Sign Language Recognition System

Chaithrashree. A [1], Mritunjay Shukla [2], Rohit Kumar Thakur [3], Shagun Dhurandhar [4], Sudipto Chakraborty [5]

[1] Assistant Professor, Department of CSE, Brindavan College of Engineering, Bengaluru, India
[2345] B.E. Student, Department of CSE, Brindavan College of Engineering, Bengaluru, India

---***---

Abstract - *Sign language is used by the deaf community to communicate with non-deaf individuals. The gestures used in sign language can be difficult for the general population to understand. However, this sign language can be converted into a form that is easy for the general public to understand. This research is based on various image and video capturing techniques, preprocessing, classifying of hand gestures and landmarks extraction, and classification techniques. To identify the most promising methods for future research, this paper examines the techniques used to generate datasets, sign language recognition systems, and classification algorithms. Many of the currently available studies contribute to classification approaches in combination with deep learning due to the growth of classification methods. This paper focuses on the methods and techniques used or applied in earlier years.*

***Key Words*: SLRS- Sign language Recognition System, CNN- Convolutional Neural Networks, LSTM- Long Short-Term Memory, SL - Sign Language, NN – Neural Network, LCBCr- Luminance, Chrominance Blue, Chrominance Redcolor.**

1. INTRODUCTION

The goal of a sign language recognition system is to identify and translate hand gestures, body language, and other non-verbal cues utilized in sign language. This paper discusses a wide range of algorithms and techniques that can be used to collect, process, and comprehend hand gestures and sign language used by the deaf. It is claimed that the system for hand gesture recognition is a method or a strategy for productive human-computer interaction. Sign languages are those that express meaning through the visual-manual modality [1]. Static sign identification from photos or videos recorded in somewhat controlled environments has been the focus of a lot of recent research.

The foundation of computer vision techniques is the way humans interpret information about their surroundings. Since sign language movements might vary, it can be challenging to create a vision-based interface that is understood by people worldwide. However, developing such an interface for a specific set of people or nations is conceivable. The detection of hand gestures and their motion when in motion depends on area selection since hand motions have distinct shape variations, textures, and movements. It is easy to identify a static hand, skin tone, and hand shape using characteristics such as finger directions, fingertips, and hand posture. The picture's background and lighting make these aspects not always reliable and accessible. It is hard to express features exactly as they are, hence the entire video frame or image is used as the input. The goal of this work is to examine and evaluate the methodologies employed in previous studies and approaches. This report also aims to determine and recommend the optimal research direction for future research.

According to Prof. T. Hikmet Karakoc, SL is comparable to the word language in that both are widely used around the world [2]. Since sign language has evolved, its vocabulary and syntax have been recognized as real languages. The reason sign language (SL) is the language of choice for the deaf is that it can be constructed by combining facial emotions, hand gestures and movements, and palm movements to convey a speaker's thoughts without the need for a voice.

Deep learning can be very effective in sign language recognition and translation tasks. Sign translation is the process of converting signed language into spoken language or text, whereas sign recognition is the analysis and interpretation of sign language gestures. Deep Learning methodologies that can be used to train models that can understand and translate sign language include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks. One popular method is using a camera to record the gestures and then applying computer vision techniques to process the video frames to extract characteristics that can be input into a deep learning model for translation or classification.

Because deep learning can automatically discover intricate patterns and relationships in large datasets, it is a good fit for sign language recognition applications. Sign language is a visual language that requires intricate hand, facial, and body motions. By examining a vast quantity of sign language data, including videos of people speaking the language, deep learning models may be trained to identify these gestures and patterns. This process allows the models to automatically identify the key elements and patterns needed for precise sign language identification.

A well-liked machine learning method for sign language recognition systems is neural networks (NN). An algorithm known as a neural network (NN) is made to identify intricate relationships and patterns in data. Neural networks' capacity to discover intricate connections between input data and predicted outputs is what gives them their effectiveness. Iteratively fine-tuning the weights and biases of the connections among the nodes in the hidden layers allows the model to become more accurate and more capable of generalizing to new data. SLRS uses a dataset of pre-segmented sign language gestures and signs to train neural networks.

CNNs are used to recognize images, including those in sign language. They work very well at distinguishing spatial characteristics in photos, which helps recognize hand motions and movements. CNNs have been used in several research to categorize individual signs or complete sentences in sign language.

Sequential data modeling is a key function of RNNs in sign language recognition, given that sequential language includes sign language. RNNs can identify temporal dependencies. through the sequential processing of the signs in sign language. RNNs have been used in certain studies to identify words and sentences in sign language (source: Starostenko et al., 2018; Iwana et al., 2020).

Long Short-Term Memory Networks are a kind of recurrent neural network that can manage the vanishing gradient problem that might occur while training RNNs on long sequences. Long-term dependencies in sign language sequences have been found using LSTMs in sign language recognition. LSTMs have been used in certain research to identify continuous sign language utterances (source: Cui and Co., 2019; Chen et al., 2021).

CRNNs combine the sequential modeling of RNNs with the spatial feature detection of CNNs.

They have been applied to the recognition of sign language to record the temporal and geographical information included in sign language videos. CRNNs have been utilized in certain research to identify continuous sign language words and gestures (source: Zhou et al., 2018; Camgoz et al., 2018).

Recording the signs is necessary for the SLRS to function. To collect these indications as input, a variety of strategies and tactics are employed. Among these techniques is the use of Microsoft Kinect sensors to collect multimodal data. Microsoft created these Microsoft Kinect sensors, a kind of motion-sensing input device. The Kinect sensor tracks human motions

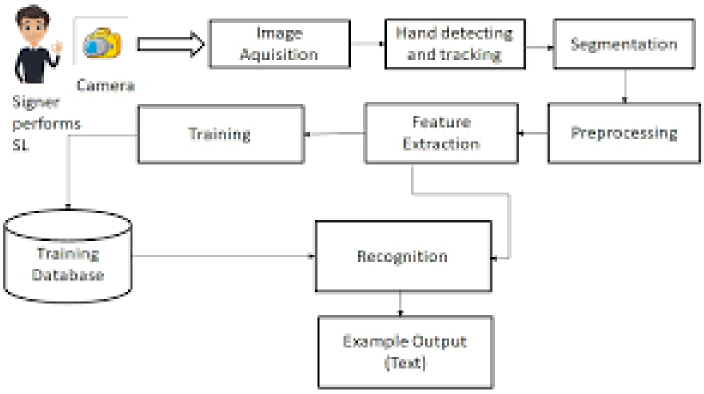

Fig 1: Block Diagram for Sign Language Recognition System

2. LITERATURE SURVEY

Year: Pre-Computer Era

Authors: William Stokoe

Title: Manual Coding System.

Methodology: Stokoe's manual coding systems aimed to represent sign languages linguistically by using symbols to denote handshapes, movements, and locations, providing a structured notation system.

Advantages: It provides a foundational understanding of linguistic structures in sign languages, facilitating cross-cultural comparisons and contributing to broader linguistic studies.

Disadvantage: Limited set of symbols, which couldn't capture the richness and nuances of sign languages comprehensively, potentially leading to oversimplification or loss of linguistic detail.

Year : 1980s-1990

Authors: Ralph K. Potter

Title: Computer Vision Technologies

Methodology: Potter's approach emphasizes leveraging computer vision for sign language recognition, using algorithms to interpret visual data of hand gestures, movements, and facial expressions.

Advantages: Focus on computer vision for sign language recognition is the potential to enhance accessibility by enabling real-time interpretation, aiding in bridging communication gaps between the Deaf and hearing communities.

Disadvantage: Computer vision systems might struggle with recognizing nuances and variations in sign language gestures due to the intricate nature of hand movements and facial expressions.

Year : 1990s-2000

Authors: Takeo Kanade

Title: Glove-Based System

Methodology: It involves developing sensor-equipped gloves to capture hand movements and gestures, and translating these inputs into recognizable sign language through computational analysis.

Advantages: Allows for a more natural and intuitive interaction by directly capturing hand movements, enhancing communication fluidity for sign language users.

Disadvantage: The system might face constraints in accurately capturing the full range of sign language gestures and nuances, potentially leading to incomplete or misunderstood interpretations due to the limited sensors' capabilities.

Year : 2000s-2010

Authors: Raja Kushalnagar

Title: Video-Based Recognition

Methodology: It involves employing advanced video analysis algorithms to interpret sign language gestures, leveraging machine learning and computer vision techniques to process visual data from videos.

Advantages: Video-based recognition allows for a more comprehensive analysis of sign language, capturing subtle nuances in hand movements, facial expressions, and body language, enhancing accuracy in interpretation.

Disadvantage: Processing video data in real-time for sign language recognition can be computationally demanding, requiring significant resources, and potentially limiting its practicality in some settings due to computational constraints.

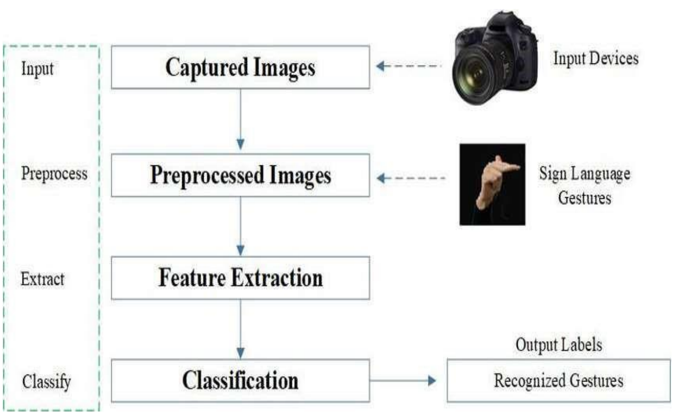

Fig 2: System Architecture

Feature	Description
Hand Shape	Describes the shape of the hand during the signing
Hand Movement	Indicates the movement of the hand during the signing
Hand Orientation	Specifies the orientation of the hand relative to the signer or viewer
Hand Configuration	Describe the arrangement of fingers and thumb during the signing

Table 1: Feature of Sign Language Recognition System

3. METHODS

Step 1: Data Collection Strategy

Design a comprehensive plan to collect diverse sign language data, considering regional variations, gestures, and expressions. Utilize multiple sources such as videos, motion capture, and crowd-sourcing to ensure a wide range of signs.

Step 2: Annotation and Labeling

Develop robust annotation protocols to label collected data accurately. Include information on hand gestures, facial expressions, body movements, and contextual cues to enrich the dataset for better recognition.

Step 3: Data Augmentation Technique

Implement techniques such as mirroring, rotation, and translation to increase the diversity and volume of the dataset. Synthetic data generation and augmentation can help address limitations due to limited original data.

Step 4: Preprocessing and Feature Extraction

Utilize image and signal processing techniques to preprocess the raw data. Extract relevant features like hand trajectories, key points, and facial expressions.

Apply dimensionality reduction and normalization to enhance model performance.

Step 5: Model Development and Training

Employ deep learning architectures like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) for sign language recognition. Train models using the annotated dataset and fine-tune them to capture intricate sign variations.

Step 6: Continuous Evaluation and Improvement

Regularly assess model performance using various metrics, including accuracy, precision, and recall. Incorporate user feedback and adjust the model to address errors and improve recognition capabilities continually.

Step 7: Integration with Assistive Technologies

Integrate the sign language recognition system into assistive technologies such as smart glasses, mobile applications, or communication devices. Ensure compatibility and usability for real-time.

4. RESULT AND DISCUSSION

An overview of earlier surveys on the techniques applied in different gesture and sign language recognition studies is given in this section. Both standard cameras and Kinect cameras are widely used to capture data in sign language. Data collection and submission to the SLR system are crucial because, once data is gathered, pre-processing is needed to guarantee correctness.

Gaussian filters remove noise from the input data. It is easy to discern the backdrop color and skin tone from the RGB color spaces. The review demonstrates how the division outcome is built by breaking up skin tone with additional limits such as edge and recognizable proof.

The vision-based approach extracts elements from images using gesture categorization. The two most used classification algorithms are SVM and ANN. When compared to ANN, SVM output was superior. Most models employ sensors to gather information about their surroundings when examining the data at hand. In vision-based applications, brain networks are typically used to process images and videos; in order, Gee and SVMs are used. This is a result of the increased availability of data sources. The classification of motions, which quickly pulls information from images, is the last step in the vision-based technique. ANN and SVM are the two most used methods for data classification. When analyzing the given data, most models get environmental information from sensors. In vision-based methods for images and videos, neural networks are commonly used, whereas HMMs and SVMs are used for classification. This is a result of the increased accessibility of data sources.

Neural network models are another key processing approach for deciphering sign language The CNN processing image passes through convolution, pooling, activation functions, and fully linked layers to understand sign language. To extract motion information from depth changes in frames and features, 3D-CNNs are implemented. LSTM-based algorithms can be utilized to extract simulation temporal sequence data from videos of sign language [13]. A step ahead in LSTM, BLSTM-3D ResNet [14], can restrict hands and palms from video progressions. HMM and SVM are two crucial classification methods in vision-based systems. In recent studies, convolutional neural networks have become more and more prominent in vision-based sign language recognition. Pre-trained models can be used right away without requiring any data to be provided or training. Because of the tremendous potential of pre-prepared models, scientists have recently given them a great deal of thought. Additionally, it lowers the cost of building and training datasets. Many analysts have used pre-prepared models like signal-based VGG16, Google Net, and AlexNet to reduce the preparation costs. To lower loss errors when training datasets, cross-entropy, and other loss functions are used. The cross-entropy multiclass category is the most widely used loss function. The widely used SGD, ADAM, and analyzer enhancers need to support the unfortunate capacity.

Because it can learn and associate on its own, the deep neural network (DNN) yields higher results but requires the greatest training sample. As of right now, it can register to execute programs on large datasets. Better program execution is taken into account while performing new computations and making improvements to existing ones.

5. CONCLUSION

In conclusion, A system that was designed to recognize alphabets and static gestures has developed into one that can also recognize dynamic motions that are in motion.

In published research, results from vision-based approaches are frequently better than those from static-based approaches. The development of large vocabularies for sign language recognition systems is currently attracting more study interest. Improvements in computer speed and the availability of datasets make it possible to access additional training for certain samples.

A lot of people are building their datasets to aid in the improvement of their recognition of sign language. The grammar and presentation of each phrase determine the sort of sign language employed in various nations.

Deep learning techniques like CNN and LSTM Models can identify the sequence of photos and video streams with a high degree of accuracy. I find that when working with video frames, LSTM models get more accurate results. They are quite good at extracting features and capturing 3D gestures. Given that LSTM can be used for videos and CNN can be used for static images, merging the two can result in a sign language recognition system that is both more potent and accurate.

6. ACKNOWLEDGEMENT

The project team would like to express their heartfelt gratitude to our alma mater Brindavan College of Engineering for giving us the opportunity and to our project mentor Prof. Chaithrashree A for her guidance, encouragement, and input which led us to undertake this project.

7. REFERENCES

[1] Sandler and Lillo-Martin's book explores the universality of language through the study of sign languages.

[2] Huang et al.'s article provides a review of the synthesis, properties, and applications of carbon dots.

[3] Khoshelham and Elberink's article examines computer vision techniques used with RGB-D sensors such as the Kinect and their applications.

[4] Morales-Sánchez et al.'s article provides an overview of the innate and adaptive immune responses to SARS-CoV-2 infection in humans.

[5] Rujana's American Sign Language Image Dataset is a collection of images used for research and training in sign language recognition.

[6] Guan et al.'s article presents an IoT-based system for energy-efficient regulation of indoor environments in buildings.

[7] Hassan et al.'s review article provides an overview of blockchain technology, including its architecture, consensus, and future trends.

[8] Makridakis et al.'s article reviews recent research on time series forecasting.

[9] Wang et al.'s article provides a comprehensive review of blockchain technology, from fundamentals to applications.

[10] Gao et al.'s article provides a review of deep learning-based image analysis techniques for agricultural applications.

[11] Zhang et al.'s article reviews recent advances in precision irrigation technologies based on IoT.

[12] Ji et al.'s survey article examines edge computing for IoT.

[13] Alsheikh et al.'s survey article provides an overview of machine learning techniques for resource management in edge computing.

[14] Al-Fuqaha et al.'s article is a survey of enabling technologies, protocols, and applications for IoT.