JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TRANSLATING SILENCE: ENHANCING UNDERSTANDING THROUGH SIGN LANGUAGE TRANSLATION

Rampur Srinath[1], Chinnaswamy CN [2], Nishchay D[3], Noel Joicil Vas[4], Prajwal HM[5], Surya S[6]

[12]Associate Professor, Department of Information Science and Engineering., NIE, Mysuru, Karnataka.

[3456]UG student, Department of Information Science and Engineering., NIE, Mysuru, Karnataka.

ABSTRACT

This comprehensive review synthesizes insights from four research studies that collectively highlight the potential of developing real-time applications to facilitate communication between individuals using sign language and those who do not. The first paper proposes an application utilizing convolutional neural networks (CNN) and Text-to-Speech translation to convert live video inputs of gestures into speech, enabling seamless real-time conversations [1]. The second paper echoes this approach, emphasizing the importance of real-time functionality and the translation of gestures to speech via CNN and Text-to-Speech technologies [2]. Similarly, the third paper reinforces the need for technical advancements to support real-time communication between signing and non-signing individuals, suggesting the use of CNN for gesture recognition and Text-to-Speech for conversion to speech [3]. The fourth paper also focuses on developing a real-time application, proposing CNN and Text-to-Speech to translate gestures from live video inputs into speech, thereby facilitating effective communication [4]. Collectively, these studies underscore the significance of integrating advanced technologies such as CNN and Text-to-Speech translators to bridge the communication gap between the speaking and non-speaking world by enabling real-time translation of sign language gestures into spoken language.

INTRODUCTION

The development of sign language technologies has become crucial to bridge the communication gap between the speech and hearing disabled communities and the speaking world. Sign languages, such as American Sign Language (ASL) and Indian Sign Language (ISL), have their own grammar and lexicon, making them complex for the general population to understand. In India, despite the significant number of people with hearing disabilities and the availability of resources

like the ISL Dictionary, a substantial communication barrier persists. The first edition of the ISL Dictionary, comprising 3000 terms, was launched in March 2018, followed by a second edition in February 2019, highlighting the government's efforts to support the learning of sign language [1]. However, the speaking masses still have little to no comprehension of ISL, underscoring the need for advanced technologies to facilitate communication. This paper proposes an interface to translate sign language into speech in real-time, focusing on capturing gestures, recognizing them using convolutional neural networks (CNN), translating signs to text, and converting text to speech [1].

Non-verbal communication, particularly through gestures, plays a vital role in human interactions. Human-Computer Interaction (HCI) technologies often rely on gestures as inputs. This model aims to convert speech to text and gestures, and vice versa, using real-time hand gesture recognition with CNN. Gesture recognition technology has evolved significantly since its inception with glove-based control, leading to various techniques for detecting hand and face gestures. Despite these advancements, challenges such as lighting conditions, background separation, and processing time persist, necessitating further improvements to enhance the system's accuracy and usability [2]. The proposed system addresses these issues by using image processing and computer vision techniques to capture hand gestures in real-time, which are then processed by CNN algorithms for accurate recognition and translation.

Statistics reveal a significant population in India with hearing and speech impairments, facing numerous challenges in daily life due to communication barriers. According to the 2011 census, over 50 lakh people in India suffer from hearing disabilities, while around 20 lakh people have speech impairments. These barriers significantly impact their education and employment opportunities, with 63% of deaf and mute individuals being unemployed and 30% never having attended school [3]. This paper discusses the importance of developing technologies to translate sign language and proposes solutions for real-time sign language recognition and translation to aid these communities in their daily interactions. By providing a means for the non-hearing and non-speaking communities to interact seamlessly with the hearing and speaking population, the proposed system aims to enhance their quality of life and reduce their dependency on others.

Communication is essential for expressing thoughts and information, yet it is challenging for the deaf-mute community, leading to social inequalities and limited opportunities. Despite various technological advancements, Indian Sign Language (ISL) remains underexplored. Sign language uses visual-manual modalities to convey meaning, incorporating gestures, facial expressions, and body language. Different countries have developed their own versions of sign language, such as ASL, Chinese Sign Language (CSL), and Japanese Sign Language (JSL), contributing to the complexities of developing a generic sign language translation system. This paper highlights the need for effective sign language translation systems and reviews various sensor-based and vision-based techniques to find the best approach for bridging the communication gap between the deaf-mute community and the general population [4].

LITERATURE SURVEY

A number of technologies have been developed for the translation of sign languages across the world. Some of these technologies make use of hardware while others are purely algorithmic software applications. Systems have been proposed and developed wherein glove-like extensions can be worn by the signer so that the gestures may be captured, identified, and consequently translated. These gloves help capture the hand movements of the signer with heightened accuracy and post which these gestures are translated through appropriate algorithms. The shortcoming with such devices is that they only capture the gestures made via hand and not the signs exhibited via expression and body language. Therefore, as an alternative to resolve this, pictures or videos are taken as inputs [1].

In such applications, the user input is taken via cameras on either mobile devices or specialized devices with inbuilt cameras are utilized. For instance, a Tamil Nadu based team worked on a mobile-based application for which video inputs are utilized [2]. One of the methods of capturing inputs for this application is to take video inputs by attaching a camera to a cap. The signer would wear this cap and her/his actions would be captured from the top view by this modified cap.

As previously discussed, all sign languages comprise of hand gestures, facial expressions, and body language. Models of this nature focus on hand-based gestures and do not widen the scope to the other aesthetics of signing. However, these models do provide results with varying accuracy within the scope of their work. If body language were to be taken into account, then the approach should be changed to accepting video inputs which capture the facial expressions as well as hand gestures of the signers. This would enable the capturing of the other factors in signing and make it possible to work on the data and recognize the signs more effectively [3].

This is one of the primary challenges faced in sign language translation, that is, the variety of factors involved in signing. Besides the basic aesthetics of signing in any of the sign languages, there are other varying factors to be considered. There exist individualistic variations in the signing of the same message. That is to say that if two people were to sign the same word or sentence, there would be variations in the signs exhibited purely out of the differences in their individual behaviors and body language. There also exist physical and structural differences in the hands of individuals. These factors matter when feeding signs in as an input to an algorithm or an application. An algorithm is only as good as the training database. Hence, it must be seen to it that the training database comprises of a variety of signs from a variety of signers to capture maximal possible variations [4].

Neural Networks or Artificial Neural Networks (ANN) are a biomimicry-based computing networks system, meaning they are based on the naturally occurring connection between neurons in animal brains. As a result, these are connectionist systems. ANN can be adapted to aid in deep learning giving rise to the entire category of Deep Neural Networks. Deep neural networks can be implemented via a variety of mathematical methods. Convolutional Neural Networks (CNN) are one such implementation. Space Invariant Artificial Neural Networks (SIANN)l Networks (SIANN), refers to a fully connected system where the networks are based on convolutions in one or more layers of neuron connectivity.

PROPOSED SYSTEM:

A heavy amount of work has been done for the translation and transcription of American Sign Language (ASL). By comparison, the work done on Indian Sign Language (ISL) has been limited, offering greater scope for innovation. Although the proposed model here could be designed to adapt to any sign language by training it on a relevant database, this model specifically focuses on translating ISL to spoken English. The implementation of this model involves several key steps: creating or compiling a database of ISL signs and gestures, recognizing gestures from input feeds using a neural network-based algorithm, processing these inputs using classification and machine learning techniques to train the model, generating a text translation of the input signs or gestures, and converting this text to speech.

The initial phase of the model's development will focus on capturing and translating stationary gestures from still images, as these are simpler and more prevalent for ISL alphabets and basic words. Complex grammatical structures, such as sentences, require dynamic gestures involving motion, which will be addressed in later phases. The dataset should ideally consist of a range of predetermined gestures performed by multiple individuals to account for variations in style, body language, and physical attributes. This approach ensures that the tool can accurately recognize and translate signs regardless of the signer.

To further enhance the model's accuracy and robustness, the dataset should include variations in graphic factors like the signer's attire and skin texture, as well as differences in image quality. A diverse dataset helps prevent overfitting and reduces bias in the trained model, leading to more reliable outcomes. The dataset will be divided into training, validation, and testing categories, with the training and validation data labeled to facilitate classification. The testing data, being a randomized mix without labels, will be used to assess the model's accuracy.

Once the dataset is compiled and segregated, the model can begin training using convolutional neural networks (CNN). The visual inputs will be processed through the CNN algorithm, which will handle input data processing, gesture classification, and model training through iterative learning. Machine learning techniques integrated into the algorithm will further refine the model, enabling it to adapt and improve with continued use.

The output of the CNN algorithm will be a text translation of the recognized sign. This text output will then be processed through a text-to-speech conversion model to generate the final audio output. Various libraries and APIs are available for text-to-speech translation, many of which can be customized to meet specific requirements. These tools allow the derived text to be converted into real-time audio output, fulfilling the system's objective.

In summary, the development of this sign-to-speech translation system for ISL involves creating a comprehensive and varied dataset, employing CNN for gesture recognition and classification, and using text-to-speech conversion to generate the final audio output. This multi-step approach, with its focus on capturing a wide range of variations in gestures and appearances, aims to produce an accurate and user-friendly translation tool for the deaf and mute community in India.

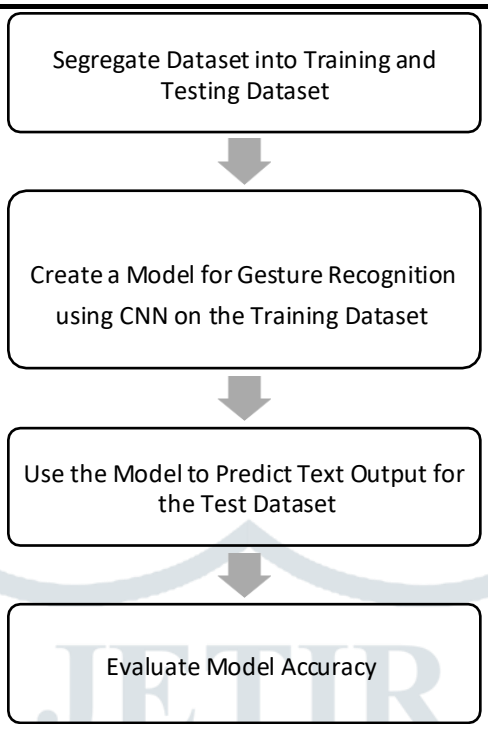

Fig. 1: Flowchart of the proposed sign to speech translator

Fig.2: Model Prediction on Live Camera Prediction: "M"

EXISTING SYSTEM:

Challenges Addressed

1. Inadequate Capture of Non-Manual Signals:

- Solution: The proposed system offers a comprehensive capture method that uses video inputs from mobile devices or specialized cameras. This allows for the integration of facial expressions and body movements with hand gestures, providing a fuller representation of sign language.

2. Limited Training Data Diversity:

- Solution: The proposed system offers the creation of a diverse dataset comprising signs and gestures from various individuals. This dataset includes variations in physical attributes and personal signing styles, ensuring the model can learn to recognize signs from diverse signers, thereby enhancing robustness and accuracy.

3. Handling Dynamic Gestures:

- Solution: The proposed system offers a phased approach, initially focusing on capturing and translating stationary gestures. As the model develops, it will incorporate video processing techniques to handle dynamic gestures, allowing for the recognition of more complex sign language structures.

4. Integration of Multimodal Data:

- Solution: The proposed system offers the use of Convolutional Neural Networks (CNNs) to process visual input data. This includes integrating image processing techniques like skin detection and edge detection, ensuring accurate capture and classification of gestures from the visual data.
-

5. Real-Time Processing and Latency:

– Solution: The proposed system offers real-time processing capabilities, ensuring it can handle video inputs and provide immediate translations. By using efficient algorithms and powerful computational resources, the system minimizes latency, initially focusing on stationary gestures to refine and optimize the model before handling dynamic gestures.

Key Features:

– Utilizes video inputs from mobile devices or specialized cameras to capture hand gestures, facial expressions, and body language. Compiles a diverse dataset with signs and gestures from various individuals, including variations in physical attributes, signing styles, and image quality.

– Initially focuses on stationary gestures, gradually incorporating dynamic gestures. Integrates image processing and computer vision techniques such as skin detection and Canny Edge detection.

– Utilizes CNNs to process visual inputs, classify signs and gestures, and generate text translations. Trains the model on labeled data that includes contextual information to understand the nuances of different signs.

– Designed to operate in real-time, processing video inputs and providing immediate translations with minimal latency. Incorporates text-to-speech libraries and APIs to convert the translated text into spoken English.

– The model is designed to be scalable, allowing for adaptation to different sign languages with appropriate training datasets. Aims to develop an easy-to-use and widely accessible application for the deaf and mute community.

CONCLUSION:

The utilization of technology in bridging communication barriers for the speech and hearing-impaired communities is pivotal in

India, where biases against these groups persist. By providing tools that facilitate interaction between these communities and the broader population, we can work towards leveling the playing field and fostering inclusivity. The proposed model offers real-time functionality and potential accessibility through mobile applications, which could significantly enhance communication accessibility. However, challenges such as ensuring translation accuracy, particularly in dynamic real-world scenarios, and the reliance on an active internet connection pose hurdles. Despite these limitations, iterative improvements driven by user feedback and the integration of machine learning algorithms hold promise in refining translation accuracy over time.

While the current system lacks context recognition and primarily focuses on translating distinct hand gestures, it serves as a valuable communication platform for the speech and hearing-impaired communities. The absence of a complete two-way translation system in Indian Sign Language highlights the need for further advancements in this field. By leveraging advanced technologies like neural networks and smart sensors, future iterations of the system can enhance accuracy and efficiency, thereby improving the quality of life for these communities. Moreover, incorporating user-friendly features and expanding the system's capabilities to include chat, gaming, and other applications could enhance its utility and accessibility.

In conclusion, the convergence of technology and communication offers immense potential in fostering inclusivity and understanding among diverse populations. By addressing the unique challenges faced by the speech and hearing-impaired communities, these systems contribute to breaking down barriers and promoting social cohesion. Continued research and development in this field are essential to further improve the accessibility and effectiveness of communication tools for all individuals, regardless of their abilities or disabilities.

REFERENCES

[1] Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma (2020). Sign Language to Speech Translation.

[2] Dr. S. Pari Selvam, Dhanuja.N, Divya.S, Shanmugapriya (2020). An Interaction System Using Speech and Gesture Based On CNN

[3] Yuvraj Grover, Riya Aggarwal, Malas, Deepak Sharma, Prashant K. Gupta (2021). Sign Language Translation Systems for Hearing/Speech Impaired People: A Review 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)

[4] Gunasagari G. S, Abhijna Yaji, Achuth M (2021). Review on Text and Speech Conversion Techniques based on Hand Gesture. Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) IEEE Xplore Part Number: CFP21K74-ART; ISBN: 978-0-7381-1327-2