



Indian Sign Language Recognition Using Deep Learning Techniques

D. Karthika Renuka ^{a,*}, L. Ashok Kumar ^b

^a Department of Information Technology, PSG College of Technology, Coimbatore-641004, Tamil Nadu, India.

^b Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore-641004, Tamil Nadu, India.

* Corresponding Author: dkr.it@psgtech.ac.in

Received: 12-03-2022, Revised: 28-04-2022, Accepted: 07-05-2022, Published: 12-05-2022

Abstract: By automatically translating Indian sign language into English speech, a portable multimedia Indian sign language translation program can help the deaf and/or speaker connect with hearing people. It could act as a translator for those that do not understand sign language, eliminating the need for a mediator and allowing communication to take place in the speaker's native language. As a result, Deaf-Dumb people are denied regular educational opportunities. Uneducated Deaf-Dumb people have a difficult time communicating with members of their culture. We provide an incorporated Android application to help ignorant Deaf-Dumb people fit into society and connect with others. The newly launched program includes a straight forward keyboard translator that really can convert any term from Indian sign language to English. The proposed system is an interactive application program for mobile phones created with application software. The mobile phone is used to photograph Indian sign language gestures, while the operating system performs vision processing tasks and the constructed audio device output signals speech, limiting the need for extra devices and costs. The perceived latency between both the hand signals as well as the translation is reduced by parallel processing. This allows for a very quick translation of finger and hand motions. This is capable of recognizing one-handed sign representations of the numbers 0 through 9. The findings show that the results are highly reproducible, consistent, and accurate.

Keywords: Speech Impairment, Voice-to-Sign Language, Multi-stream CNN-LSTM, HMM

Introduction

A sign is a hand motion or both hand movements combined with a facial expression that communicates a specific message. Because some people would find it difficult to learn and understand sign language, this project provides a solution by developing a voice-to-sign language

translation system that combines speech and image processing methods. Speech processing, which includes voice recognition, is the study of recognizing spoken words, regardless of who is speaking. The primary strategy of the proposed model is template-based recognition [1]. Convert Language to Sign Using Procedure Sign language is the natural means of communication for those that are deaf or even have speech impairments (SL). Although communication among deaf, hard of hearing, and hearing signers is not difficult, the deaf population faces significant challenges in the educational, social, and workplace sectors. The primary goal of this study is to develop new vision-based technology for continuous sign language recognition and text translation [2]. To help the deaf and/or speech-impaired connect with hearing people, this paper proposes a mobile active application program for machine translations of Indian sign language into English speech. This sign language interpreter should be able to translate values 0 through 9. Although facial expressions contribute significant information to the emotional component of the sign, their analysis wasn't included in this project effort because it would complicate an already difficult problem. Because our technology is designed to listen to the deaf, it could serve as a translator between the deaf and those who do not understand sign language, eliminating the need for a middleman. People who are deaf or hard of hearing would converse using their natural speech patterns. Someone who is both deaf and stupid is said to be unable to hear or speak. Deaf and dumb people are indeed a closed-off, introverted culture whose numbers are steadily increasing. Deaf education has only been around for about a century. When there is no proper language, the deaf-dumb prefer sign language for education because it is the first form of communication in Hinduism. Because ASR devices can only detect a small number of discrete words, the user must stop after each word as well as repeat it numerous times in required to practice the system. Expert understanding of speech variations is hand-coded into a system. In which statistical learning techniques, typically Markov Models, are used to automatically model changes in speech. To overcome the disadvantages of HMM, machine learning techniques such as neural nets and genetic algorithms/programming could be introduced.

Related Works

Cascaded Segment on Sensing Networks for Text-based Traffic Sign Detection simplifies Text Box architecture by creating traffic sign-specific designs. On the one hand, the traffic sign detector is language-insensitive because it does not detect based on text. The text detector, on the other hand, can be fine-tuned with any other language. It was created and modified as an LSTM model for continuous sign language recognition using leap motion. CNN employs a variant of multilayer perceptrons that is designed to require minimal preprocessing [3]. The input gate regulates the flow of activations into memory cells. The output gate regulates the flow of cell activation output to the rest of the network. Finding Sequential Parallelism Using Multi-stream CNN-LSTM-HMM and Weakly Supervised Learning The mixed HMM modeling research for sign language recognition has been extended to sign language. An Effective Wireless Mega Capacitive Sensor System Action Recognition with Intelligence at the Edge Using Gloves ML categorization models are built on a Raspberry Pi using the machine learning packages like

scikit-learn to achieve AI on the edge [4]. Over 100 iterations, the modified and decoded datasets are used to train and test machine learning models, retrieving model accuracy, training time, and training time for comparisons. In each iteration, 3000 samples were retrieved from each gesture and shuffled into 2,400 training samples (80%) and 600 testing samples (20 percent).

Focused on a Chinese Sign Language Recognition Method Key frame-oriented Clips (kccs) with adaptive lengths in SLR beyond frame level were proposed to promote efficiencies and minimize analysis problem [2]. Key frames with depth and trajectory information are utilized to extract multimodal kccs. KCC sign trajectory data is used in all approaches except the CNN approach, whereas KCC sign depth data is utilized only in the 3D-CNN strategy. We can learn that the KCC strategy can accomplish a substantial optimum efficiency by lowering redundant data by 3-7 times. Hand tapping gestures-Based non-touch character input system the kinect identifies a user at a range of 0.8m 4.0m. The skeleton data includes locations for 20 various joint positions on the user's body. The P-hand hand area is a small region of the user image containing all of P-fingertips. Hand's The system utilizes opencv's Find contours function, that utilises the Suzuki85 algorithm. We find the distance among each point on the outline and the centre of the P-hand and compare it to the ranges between the nearest successive 40 points on the outline. The feature vector obtained after applying the feature fusion method is given to the classification model in DCA-based Unimodal Feature-level Fusion of Orthogonal Moments for Indian Sign Language Dataset. The accuracy, feature vector size, and computational time of the results are all assessed. The accuracy is defined as the percentage ratio of correctly identified images to the total set of input images in a single category during classification test results, and the computing time is the time needed to develop and test the SVM-based classifier model [5]. A sensor for identifying hand gestures using reflected impulses. Gesture evidenced waveforms were assessed at a length of 15 centimetre from the transmitter. The distance between the arm and the transmitter exposes a lag time and linearly weights the amplitude of the waveform. We evaluated different gestures in American Sign Language ("S", "E", "V", "W", "B", and "C") from 5 persons. The mirrored output waveform of the 3 gestures ("S", "V", and "B").

Proposed Methodology

Sign language is a type of structured gesture that has long been utilized by the differently abled. The study's goal is to create a mobile application that will assist parents, teachers, and differently abled students in learning Indian Sign Language. Fix the limitations of currently available mobile sign language learning applications. One critical aspect of any methods is understanding the significance of this work in society [6]. The current application was developed using the Android operating system, which is designed for smart phones and tablet computers. This operating system must be capable of providing access to various helpful toolkits which can be utilized to develop other downstream applications. Conversation allows individuals to interact and share emotions and attitudes [7]. The differently abled community faces a number of challenges when interacting with the general society. There is a method that could transform sign

languages to intelligible way that provide communication with people. These techniques only allow differently abled people to communicate utilising sign language via smart phones [8]. Smartphone based differently abled interaction model that enables the needs of differently abled developing speech enabled application in mobile that will transform their sign language into texts [9]. Ignorant differently abled people can only send and receive messages via sign language, which causes significant problems in their day to day life.

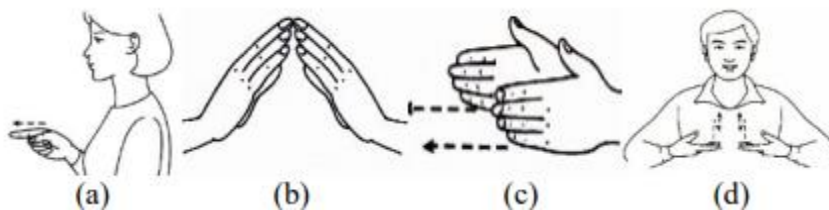


Figure 1. Sample Hand Gestures (a) you (b) Home (c) Road (d) Happy

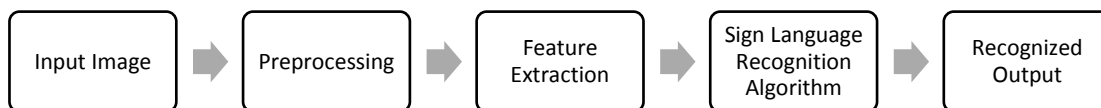


Figure 2. Work flow of Sign Language Detection

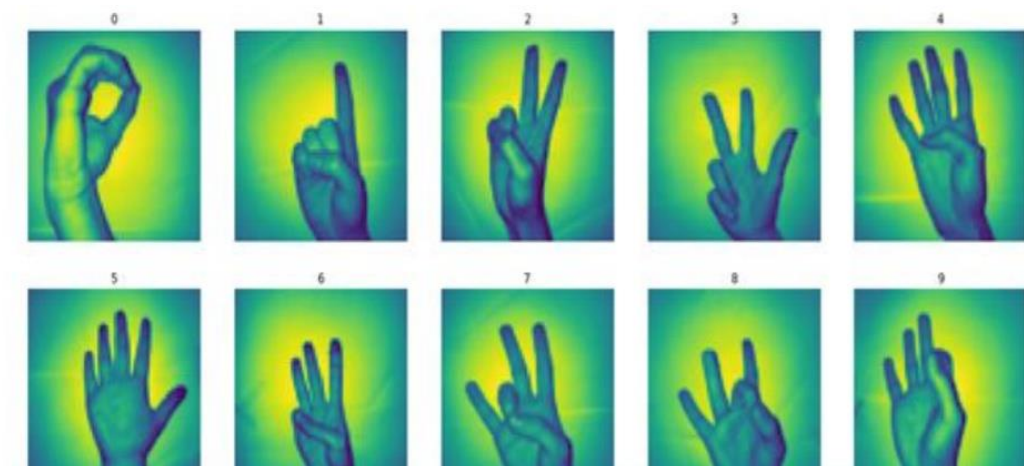


Figure 3. Classification Using CNN

Result Analysis

Convolution Neural Network based classification performance is analysed using confusion matrix. Accuracy is the performance metric used to evaluate the performance of system

$$Accuracy = (TP + FP)/(TP + TN + FP + FN)$$

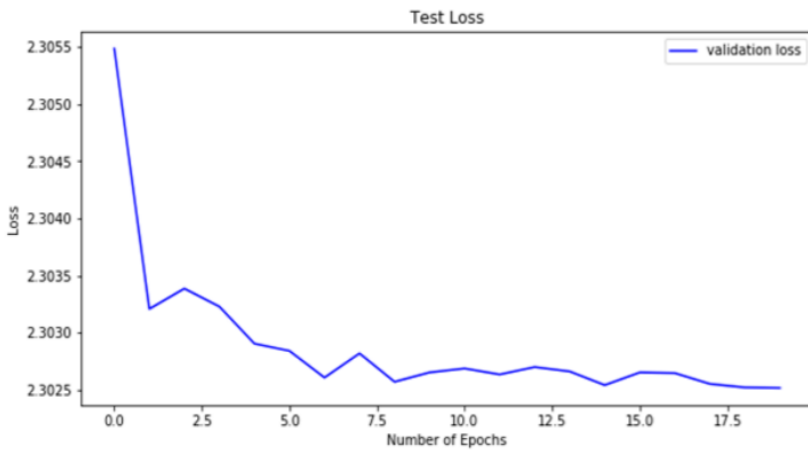


Figure 4. Value of Loss

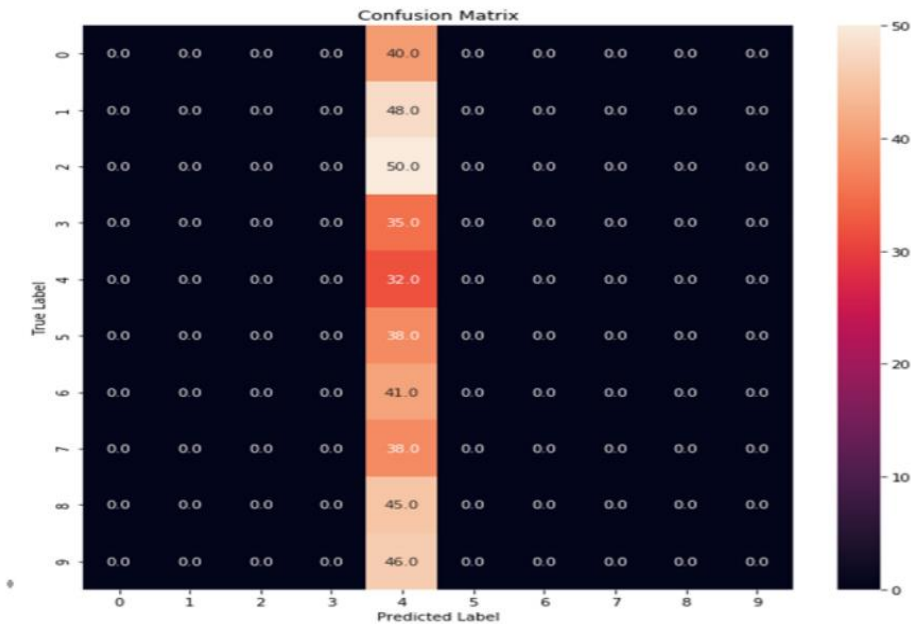


Figure 5. Confusion Matrix

Conclusion

A practical application for poorly educated hearing-impaired people is suggested. This development intends to help the hearing-impaired people by offering them with an interesting communication and collaboration device. This work proposed a smart phone-based application that lets hearing impaired people in our society to interact. It also produces assistive technology for the hearing-impaired people in a variety of fields such as restaurants, hospitals, and transportation. Furthermore, this implementation includes a translator which transform Indian to English Sign Language.

References

- [1] Chakraborty, B.K., Sarma, D., Bhuyan, M.K., & MacDorman, K.F., (2017). Review of constraints on vision-based gesture recognition for human-computer interaction. *IET Computer Vision*, 12(1), 3-15. <https://doi.org/10.1049/iet-cvi.2017.0052>
- [2] Shin, J., & Kim, C.M., (2017). Non-touch character input system based on hand tapping gestures using Kinect sensor, *IEEE Access*, 5, 10496-10505. <https://doi.org/10.1109/ACCESS.2017.2703783>
- [3] Yu, Y., Chen, X., Cao, S., Zhang, X., & Chen, X., (2019). Exploration of Chinese sign language recognition using wearable sensors based on deep belief net, *IEEE journal of biomedical and health informatics*, 24(5), 1310-1320. <https://doi.org/10.1109/JBHI.2019.2941535>
- [4] Yang, X., Chen, X., Cao, X., Wei, S., & Zhang, X., (2016). Chinese sign language recognition based on an optimized tree-structure framework, *IEEE journal of biomedical and health informatics*, 21(4), 994-1004. <https://doi.org/10.1109/JBHI.2016.2560907>
- [5] Pan, J., Luo, Y., Li, Y., Tham, C.K., Heng, C.H., & Thean, A.V.Y., (2020). A wireless multi-channel capacitive sensor system for efficient glove-based gesture recognition with AI at the edge, *IEEE Transactions on Circuits and Systems II: Express Briefs*, 67(9), 1624-1628. <https://doi.org/10.1109/TCSII.2020.3010318>
- [6] Oliveira, T., Escudeiro, N., Escudeiro, P., Rocha, E., & Barbosa, F.M., (2019). The virtualsign channel for the communication between deaf and hearing users, *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 14(4), 188-195. <https://doi.org/10.1109/RITA.2019.2952270>
- [7] Koller, O., Camgoz, N.C., Ney, H., & Bowden, R., (2019). Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos, *IEEE transactions on pattern analysis and machine intelligence*, 42(9), 2306-2320. <https://doi.org/10.1109/TPAMI.2019.2911077>

- [8] Kim, S.Y., Han, H.G., Kim, J.W., Lee, S., & Kim, T.W., (2017). A hand gesture recognition sensor using reflected impulses, *IEEE Sensors Journal*, 17(10), 2975-2976. <https://doi.org/10.1109/JSEN.2017.2679220>
- [9] Joshi, G., Vig, R., & Singh, S., (2018). DCA-based unimodal feature-level fusion of orthogonal moments for Indian sign language dataset, *IET Computer Vision*, 12(5), 570-577. <https://doi.org/10.1049/iet-cvi.2017.0394>

Funding

No funding was received for conducting this study.

Conflict of interest

The Authors have no conflicts of interest to declare that they are relevant to the content of this article.

About The License

© The Author(s) 2022. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License