture recognition. Future work aims to advance both static and dynamic recognition, contributing to human-machine interactions and technology evolution.

Conflict of Interest

None.

References

[1]. Sanders, Nathan, and Donna Jo Napoli. "Reactive effort as a factor that shapes sign language lexicons." _Language_ (2016): 275-297.

[2]. Kumar, D. Manoj, et al. "EasyTalk: A translator for Sri Lankan sign language using machine learning and artificial intelligence." _2020 2nd International Conference on Advancements in Computing (ICAC)_. Vol. 1. IEEE, 2020.

[3]. Fitzpatrick, Elizabeth M., et al. "Sign language and spoken language for children with hearing loss: A systematic review." _Pediatrics_ 137.1 (2016).

[4]. Marzouk, Radwa, et al. "Atom Search Optimization with Deep Learning Enabled Arabic Sign Language Recognition for Speaking and Hearing Disability Persons." _Healthcare_. Vol. 10. No. 9. MDPI, 2022.

[5]. Atkinson, Joanna, et al. "Detecting cognitive impairment and dementia in deaf people: the British Sign Language cognitive screening test." _Archives of Clinical Neuropsychology_ 30.7 (2015): 694-711.

[6]. Dhulipala, Shivanarayna, Festus Fatai Adedoyin, and Alessandro Bruno. "Sign and human action detection using deep learning." _Journal of Imaging_ 8.7 (2022): 192.

[7] Takahashi, Kazuhito. "Kazuhito00/Hand-Gesture-Recognition-Using-Mediapipe: MediaPipe (Python) GitHub, github.com/Kazuhito00/hand-gesture-recognition-using-mediapipe.

[8]. Fernando, K. S. S., and S. D. H. S. Wickramarathne. "Sri Lankan Sign Language Tutor." (2013).