Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Real-time Conversion of Sign Language to Text and Speech

Raksha Beriya , Prof. K. K. Chhajed
Department of Computer Science & Engineering, P. R. Pote (Patil) Education & Welfare Trust's Group of Institutions, College of engineering & Management, Amravati

Abstract:

While the rest of the population communicates vocally, the speech-impaired community uses sign language exclusively. To read the static and dynamic signs in Indian Sign Language and translate them into speech, this effort intends to close the communication gap. Data regarding the actions is gathered using a sensor glove equipped with flex sensors to measure how each finger bends and an IMU to determine the hand's orientation. After being wirelessly sent, this data is then categorised into appropriate speech outputs. Since LSTM networks can learn long-term dependencies, they were investigated and used for the classification of gesture data. The created model will have classification accuracy for 26 motions, demonstrating the viability of employing neural networks for sign language translation.Recognition of hand gestures is a key to getting past many obstacles and making life easier for people. Numerous applications might be made use of when machines are able to comprehend human activity and its meaning. The identification of sign language is one area of particular attention. This study presents a comprehensive analysis of current advances in hand gesture and sign language recognition methods. The strategies examined are appropriately divided into five stages: data collection, pre-processing, segmentation, feature extraction, and classification. At each level, alternative algorithms are explained, and their advantages are contrasted. We also talk about the difficulties and restrictions that general gesture recognition research, as well as sign language recognition specifically, faces. It is hoped that the study will serve as a thorough introduction to the field of automated gesture and sign language identification for readers and will help future research efforts in this area.

Keywords: sign language, neural network, image processing, deep learning

I. INTRODUCTION

Around 7 million Indians are deaf and mute, while 1.9 million have speech impairment, per the 2011 census. 30% of them have never attended school, and 63% of them are unemployed [1]. Due to their impairment, they can only express themselves to one another through gestures and signals. Most people use verbal communication, while those with speech impairments rely on sign language. There is a communication gap as a result. Because of this, individuals are at a disadvantage and are unable to access the same educational and employment prospects.The speech-impaired people in India uses a standard set of sign language gestures known as the Indian Sign Language (ISL). These ISL motions can be divided into static and dynamic types. When making a sign, a static gesture is one in which the hands remain still. The majority of the ISL motions for alphabets is static indicators. On the other hand, dynamic gestures, which make up the majority of ISL gestures, involve hand motions while being performed.

The communication gap could be effectively closed by a technology that can translate Indian Sign Language into voice, enabling the speech-impaired to interact with the public. This would facilitate their social integration and lifestyle change.The gadget would need to be able to recognise both static and dynamic motions in ISL to be able to translate it. Standard neural networks can categorise static motions, but they cannot categorise dynamic gestures. Dynamic gestures produce sequential data since the reading at each time point depends on the readings taken at earlier points in time. Standard neural networks cannot be used for the classification of sequential data because they demand that each reading be independent from the other readings.

Recurrent neural networks (RNNs) are connectionist models that can process sequential data one element at a time while selectively passing information across sequence stages [4]. As a result, they can model input and/or output made up of sequences of non-independent pieces.Recurrent neural networks can simulate sequence and temporal dependencies simultaneously on various scales [4]. However, it has been noted that when the time range of dependencies grows, the gradient descent of error criterion may become increasingly ineffective [5]. Long Short-Term Memory (LSTM), a unique kind of RNN capable of learning long-term dependencies, can be used to overcome this inefficiency. With noisy, incompressible input sequences, LSTM is a recurrent network design that can train to bridge time gaps more than 1000 steps without sacrificing short time lag performance [6]. In order to create a device that can translate sign language in real-time, LSTM-based neural networks can be utilised to classify sensor data corresponding to sign language movements to their appropriate speech labels because of their capacity to learn long-term relationships. In this research, we describe the construction and operation of a wearable glove that can convert ISL signs into speech. The embedded sensors in the glove use Bluetooth to send data to a processing device, such as a PC or smartphone. An LSTM-based neural network is then used to classify this data to produce output for both text and speech.

1.1 CHALLENGES IN GESTURE RECOGNITION

Complex procedures including motion modelling, motion analysis, pattern recognition, and machine learning are used in the recognition of gestures. There are both manual and automatic ways in it. Predictive ability is impacted by environmental factors including background lighting and movement speed. The gesture appears differently in 2D space due to the multiple vantage angles. In some studies, the signer is required to wear a wristband or coloured glove to help with hand segmentation, as seen in [3, 8, 10]. The segmentation process is made simpler by using coloured gloves. Dynamic gesture recognition is expected to have issues with temporal variance, spatial complexity, movement epenthesis, repeatability, connection, and numerous properties including change in orientation and region of gesture. The effectiveness of a gesture recognition system in overcoming the difficulties can be measured using several evaluation criteria. Scalability, robustness, real-time performance, and user independence are some of these requirements.

II. RELATED WORKS

Various studies on various sign language recognition techniques have been conducted in the past.

The usage of electromechanical devices comes first. The natural signing skill of the signer is impacted by this type of system. The usage of coloured gloves falls under the second category and employing any technology that can interfere with a signer's natural signing ability falls under the second category as well [3].

A brand-new approach for developing SLR systems based on EMG sensors and a data glove was presented by Al-Ahdal and Tahir [1]. This technique uses electromyography signals from the hand muscles to determine the word boundaries for continuous SLR streams of words.

The colour gloves strategy was put forth by Iwan Njoto Sandjaja and Nelson Marcos [2] and uses a multi-color tracking algorithm to extract significant data from the movie.

Ibraheem and Khan [4] have studied several gesture recognition methods, including more modern ones.

For hand tracking and segmentation, Ghotkar et al. [5] employed the Cam shift approach and the Hue, Saturation, Intensity (HSV) colour model. A genetic algorithm is employed to recognise gestures.

A straightforward approach for recognising sign language was created by Paulraj M P et al. [7] utilising skin colour segmentation and artificial neural networks.

Gaikwad, P.B. et al. [8] suggest a hardware approach combining accelerometers and flex sensors inside gloves to recognise the indications. According to [8], one sign can be distinguished from others by the degree of bends and the speed of the hand movements. Although statistically accurate, such a device has built-in limitations due to its cost, upkeep, and use.

In one of the early pieces of work, Starner, T.E. et al. [9] suggest a visual solution. The system can recognise a portion of the vocabulary used in American Sign Language utilising a single camera module and a Hidden Markov Model.

The method needs the signer to wear particular-coloured gloves so that the camera can follow their hands because it does not operate on bare hands. Its limitation of requiring gloves, however, prevents regular use.

Using active contour models [10] suggest a solution to the difficult challenge of tracking hands without gloves. This energy-saving method is used to follow a signer's hands and head in real-time video. Error back propagation algorithm is used to classify signs using a straightforward feed forward neural network. The research does not seek to identify continuous indicators in a streaming video; instead, it primarily concentrates on identifying single signs[11-15].

III. PROPOSED METHODOLOGY

The suggested system's block diagram is displayed in Fig. I. Here, a consistent white background and a web camera are used to input hand movements into our system. The right hand is used to create 26 combinations of Indian signs that are then recorded in the training database.

These recorded input movements are pre-processed. After that, hands are segmented to isolate the object from the background. Certain elements are used to represent the segmented hand image. These features are utilised in conjunction with the PCA method, which produces the best results, to recognise gestures. The undesirable outcome was transformed into the appropriate text and speech form.

The process of recognising signs consists of four main parts.

Data collection, pre-processing, segmentation, feature extraction, sign recognition, and sign to text and speech conversion are among them.

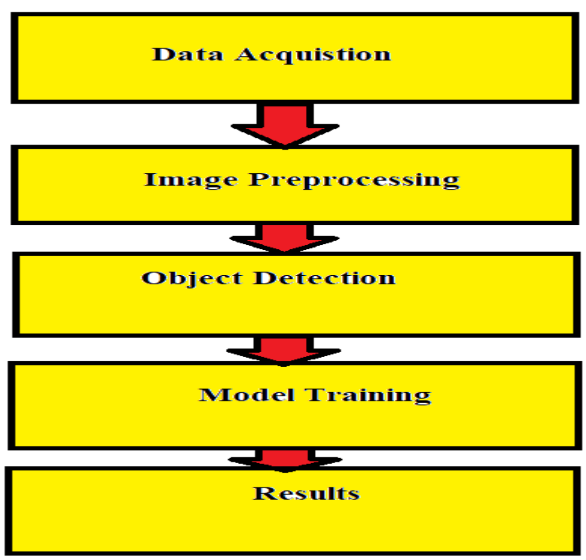

Fig 1: Flow Diagram

CONCLUSION

The suggested approach provides output in text and audio formats, assisting in decreasing the communication gap between deaf-dumb and sighted persons. This effort will eventually be expanded to include all the Marathi and Hindi signs' phonemes.

REFERENCES

[1] M. Ebrahim AI-Ahdal&NooritawatiMdTahir," Review in SignLanguage Recognition Systems" Symposium on Computer &InfOlmatics(ISCI),pp:52-57, IEEE ,2012

[2] IwanNjotoSandjaja, Nelson Marcos," Sign Language NumberRecognition" Fifth Intemational Joint Conference on INC, IMSand IDC, IEEE 2009

[3] Pravin R Futane , Rajiv v Dharaskar," Hasta Mudra aninterpretatoin of Indian sign hand gestures", internationalconference on digital object identifier, vol.2, pp:377-380, IEEE,2011.

[4] Noor Adnan Ibraheem and RafiqulZaman Khan," Survey onVarious Gesture Recognition Technologies and Techniques",IntelllationalJoulllal of Computer Applications (0975 - 8887)Volume 50 - No.7, July 2012

[5] Archana S. Ghotkar, RuchaKhatal,SanjanaKhupase, SurbhiAsati&MithilaHadap," Hand Gesture Recognition for IndianSign Language" Intelllational Conference on Computer Communication and Informatics (lCCCI), pp: 1-4. IEEE,Jan 2012.

[6] Paulraj M P, SazaliYaacob, MohdShuhanaz bin ZanarAzalan,RajkumarPalaniappan," A Phoneme Based Sign LanguageRecognition System Using Skin Color Segmentation" 6thIntelllationalColloquiwn on Signal Processing & Its Applications(CSPA), pp:I-5,IEEE,20IO.

[7] Nasser H. Dardas and Emil M. Petriu" Hand Gesture Detectionand Recognition Using Principal Component Analysis"intelllational conference on computational intelligence for measurement system and application (CIMSA), pp: 1-6, IEEE,2011.

[8] Gaikwad, P.B. and Bairagi, D.V., 2014. Hand gesture recognition fordumb people using indian sign language. International Journal ofAdvanced Research in computer Science and Software Engineering,pp.193-194.

[9] Starner, T.E., 1995. Visual Recognition of American Sign LanguageUsing Hidden Markov Models. Massachusetts InstOf Tech CambridgeDept Of Brain And Cognitive Sciences

[10] Kishore, P.V.V. and Kumar, P.R., 2012. Segment, track, extract, recognizeand convert sign language videos to voice/text. (IJACSA) InternationalJournal of Advanced Computer Science and Applications, 3.

[11]Wankhade, Reshma B., et al. "Student Attendance Monitoring System Using Facial Recognition." International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Volume 3. Singapore: Springer Nature Singapore, 2022.

[12] Nanda, S. K., Ghai, D., &Pande, S. (2022). VGG-16-based framework for identification of facemask using video forensics. In Proceedings of Data Analytics and Management: ICDAM 2021, Volume 2 (pp. 673-685). Springer Singapore.

[13] Wankhade, Reshma B., et al. "A Review on Essential Resources Utilized for Face Recognition." Proceedings of the International Conference on Innovative Computing & Communication (ICICC). 2021.

[14] Aherwadi, Nagnath B., et al. "Criminal Identification System using Facial Recognition." Proceedings of the International Conference on Innovative Computing & Communication (ICICC). 2021.

[15] Gope, Birjit, et al. "Handwritten digits identification using MNIST database via machine learning models." IOP Conference Series: Materials Science and Engineering. Vol. 1022.No. 1.IOP Publishing, 2021.