SMART SIGN: Live Sign Language Interpretation For Barrier-Free Video Conferencing

Mitali Vishwekar[1], Daniel Injeti[2], Dr. Rakhi O. Gupta[3], Nashrah Gowalker[4]

[1]Student, I.T Department, Kishinchand Chellaram College, HSNC University, Mumbai, India
[2]Student, I.T Department, Kishinchand Chellaram College, HSNC University, Mumbai, India
[3]Co-ordinator, I.T Department Kishinchand Chellaram College, HSNC University, Mumbai, India
[4]Assistant Professor, I.T. Department Kishinchand Chellaram College, HSNC University, Mumbai, India

Abstract
In the digital era, effective and inclusive communication is essential, particularly in video conferencing platforms widely adopted post-pandemic for professional and social interactions. However, these platforms often lack accessibility features for individuals with hearing and speech impairments. This research presents a lightweight, real-time sign language translator designed for seamless integration with popular video conferencing platforms like Google Meet, Microsoft Teams, Zoom, and Discord. Utilizing deep learning algorithms, image processing techniques, and virtual camera technology, the system translates sign language into written captions, displayed in real-time. By fostering accessibility and inclusivity, this innovation empowers deaf and mute individuals to participate actively in virtual meetings, bridging communication gaps and advancing equal opportunities for all.

Keywords: Sign Language Translation, Hand Gesture Recognition, video-conferencing, Sign Language Translator for Meeting Apps, Real Time Sign Language Translation, Google Meet, Microsoft Teams, Zoom, Alphabetical sign language, Sign Language Interpreter.

Introduction

A. Historical Development of Sign Language

IN Early Theories and Naturalness of Gestures. Plato's Cratylus (5th century BC) introduced the concept of gestures as a natural communication method, highlighting their intuitive connection to human expression when speech was unavailable. The Birth of Formal Sign Language Systems was the pioneering efforts of Pedro Ponce de Leon in the 16th century, who created the first manual alphabet, and Juan Pablo Bonet's seminal 1620 publication on sign language phonetics marked the beginning of structured communication methods for the Deaf.

B. Global Diversity and Cultural Importance of Sign Language

Sign Language Usage around the World is More than 300 unique sign languages are actively used by approximately 70 million Deaf individuals globally, reflecting rich cultural and linguistic diversity. The Emergence and Significance of American Sign Language (ASL) has Originated in the early 19th century at the American School for the Deaf in Connecticut, ASL has become a critical component of Deaf culture and identity in the United States and Canada.

C. Advocacy and Legal Recognition of Sign Language

Acknowledgment as Full-Fledged Languages. The United Nations Convention on the Rights of Persons with Disabilities recognizes sign languages as equal to spoken languages, promoting the linguistic identity of Deaf communities. Raising Awareness through the International Day of Sign Languages. Established by the UN General Assembly in 2017, September 23 highlights the importance of sign language in achieving human rights for the Deaf community.

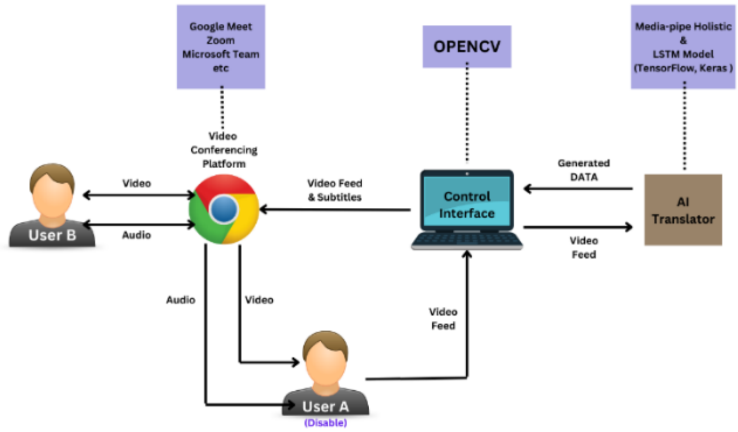

Figure 1: Flow Diagram of Idea Purposed.

D. ASL as a Technological Focus and Bridge to Inclusivity

Rich Datasets for Research and Development in which ASL's extensive datasets and historical prominence make it a leading sign language for the development of innovative technological solutions. Real-World Impact of Real-Time Translation Systems with Modern tools empower Deaf individuals by enabling seamless interaction with hearing counterparts, fostering inclusivity in social and professional settings.

E. Technological Advancements in Sign Language Translation

Early Innovations in Translation Tools was the development of tools such as the robotic hand RALPH (1977) and motion-detecting gloves like the CyberGlove laid the foundation for automated sign language interpretation. Modern Applications Leveraging Machine Learning and Computer Vision Advances in deep learning, image processing, and computer vision now enable real-time translation systems that significantly improve communication accessibility.

LITERATURE REVIEW

A. Paper Title 1: Real-Time Sign Language Detection using Human Pose Estimation.

The central theme of this paper is to develop a sign language recognition model that can operate in real-time directly from the web browser, specifically for video conferencing software. This model uses unidirectional LSTM with one layer and 64 hidden neurons on the DGS corpus dataset with 301 videos. It is observed that using a more diverse set of landmarks with a single point per hand, the system performs much better. Here, the parameters are increased by calculating 51, 842 features, with an accuracy of

91.53% and latency of 3537 ms. Both wrists have higher weights and are asymmetric, concluding that the selection is based on the user using the dominant hand for the gesture. 4,138 signing sequences totaling 11.35 seconds in duration make up the test set. Many mistakes might occur when testing the model. Using the Pose-All model resulted in fewer errors of most types except the bridge and skipped. Here the processing is done on the device, which is faster than processing on the videoconferencing servers. This paradigm has drawbacks since it can be too sluggish to use on mobile devices or devices without hardware acceleration like a GPU. According to tests, Google Meet, Zoom, and Slack can be tricked into thinking a person is speaking by broadcasting audio at 20KHz, which is inaudible to people, but if the audio is cropped to a range that is audible to humans, the system may fail. [9]

B. Paper Title II: Real-Time American Sign Language Recognition with Faster Regional Convolutional Neural Networks.

The article describes the creation and application of convolutional neural network-based ASL translators that can categorize spatial data. The proposed system contains a pre-trained GoogLeNet architecture that was developed using the ASL datasets from Surrey University and Massey University as well as the ILSVRC 2012 dataset. The pre-trained model is modified for this purpose via transfer learning to accommodate more specialized or niche data. CNNs and RNNs are used in the system to extract spatial and temporal characteristics, and a CNN is used to categorize ASL letters. To extract more pertinent characteristics from the frame, backdrop, and other body elements are eliminated after extracting the frames for each gesture from many video sequences. This method can be extended to sentence-level sign language translation, and future work can focus on combining the CNN and RNN models into a single model for better performance. The system shows promising results in classifying ASL letters with high accuracy, and further development could improve the system's overall performance.[10]

C. Paper Title III: Relevant Features for Video-Based Continuous Sign Language Recognition.

This study describes the development of a continuous sign language recognition system utilizing Hidden Markov Models (HMM). A vocabulary of 97 German Sign Language (GSL) signals was used in the system's architecture to enable it to detect phrases in sign language. The system employs feature vectors that represent manual sign settings as input for training and recognition, and beam search is used to reduce computing complexity while performing the recognition job. A monochrome video camera was integrated into the device to capture images, and the device utilized plain-colored cotton gloves to accomplish real-time picture segmentation and feature extraction. The researchers studied the effects of various characteristics on the identification outcomes. The experiments conducted revealed that the system's precision rate was 94%, utilizing a lexicon of 52 gestures and all available characteristics. A unique Hidden Markov Model (HMM) represented each signal in the system. A language model will be incorporated into the system's recognition process in the future, and it is anticipated that the incorporation of syntactic and semantic data would further enhance recognition performance. The findings of this study are encouraging because they may allow deaf and hard-of-hearing people to continuously and automatically recognize phrases written in sign language, boosting their capacity to interact with the hearing community. [11]

D. Paper Title IV: Glove Based Sign Language Interpreter for Deaf and Aphonic Peoples:

The aim of this study is to create a system utilizing a glove that can effectively translate communication for individuals who are deaf and mute, allowing them to interact with others. The user's hands and fingers may be tracked as they move and are positioned using the accelerometers included into the glove. The gadget can understand hand gestures and translate them into words by measuring these motions. Both hearing and speaking persons and deaf-mute people may communicate more affordably and conveniently with the help of this method. The initiative focuses especially on Indian Sign Language (ISL), where names or English phrases without a conventional sign are written out using fingerspelling. This system identifies the user's gestures as they are being made since it employs real-time signal input and is data dependent. In this system, a microcontroller, an accelerometer, a speech module, an LCD, a speaker, and a regulated power supply are all employed as hardware elements. Communication between persons who are deaf-mute and those who are not can be facilitated by the glove-based interpretation system created as part of this research. The system's ability to detect complicated motions and signals can be expanded in the future, and it can also be integrated with other communication technologies.[12]

E. Paper Title V: Sign Language Translator

The approach outlined in this paper is intended to get over the communication difficulty experienced by deaf and mute people who are unable to verbally communicate. Instead, they use non-verbal communication techniques like sign language. Nevertheless, not everyone is able to comprehend sign language, which can make communication challenging. The technology can transform sign language fingerspelling into a common language that is simple for others to comprehend in order to solve this problem. When communication must happen fast if there is no translator present, this can be quite useful. The finger spells may be swiftly and correctly translated into words by the system, improving communication. Moreover, the system can fingerspell words that have been typed or printed. This function can be particularly helpful when conversing with deaf and mute people who are conversant in sign language but are unable to read. Effective communication between deaf and mute people and those who don't understand sign language is made possible by this two-way communication channel.[13]

Objective of Study

The primary objective of this study is to develop a lightweight, real-time sign language translator for video conferencing platforms, enabling seamless and inclusive communication between Deaf and hearing individuals. Specifically, this study aims to:

A. Communication Barriers:

Mute and Deaf individuals often face difficulties in video calls due to the lack of intuitive tools that facilitate sign language communication. Unlike spoken language, sign language relies on visual gestures and facial expressions, which are not directly understood by those unfamiliar with the language. This gap forces many Deaf individuals to rely on text-based communication or third-party interpreters, both of which disrupt the flow and intimacy of conversations. The absence of integrated solutions hinders full participation in professional and social interactions.

B. Real-Time Translation Needs:

Timeliness is critical in effective communication, yet many existing sign language translation tools fail to deliver accurate outputs without delays. Even a few seconds of lag in translation can cause confusion and disrupt the flow of live conversations. This issue is further compounded in scenarios with rapid exchanges, such as group discussions or professional meetings. Real-time translation requires not only quick processing of gestures but also accurate interpretation of context, making it a highly challenging yet essential feature.

C. Limited Support for Multiple Sign Languages:

While over 300 sign languages exist globally, most existing systems primarily support dominant languages like ASL or BSL. This leaves millions of individuals who use other sign languages without access to translation tools. Furthermore, sign languages often have regional dialects and variations, making it challenging to develop comprehensive systems that cater to diverse linguistic needs. Addressing this limitation requires extensive datasets and adaptable algorithms that can accommodate the linguistic richness of global sign languages.

D. Integration Challenges with Video Conferencing Platforms:

Most video conferencing platforms lack built-in tools for sign language translation, requiring users to rely on external applications or devices. These setups often involve additional hardware, complex configurations, or virtual camera systems, making them inconvenient and inaccessible for many users. Seamlessly integrating translation tools directly into platforms like Zoom or Microsoft Teams would significantly enhance usability and ensure wider adoption.

E. Accuracy and Reliability Issues:

Accurate recognition of sign language is a significant hurdle due to variations in signing styles, environmental factors like lighting and camera angles, and the intricate combination of hand movements, facial expressions, and body language. Many systems fail to consistently deliver reliable results in real-world conditions, particularly in dynamic video conferencing environments. High accuracy and reliability are essential for fostering trust in these systems and ensuring effective communication.

F. Inclusivity in Digital Communication:

Without accessible tools, people with hearing and speech impairments are often excluded from digital communication, perpetuating social and professional inequalities. Video conferencing has become a critical medium for education, business, and social interaction, but its lack of inclusivity for Deaf and mute individuals limits their participation and contribution. Bridging this gap with inclusive technologies can empower these individuals, ensuring equal opportunities in an increasingly digital world.

TECHNOLOGIES USED

This project proposes a real-time sign language translator for video conferencing platforms, enabling seamless communication for mute and Deaf individuals by converting gestures into text captions. Utilizing OpenCV for video input, Media-Pipe Holistic for gesture tracking, and an LSTM model with TensorFlow for sequential recognition, the system ensures high accuracy, supports diverse sign languages, and fosters inclusivity in virtual meetings. Following are the technologies user and Purpose solution after it:

A. OpenCV:

We will utilize OpenCV for input video sources since it is a large library that provides many methods for image and video processing. We capture video from the camera using OpenCV. A camera is utilized as an input source in the video conferencing platform. Our camera's primary role is to integrate with any video conferencing platform.

B. Media-pipe Holistic:

Sign language recognition simultaneously recognizes a person's posture, facial features, and real-time hand tracking. Holistic is one of the pipelines with enhanced face, hand, and posture elements, providing holistic tracking where the model can recognize hand and position of the body alongside facial landmarks at the same time. Media-Pipe Holistic will be used to recognize and track multiple human body parts and their movements in real-time video streams. This enables the tool to capture hand gestures and movements during sign language and pass them through the neural network to identify the corresponding sign language phrase. By using Media-Pipe Holistic, the tool is able to accurately capture the movements of the hands and other body parts, which is essential for accurate sign language recognition.

C. LSTM model:

A type of recurrent neural network (RNN) that can easily remember past inputs that is useful for processing sequential data. LSTM artificial neural network is mostly utilized in artificial intelligence and deep learning. For LSTM we will use tensorflow and keras. We will import dependencies for LSTM that are sequential models, LSTM layer and dense layer. After building LSTM we will train our LSTM Learning Model. Then we will make sign language predictions to check the performance of a model after that we save our model weights.

D. TensorFlow:

TensorFlow, a deep learning framework, was used to create, test, and operate our LSTM layers. TensorFlow is a prominent open-source toolkit for constructing machine learning and deep learning models that was launched in 2015 by the Google Brain team. It is based on the Python programming language and builds models by doing numerical computations on data flow graphs. These are the key characteristics of TensorFlow. It performs well with multidimensional arrays. It enables computing scalability across machines and massive data volumes. It allows for quick debugging and model creation. It has a big community and includes a Tensor-board to display the model. Face identification, language translation, fraud detection, and video detection are some of its uses.

E. Proposed Solution:

This project aims to create a real-time sign language translator for video conferencing platforms, helping mute and deaf individuals communicate effectively. Key aspects is to Converts sign language gestures into text captions instantly during video calls. Supports various sign languages to cater to diverse users. Make it simple to set up and use with existing video conferencing tools. Utilizes advanced machine learning for precise and reliable gesture recognition to get High Accuracy. Promotes equal participation for hearing and speech-impaired individuals in virtual meetings.

COMPARATIVE STUDY

To understand the effectiveness and uniqueness of our real-time sign language translator for video conferencing platforms, a comparative study with existing solutions is essential. Here are some of the top Company with their project features and limitations that makes our project a unique as follows:-

PLATFORM	FEATURES	LIMITATIONS
Google Meet	This platform supports live captions, you have to hire a sign language interpreter to the call.	they don't have built-in sign language interpretation
ZOOM	This platform supports live captions, you have to hire a sign language interpreter to the call.	they don't have built-in sign language interpretation
Microsoft Teams	This platform supports live captions, you have to hire a sign language interpreter to the call.	they don't have built-in sign language interpretation
Convo	Convo is a Video Relay Service (VRS) provider that offers live interpretation for the	They don't have built-in sign language interpretation the communication is done

	Deaf and hard of hearing.	via video calls with the help of interpreters.
Purple Communications	Purple Communications is a Video Relay Service (VRS) provider that offers live interpretation for the Deaf and hard of hearing.	They don't have built-in sign language interpretation the communication is done via video calls with the help of interpreters.
ZVRS	ZVRS is a Video Relay Service (VRS) provider that offers live interpretation for the Deaf and hard of hearing.	They don't have built-in sign language interpretation the communication is done via video calls with the help of interpreters.
Sorenson	Sorenson is a Video Relay Service (VRS) provider that offers live interpretation for the Deaf and hard of hearing.	They don't have built-in sign language interpretation the communication is done via video calls with the help of interpreters.

.

CONCLUSION

This project successfully demonstrates the potential of real-time sign language translation to bridge communication gaps for mute and Deaf individuals in virtual environments. By integrating advanced technologies like OpenCV, MediaPipe Holistic, and LSTM neural networks, the system achieves precise and reliable gesture recognition, transforming sign language into text captions seamlessly during video conferencing. The solution supports multiple sign languages, ensuring accessibility for diverse users, and is designed for easy integration with popular video conferencing platforms. By promoting inclusivity and empowering individuals with hearing and speech impairments, this project represents a significant step toward equal participation in digital communication and sets the foundation for further advancements in assistive technologies.

BIBLIOGRAPHY:

1. "Open your eyes: Deaf studies talking," Dirksen Bauman (2008). This publication is available through the University of Minnesota Press and has an ISBN of 978-0-8166-46197
2. Kim Nielsen (2012) published "A Disability History of the United States," a book that examines the history of disability in America. The book is available through Beacon Press and has an ISBN of 978-080702204-7.
3. "Reduction de las letras y Arte para enseñar á ablar los Mudos" this book was written by J. de Pablo Bonet in 1620. This book is regarded as the first modern treatise on sign language phonetics and features a manual alphabet for the hearing-impaired. A facsimile of the original book is available online (in Spanish) through the “Miguel de Cervantes" Virtual Library.
4. In his work on cryptography titled "Mercury, the Swift and Silent Messenger" (1641), John Wilkins discussed fingerspelling as a method of secret communication using sign language. Wilkins also provided an example of a fingerspelling system.
   "Chirologia: or the natural language of the hand," a book by John Bulwer, was published in London

in 1644. Bulwer mentions in this book that deaf individuals use alphabets, although he presents a different system that focuses on public speaking.

5. Jaffe, DL (August 1994). "Evolution of mechanical fingerspelling hands for people who are deaf-blind". This paper was published in the Journal of Rehabilitation Research and Development. 31 (3): 236–244. PMID 7965881.
6. "Sign Language Recognition and Translation: A Multidisciplined Approach from the Field of Artificial Intelligence". This paper was written by Parton, B. S. (12 October 2005) and was published in the Journal of "Deaf Studies and Deaf Education". 11 (1): 94–101. doi:10.1093/deafed/enj003. PMID 16192405.
7. Boris Mocialov, Helen Hastie, Graham Turner, Katrin Lohan (2017). "Towards Continuous Recognition of Sign Language Using Deep Learning" (PDF). Robot Assistants and Meaning: The Gap Smart Devices Leave" (IEEE-RAS International Conference on Humanoid Robots). S2CID 5525834.
8. Moryossef, A., Tsochantaridis, I., Aharoni, R., Ebling, S. and Narayanan, S., (2020), August. Real-time sign language detection using human pose estimation. In the European Conference on Computer Vision (pp. 237-248). Springer, Cham.
9. S. Dinesh, S. Sivaprakash, M. Keshav, K. Ramya (2018). Real-Time American Sign Language Recognition with Faster Regional Convolutional Neural Networks.
10. Bauer, B., & Hienz, H. (2000). Relevant features for video-based continuous sign language recognition. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (pp. 440-445). doi: 10.1109/AFGR.2000.840672. Real Time Sign Language Translator for Video Conferencing Platforms https://iaeme.com/Home/journal/IJAIML 22 editor@iaeme.com
11. G R Meghana, Harshitha H S, L Shyla & Madan Kumar S M (2020). Glove-Based Sign Language Interpreter for Deaf and Aphonic Peoples.
12. Sujay R, Somashekar M & Aruna Rao B P (2022). Sign Language Translator.