International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Conversion of Sign Language into Text

1st Neelakantappa B B
Computer Science and Engineering
Malnad Colllege of Engineering
Hassan,India

2nd Akshatha N
4MC21CS010
Computer Science and Engineering
Malnad Colllege of Engineering
Hassan,India

3rd Amrutha M R
4MC21CS011
Computer Science and Engineering
Malnad Colllege of Engineering
Hassan,India

4th Anusha S K
4MC21CS018
Computer Science and Engineering
Malnad Colllege of Engineering
Hassan,India

5th Arshiya Tasneem
4MC21CS020
Computer Science and Engineering
Malnad Colllege of Engineering
Hassan,India

***Abstract*—Abstract–This project is based on mark creation of new system which will be capable to translate sign language in the text which will helpful for the hearing impaired and Speech impaired people for better communication. With reference to state of art computer vision processes along with machine learning and deep learning the proposed system detects the sign language gesture in real time. It is for this reason that the primary objective of this solution is to reduce exclusion of sign language users from other individuals within their community and develop the most convenient, efficient, and stable means of turning gestures into words.**

***Index Terms*—Sign language recognition, text conversion, computer vision, machine learning, deep learning, real time gesture recognition, assistive technology for communication, use of artificial sign language for hearing and speech impaired, human computer interface.**

I. Introduction

Sign language is an efficient, physical method of communicating which is recognized by millions across the deaf and Hard of hearing communities. Although it plays a crucial role in facilitating signers' interpersonal communication, numerous challenges emerge when digital signers interact with nonsigners, one is effectively excluded socially due to shortage and ineffective means of communication as is the case with digital signers. These challenges could be seen as strong proofs of the existence of the gap together with the necessity to adopt new methods in order to change the relationships between these entities constantly communicating with each other.

This project is the powerful and efficient signed language to text transliteration system for enhance the communication context and society integration. This system uses computer vision integrated with ML and DL to identify and analyse precisely the gestures used by individuals in sign language. As such, the system confirms accurate and genuine representation of varying gestures into textual language by the use of motion tracking, hand gesture recognition and natural language processing techniques.

Identify applicable funding agency here. If none, delete this.

New breakthroughs in AI in recent times have enhanced a boost in this technique and in many ways have paved the way for reducing limitations that are associated with gesture recognition. CNNs and RNNs introduce great practicability in solving such complexity perspectives of the sign language as the variations of handshapes, movement path, orientation, and facial expressions. The performance of transfer learning models gets to another level by the system flexibility that enables it to transform into another dataset, and language. The utility of such a system is virtually unlimited. It could alter the sphere of education allowing deaf students to manage communication in the group in some of the few inclusive classrooms. It could also transform workplaces, retail and shopping, receptionists and heath care, making the impaired person to have the best shot if not in.

But even then it has a problem with it. Some of the current challenges touching on the modern day systems include; They may lack the capacity to process in real time, they may not efficiently accommodate a large number of customers, they may not be able to differentiate between different sorts of sign languages or even sing regional variations. To address these gaps, this project used the scalable and flexible modular architectures apart from pre-processing operations which enhance recognition performances.

This work shall justify the existing playing field of technologies for the deprived Deaf and hard-of-hearing subjects by presenting a new design approach which considers technological improvement and usability. Besides closing the communication gap, it is also intended to enable people to stand on their own and create an equal society in the connection world.

II. Literature Survey Overveiw

A. Historicall developing

SLR started with gesture-based systems but eventually focused on machine learning models, like HMMs and KNN. Isolated gestures were the basis of early system models; however, due to increase in accuracy and flexibility for being used to identify dynamic gestures and multi-linguistic systems by deep learning models like CNNs and RNNs, it has gained much importance in contemporary times.

B. *Technological advancement*

The newest SLR systems use deep learning models, particularly CNNs, RNNs, and Vision Transformers (ViTs), to perceive complex, dynamic gestures. Hybrid models, like the fusion of CNN and Transformer, have achieved better accuracy. Transfer learning is also applied to deal with problems arising from multi-cultural sign language recognition to promote generalization across different languages.

C. *Common Challenges*

Some of the challenges are cultural variability in sign languages, data quality and preprocessing problems, dynamic gesture recognition, processing requirements in real-time, and independence of users. Variability in lighting, camera angles, and background noise complicates the recognition process especially in natural settings

D. *Researcg gap*

There are also limitations in cross-cultural recognition, real-time optimization of systems, and generalization across sign languages. More effort still has to be made into transfer learning and operating noise-tolerant systems, as well as fusion multimodal systems that integrate gestures with speech or text for better communication

E. *Importance of Present Research*

Current research is necessary for improving the communication of deaf people, advancing the inclusion of technology, and allowing real-time, scalable translation of sign languages. It contributes to AI-based systems for human-computer interaction and supports the access of hearing-impaired individuals to education and employment.

F. *Ethical and Privacy Issues*

It has been well established that the current SLR on integrating visual data with other sensory inputs, such as audio, haptic feedback, and text, has provided real-time contextualization of gestures in communication. Such integration is said to make the accuracy rate in complex environments much better. Some examples of this include the gesture recognition in which speech could be used for the explanation of unclear signs; accessibility tools such as virtual and augmented reality-based experiences

III. METHODOLOGY

A. *Data Collection*

It starts with the input of the training data set of each of the depictions all the symbols that are to be recognized by the model. It is even taken from different angles to guarantee that the model captures all the forms of the sign language gestures. Even all these lighting conditions and backgrounds together with variability of the orientation of objects are offered on purpose in the dataset. If thus subjected to different incidence through which the model shall learn from and through the model can identify gestures in any given environment regardless of inputs from the expert or the skilled.

B. *Feature Extraction*

Proposals for chapters and sections are made to offer an overall framework of the division of important aspects in gestures in the light of improving the performance of the given model. The features needed for implementation are derived with the help of OpenCV and Mediapipe in relation to the hand keypoints, motion trajectory and texture parameters. As seen in Table 2, these features allow the identification of these gestures with reasonable accuracy, if at all, only when performed in other settings or in suboptimal conditions.

C. *Model Development*

The recognition task use deep learning approaches with more focus on CNN on the obtained data. CNNs is declared applicable in image and video analysis mainly because it expands the spatial pyramid for hierarchal structures detection in image data . The pictorial data is labelled sequentially; the proper sign in the act is equivalent to a frame. Data preprocessing in other words, Cleaning, normalizing, augmenting and so on is the preparation phase of the model for the learning phase and the learning phase optimizes the model's accuracy and learning rate and the speed at which it gets the answer. When dealing with temporal features of gesture sequences, LSTMs or GRUs are used and therefore the dynamic motion is well captured.

D. *Gesture to text mapping*

Actually, the connection between the observed motion gesture and its textual counterpart – the affiliate description – is made very specific. This requires construction of annotated sets that would link gestures to textual labels that are significantly accurate and pertinent contexts. However, to form linguistically correct and semantically meaningful sequences, NLP methods are used on these gesture sequences. There is also provision for recognition of multiple language thus enabling easy translation of different sign languages ideal for users of different language.

E. *System Integration*

This description also describes how the gesture recognition model is implemented in a TSO system and how the TSO system utilizes the model. The gesture recognition component is assisted by a text to natural language processing unit transforming each gesture to text.

F. *Testing and Feedback*

The final test that carried out in the course of the design process is the usability test which is wide enough. Regarding the effectiveness of the developed system people offer some opinions connected with the relevance of the questions that are posed to the system, the usability of the system and the effectiveness of the system. This input is employed to counteract such diverse obstacles as latency, wrong recognition of gesture to response time. These modification and refinements provide the guarantee that the system delivers the required solution to translating sign language for contingencies in real-world scenarios.

G. *Accessibility and Ethical Considerations*

Compared to accessibility and privacy the latter is quite important and based on this an inclusive and ethical solution is made for the public. The idea it is supposed to eradicate categorization based on the performance by gender and culture and, therefore, the model embraces all. The activity of data gathering and archiving including user data is performed with due reference to applicable legal requirements like the GDPR.

IV. IV. METHODOLOGY

A. *Hardware Components*

*1) **Camera (Webcam/High-Resolution Camera)**:* For motion capture not only do you need a system to track the hand which is still and in motion for image processing you also need sufficient resolution. As for the camera you are welcome to use any USB or stand alone camera that can be integrated with this system.

*2) **Motion Sensors**:* It is optional to identify motion of a hand around besides switch on of the system if somebody is within it.

*3) **Connectivity Modules** :* These modules are capable of conveying the actual time data to another cloud service or server for further anlysis and result.

B. *Software Components*

*1) **Python**:* This is the language that you use, to put down on paper the code that your system requires. It has proved out to be a fundamental software application since, among many other activities, you can code all the algorithms, train and also deploy machine learning, image processing, and data set managing.

*2) **TensorFlow** :* This framework is employed in building signs language recognition models, training as well as deployment. First, ever since machine learning is in vogue, it could be mentioned that this type of architecture is one of the leading software tools needed for accomplishing any venture that seeks to adopt this phenomenon.

***OpenCV**:* Morphology is an extensively popular big library for image enhancement, separating and features retrieval for computer vision proposals; predictability serves an important software tool responsibility in any direction or corner of project that are linked or correlated to the line of computer vision.

*3) **Numpy**:* In fact, it is a basic package in machine learning used for Number Gimmicks and multi-dimensional array. It is required when operating with data and their operation which is a major practice throughout the model building process.

*4) **Matplotlib**:* This library is used for making the plots the other form of visualization. Not only it enabled to observe model behaviour and data sets and even pointers, and on a clean interface, but it is also considered to be a rather helpful software tool to compare a huge amount of data and different models.

*5) **IDE/Code Editors**:* Other computing tools they include Jupyter, Notebook, Visual Studio Code which are also used in writing, testing and debugging in python codes. They provide the premises on which you will be using in performing your task.

*6) **Dataset Preparation**:* To begin with, with the help of NumPy and Open CV for pre processing your datasets are going to be generated locally using Python scripts. For mids in pms, it is a step within the project which involves the use of software tools to format data for training models or for use in model assessment.

*7) **ResNet**:* In order to overcome the low accuracy of gesture recognition addressing the issue we use transfer learning.

*8) **spaCy**:* Combined with more properly known gestures in order to achieve further comprehensibility and readability of the text by the format and processing of the text.

V. SYSTEM WORKFLOW OVERVIEW

The system effectively resolves the communication barrier between the hearing and speech impaired people as it translates the sign language gesture into text simultaneously. Therefore, it facilitates social interaction on education, employment, and social-communal platforms, including for those using sign language, and those who do.

A. *Input Module*

It tracks hand and body movement with great efficiency and high rates of accuracy. Realtime movement capture can be captured by an RGB or depth camera Click here to view the methodology. Several applied in precision include the Leap Motion Controllers, or motion sensor gloves. Video preprocessing tools also filter out noise from video apart from being able to isolate gesture out of a video background and also analyse it further.

B. *Gesture Recognition Module*

This module then further takes the captured gestures and proceeds to interpret meaning from the same. Pre-processing techniques which are used in the computer vision include erasing the background, processing the gestures to a common scale, and erasing the body parts and hands to isolate them. Such algorithms employed in this module are CNN others are as follows; the spatial and temporal features similar to key point detection as in MediaPipe. With the right words or letters to map these features that have been extracted, it will be possible to map further with the use of other algorithms such as Random Forest and SVMS aside from deep learning algorithms such as recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks. The system assumes a pre-trained model, tuned for use in particular datasets specific to the selected sign language for instance the American Sign Language.

C. *Text Conversion Module*

When signs are detected, the module of text translation translates those signs into written forms. The system can process either word- and letter-based sign language; in this way, different styles are considered. Real-time or batch input processing assures that communication continues with no disruption at all. Then, it assures the output of those recognized signs through proper textual display in order to make a meaningful relationship between signed and written forms.

D. *Output Module*

The output module will not only display the converted text on a screen but also send it to devices like a smartphone, computer, or dedicated hardware. This module may also be capable of text-to-speech because all identified gestures can then be voiced for increased accessibility. The module will make the output clear, direct and easy to understand for non-sign language users

VI. RESULTS

A. *Performance Evaluation*

While assessing the enhancements to the suggested system, beneficial outcomes such as accuracy, precision, the rate of recall and F1-score were taken into consideration; therefore, it was possible to provide a solution in addition to understanding the potential advantages. For the classification of raw data from the test dataset for gesture recognition capability of the proposed system irrespective of the condition, classification accuracy attained is 89.5 percentage. Nevertheless, improvements in the accuracy and F1 score support these performances, and show even more significant limitation of the false positive and false negative rates in the system.

B. *Comparision with Other Systems*

This shall be a marked improvement in speed, effectiveness, as well as the usability of this model compared with most models used in earlier sign language recognition. Other systems prior to this one have lower overall accuracy as they rely on a smaller amount of data and employ comparably less complicated algorithms; this system employs CNN-LSTM structures totaling smart hand tracking methods and raises the set accuracy bar to 89.5 percentage. The speed of processing is also unusual in comparison with more or less previous models which consume much computation; this kind of design works fluent in average hardware with a latency in real time The choice of cheapest devices like web-cams and usage of OpenCV library diminished cost of corresponding hardware installation in comparison with similar systems. Due to simplicity of the user interfacing, this had improved the use since as compared to the actual difficult setting of the traditional system.

C. *Challenges and Limitations*

Still, considering these said changes and having an almost ‘bullet-proof’ design, the system encounters several challenges. Nevertheless, one can mention the only significant drawback that has been possible to identify – our model’s performance greatly depends on the lighting conditions and if the image although was taken under average lighting – hand key point detections may be imprecise. Moreover, and more seriously, is the refining of complex gestures that make use of a couple of fingers of both hands or any other combination and/or timing, which remains a hot potato given poor training data. Still there is one more issue which is related to working with the streams of continuous video inputs as sequences, frames per second rather than frame latency impacts actual working time. Such limitations require to be tackled in order to sustain the increase in reliability and efficiency of the system and the satisfaction of the final beneficiaries.

D. *Future Improvements*

However, to overcome these challenges, the following few enhancements are suggested for the future: Input of more gesture data from the domain, to more gestures, to more users, at different ages and sexes, and in different environmental situations will improve reliability of the model. Enhancing the complexity of these models only like s3D CNNs, transformer-based architectures, or the use of attention mechanism may also optimize the recognition of complex and temporal-varying gestures. Furthermore, making the system general so that the two sign languages we developed detectors for, to be used in other domains, using domain adaptation techniques to detect not only ASL or BSL but also more ISL and any blend of them with other sign languages across the globe will make the system popular. Currently, using concrete types of devices such as edge AI devices or dedicated GPUs, it is possible to reduce latency for a sequence of consecutive video streams to enhance real-time performance.

VII. CONCLUSION

As a result, the sign language to text transference in the past was one of the most striking developments that ever occurred in the communication sector, which brought the course of vision-impaired people closest to non-impaired ones. With the coexistence of computer vision technology, machine learning development, and deep learning, the process utilizes state-of-the-art technology in order for the system to accurately record signs from a sign language to various formats of text. Such activities not only make communication easier but also help achieve a consistent environment in education, work, and other sectors. The proposed enhancements comprise the fusion of different datasets, a manifestation of more complex architectures such as 3D CNNs and transformers, and multi-language sign language recognition for enhancing the energy efficiency of the device. There is also an integration offered which connects the system to edge AI devices or to GPUs, which would help to offset the delay in data processing and improve the accuracy of the system. In the final analysis, this project is indicative of the ability of AI-powered systems to create a more inclusive international platform where the voice of people with disabilities is clearly heard and their views are fully appreciated. The project will enable a global switch over only if the hurdles that have been identified are fixed and the upcoming technical advances are wisely employed, making the project the future of worldwide sign language recognition.

VIII. References

[1] Poorvi Desai, Preksha Mohre, Soumya Bani, Varada S. Naik, and Sudha P. R., ”Conversion of Sign Language to Text for Deaf and Dumb.” [2] Abu Saleh Musa Miah, Md. Al Mehedi Hasan, Yoichi Tomioka, and Jungpil Shin, ”Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network.” [3] P. Subha Rajam and Dr. G. Balakrishnan, ”Real Time Indian Sign Language Recognition System to Aid Deaf-Dumb People.” [4] Camelia Avram, Laura-Flavia Pacurar, and Dan Radu, ”Sign Language to Text Conversion Using CNN (For Deaf and Mute People).” [5] Abhay Kumar, Anurag Gupta, Bipin Chaurasia, and Chaitanya Mishra, ”Conversion of Sign Language to Text for Deaf and Dumb.” [6] Medhini Prabhakar, Prasad Hundekar, Sai Deepthi B. P., Shivam Tiwari, and Vinutha M. S., ”Sign Language Conversion to Text and Speech.” [7] Nada Bahaa Ibrahim Ahmed, ”Advances, Challenges, and Opportunities in Continuous Sign Language Recognition.” [8] Mahesh Kumar N. B., ”Conversion of Sign Language into Text.” [9] Tengda Zhao, Jinxin Ma, Hongkai Chen, Kaixin Liu, Huajie Shao, and Qian Wang, ”Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System.”