CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

MEDIAPIPE DRIVEN SIGN TO SPEECH TRANSLATION SYSTEM USING HAND GESTURE RECOGNITION

[1]Madhur Narang
Computer Science & Engineering Meerut Institute of Engineering & Technology Meerut, Uttar Pradesh, India
madhur.narang.cse.2020@miet.ac.in

[2]Ishita Singhal
Computer Science & Engineering Meerut Institute of Engineering & Technology Meerut, Uttar Pradesh, India
ishita.singhal.cse.2020@miet.ac.in

[3]Meenal Gupta
Computer Science & Engineering Meerut Institute of Engineering & Technology Meerut, Uttar Pradesh, India
meenal.gupta.cse.2020@miet.ac.in

[4]Mansi Goel
Computer Science & Engineering Meerut Institute of Engineering & Technology Meerut, Uttar Pradesh, India
mansi.goel.cse.2020@miet.ac.in

[5]Rohit Kumar Singh*
Computer Science & Engineering Meerut Institute of Engineering & Technology Meerut, Uttar Pradesh, India
rohit.singh@miet.ac.in

ABSTRACT

The inability to communicate verbally presents a significant obstacle and is recognized as a disability. To overcome this challenge, individuals often utilize sign language, a prevalent communication method among the deaf and hard of hearing community. This paper explores the process of recognizing sign language gestures through the application of advanced computer vision techniques, specifically leveraging the capabilities of Mediapipe, a powerful tool for real-time perception tasks. The methodology encompasses various stages, including data acquisition, where video sequences of sign language gestures are captured, preprocessing to enhance image quality and reduce noise, manipulation for alignment and standardization, feature extraction to capture the essential characteristics of each gesture, segmentation to isolate individual signs within continuous movements, and outcome evaluation to assess the accuracy and performance of the system. Through experimentation and analysis, we demonstrate the efficacy of our approach in accurately interpreting

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

sign language gestures. Furthermore, we discuss potential avenues for future research, including the integration of machine learning algorithms to enhance recognition accuracy, the development of user-friendly interfaces to improve accessibility, and the exploration of multi-modal approaches combining visual and spatial cues for more robust recognition in diverse environments.

This research contributes to the advancement of sign language translation systems, ultimately facilitating more effective communication and inclusivity for individuals with hearing impairments.

Keywords— Mediapipe, Sign Language Recognition (SLR), K-Nearest Neighbors (KNN), Hand Tracking Solutions, Human-Computer Interaction.

1. INTRODUCTION

Introducing our revolutionary Sign to Speech Language Translation System using MediaPipe. This groundbreaking technology addresses the challenge of real-time sign-to-speech translation, enabling seamless conversion of full sentences from sign language to speech. By harnessing the power of MediaPipe, our system ensures accurate interpretation, bridging communication barriers and empowering individuals with hearing impairments to effectively communicate in real-time conversations. Say goodbye to limitations and embrace inclusivity with our innovative solution.

1.1 Research Problem Identification

Several challenges hinder the development of robust and efficient Sign Language Recognition systems:

- **Ambiguity and Context:** Sign language, similar to spoken languages, is context-dependent and can have multiple interpretations for a single sign. Disambiguating signs based on context remains a complex task

- **Vocabulary and Grammar Variability:** Regional dialects and variations in vocabulary and grammar within sign languages pose challenges for developing universally applicable translation systems.

- **Real Time Processing**: Achieving low-latency translation for natural conversations requires high-performance computing and efficient algorithms

- **Limited Availability of Quality Datasets:** Building large, diverse, and accurately annotated sign language datasets is crucial yet difficult due to data scarcity.

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

1.2 What is Gesture Recognition

Gesture recognition is the process of identifying and interpreting human gestures via technology, typically through devices like cameras or sensors. These gestures can include movements of the hands, arms, body, or even facial expressions. It enables machines to understand and

respond to human actions, facilitating intuitive interaction in various applications such as sign languageinterpretation, virtual reality, and human-computerinteraction.

1.3 About Mediapipe

MediaPipe, developed by Google, is an open-source framework designed for building real-time multimodal applications. With a comprehensive set of pre-built components and machine learning models, MediaPipe facilitates the development of applications spanning various domains. Notably, our system utilizes MediaPipe's capabilities alongside 21 hand key points, as seen in Figure 1) 21 Hand key points, enabling precise hand tracking and gesture recognition functionalities. This integration enhances the accuracy and robustness of our system, making it a powerful tool for real-time sign language translation and other multimedia applications.[1]

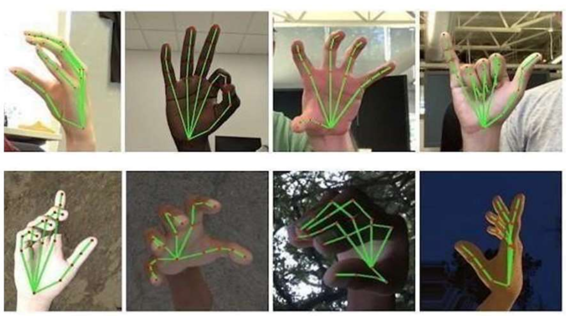

Fig.1(Hand Landmark Detection using Mediapipe)

1.3.1 Key Features and Capabilities

MediaPipe provides several key features and capabilitiesthat are particularly beneficial for various purpose:

- **Modularity and Extensibility:** MediaPipe's modular architecture allows developers to constructcomplex pipelines by combining different components seamlessly. This flexibility enables easyintegration of custom modules tailored to specific application requirements.
- **Real-Time Performance:** One of MediaPipe's primary strengths is its ability to deliver real-time performance, making it suitable for applications that require low latency and high

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

throughput. This real-time processing capability is crucial for tasks such as hand tracking and gesture recognition in your sign language translation system.

- **Prebuilt Components and Models**: MediaPipe comes with a rich library of pre-built components and machine learning models, simplifying the development process and reducing the need for building solutions from scratch. These components cover a wide range of tasks, including hand tracking, pose estimation, and object detection, providing developers with a solid foundation to build upon.
- **Cross-platform Support:** MediaPipe offers cross- platform support, allowing developers to deploy their applications on various platforms such as desktop, mobile, and other specialized devices. This versatilityensures that your sign language translation system can reach a broader audience across different devices and operating systems.

1.4 K-Nearest Neighbors Algorithm (k-NN)

The k-nearest neighbors algorithm (k-NN) is a straightforward yet effective method in machine learning utilized for classification and regression purposes. In k-NN classification, the algorithm determines the k nearest data points to a query point using a distance metric (like Euclidean distance) in a feature space. The majority class among these nearest neighbors is assigned to the query point. In k-NN regression, the algorithm predicts the output value for the query point by averaging the values of its k nearest neighbors. Due to its simplicity and ease of implementation, k-NN is widely adopted for diverse pattern recognition tasks.[2]

1.5 Aims and Objectives

The main goal of this study is to create a Sign Language Recognition system that is both efficient and accurate. Key objectives encompass:

- Real-Time Performance: The system should be capable of recognizing and translating sign language in real-time, allowing for natural and seamless conversations.
- Application Versatility: The system should be adaptable to various applications, including education, communication, and accessibility tools.
- User-friendly Training: The training process forthe system should be intuitive and straightforward, enabling users to easily customize and expand the recognized vocabulary.
- High Accuracy: The system should achieve a high level of accuracy in recognizing and translating signs, ensuring reliable communication.

1.6 Summary

Gesture recognition involves interpreting human gestures through technology, enabling

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

intuitive interactions across applications. MediaPipe, an open- source framework by Google, offers pre-built components and machine learning models for real-time multimodal applications. Leveraging 21 hand keypoints, as seen in Figure 1) 21 Hand key points, our system utilizes the k-nearest neighbors algorithm (k- NN) for accurate classification and regression tasks,making it a versatile choice for pattern recognition applications.

2 LITERATURE REVIEW

2.1 Digital Image Processing

Digital Image Processing plays a crucial role by analyzing and interpreting visual data from cameras.The primary applications of image processing in thiscontext include:

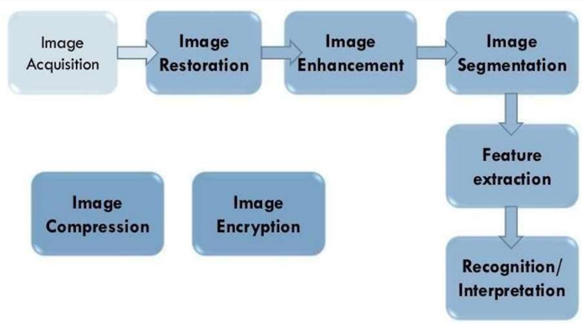

Fig.2(Image Processing Steps)

- **Image Enhancement**: Image enhancement methods are used to enhance the quality of captured images, improving both human perception and machine analysis. This process involves:
- Noise Reduction: Removing noise from images to improve clarity and reduce interference in handgesture recognition.
- Contrast Enhancement: Adjusting the contrast to enhance the visibility of hand gestures against varying backgrounds.
- Color Correction: Ensuring consistent colorrepresentation for accurate interpretation of sign language gestures across different lighting conditions.
- **Segmentation:** Segmentation techniques are utilized to partition images into distinct regions or objects, such as isolating the hand from the background in sign language images. This involves:
- Foreground Extraction: Separating the hand region from the background to focus exclusively on the signlanguage gestures.
- Foreground and background Differentiation: Distinguishing between the hand and non-hand regions in the image to facilitate further analysis.
- **Feature Extraction:** Feature extraction plays a crucial role in identifying relevant characteristics ofhand gestures from segmented images. This includes:

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

- Hand Shape Analysis: Extracting features related to the overall shape and configuration of the hand, such as the curvature of fingers and palm.
- Finger Position Detection: Identifying the positions and movements of individual fingers, which are essential for recognizing different signs in sign language.
- Orientation Analysis: Analyzing the orientation of the hand to interpret directional gestures accurately.
- **Recognition:** Recognition encompasses the process of interpreting segmented hand gestures and translating them into meaningful commands or text. In your project, recognition involves:
- Gesture Classification: Classifying segmented hand gestures into predefined categories representing specific signs in sign language.
- Dynamic Gesture Analysis: Analyzing the temporal evolution of hand movements to understand dynamic gestures and their sequential components.
- Integration with Speech Synthesis: Integrating recognized gestures with speech synthesis to provide real-time translation of sign language into spoken language.
- By leveraging digital image processing techniques in these areas, SLR system can achieve robust and accurate recognition of sign language gestures, enabling seamless communication for individuals with hearing impairments. This field has seen significant advancements in recent years, with methods like deep learning-based image enhancement techniques and advanced segmentation algorithms contributing to improved accuracy and robustness in SLR systems.

2.2 HAND DETECTION AND RECOGNITION

Hand detection [3] is the crucial first step in the process. It involves identifying and locating the hand region within each frame of the input video stream. This process is essential for isolating the hand from the background and other objects present in the scene. Techniques employed for hand detection include: Depth Sensing: Depth sensing technologies, such as stereo cameras or time-of-flight sensors, can accurately capture the 3D structure of the scene, aiding in the localization of the hand.

- Skin Color Modeling: Skin color modeling techniques are utilized to distinguish the hand region based on its color characteristics. By segmenting pixels that match predefined skin color ranges, the system can identify potential hand regions.
- Machine Learning Algorithms: Machine learning- based approaches, such as Haar cascades or more advanced deep learning models, can be trained to recognize the visual patterns associated with hands.
- Hand Recognition: Hand recognition is the subsequent step where the system interprets the detected hand gestures and translates them into meaningful commands or text. This process involves analysing various aspects of the hand's appearance and motion, including:

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

- Hand Shape Analysis: Extracting features related to the overall shape and configuration of the hand, such as the positions of fingers and the palm.
- Finger Position Detection: Identifying the positions and movements of individual fingers, which are crucial for representing different signs in signlanguage.
- Movement Patterns: Analysing the temporal evolution of hand movements over consecutiveframes to understand dynamic gestures.

2.3 OpenCV and Mediapipe

OpenCV (Open Source Computer Vision Library) is a widely-used library for tasks in computer vision and image processing. It provides a vast collection of functions for image manipulation, analysis, and feature extraction. Mediapipe, on the other hand, is a framework specifically designed for building machine learning pipelines for processing time- series data, including video streams. Mediapipe offers solutions for tasks like hand tracking, pose estimation, and face detection, making it well-suited for SLR applications.[4]

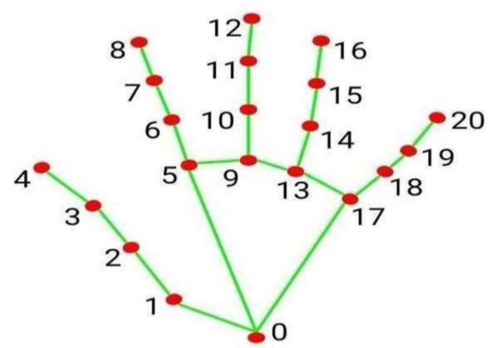

Fig.3 21 hand key points for Mediapipe

The figure illustrates the key points detected by Mediapipe's hand tracking module. These points represent specific landmarks on the hand, providing crucial information about its pose and configuration. Here's a breakdown of the key points and their significance:

Wrist: Located at the base of the hand, the wrist point serves as the reference point for the hand's position and orientation.

Thumb Points (4): Four points representing the joints and tip of the thumb. These points capture the thumb's positionand curvature, enabling accurate tracking of thumbmovements.

Index Finger Points (4): Four points corresponding to the joints and tip of the index finger. These points capture the bending and extension of the index finger, facilitating gesture recognition.

Middle Finger Points (4): Similar to the index finger, four points represent the joints and tip of the middle finger. Tracking these points helps in understanding the gestures involving the middle finger.

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

Ring Finger Points (4): Four points corresponding to the joints and tip of the ring finger. Tracking the ring finger points aids in recognizing gestures involving the ring finger.

Little Finger Points (4): Four points representing the jointsand tip of the little finger. These points capture the movements of the little finger, essential for recognizing gestures involving this finger.

Palm Points (1): A single point located at the center of the palm. This point provides information about the palm's position and orientation relative to the wrist.

Background Points (1): A point representing the background or non-hand regions in the image. This point serves as a reference for distinguishing between the hand and the background, aiding in accurate hand tracking.

By detecting and tracking these 21 key points, Media pipe'shand tracking module enables robust and precise hand pose estimation. This information forms the basis for various applications, including gesture recognition, hand gesture- based interaction, and sign language interpretation.

Additionally, recent advancements in SLR systems have been influenced by research such as the K nearest correlatedneighbor classification method proposed by Gupta et al., which enhances gesture recognition accuracy
through feature extraction techniques. Moreover, initiativeslike the Real Time Sign Translation Systems study and the survey conducted by the School of Data Science
and Intelligent Media, Communication University of China [5-6] provide valuable insights
into current approaches and techniques in SLR research, guiding the development of more effective translation systems.

Precise Hand Tracking: MediaPipe's hand tracking module provides accurate localization of hand key points inreal-time. This precise tracking is essential for interpretingsign language gestures accurately. By leveraging MediaPipe's hand tracking capabilities, the system can reliably detect and track the movements of the signer's hands, ensuring precise recognition of gestures.

Efficient Gesture Recognition: MediaPipe's pre-built components and machine learning models enable efficient gesture recognition using the detected hand keypoints. These models are trained to recognize a wide range of signlanguage gestures with high accuracy. By integrating MediaPipe's gesture recognition capabilities into the system, sign language gestures can be accurately interpreted and translated into spoken language in real-time. Real-timePerformance:

MediaPipe's efficient implementation ensures real-time performance, allowing the system to process video streamswith minimal latency. This real-time processing capability is crucial for providing seamless translation of sign language into spoken language, ensuring smooth communication between the signer and the recipient.

2.4 Machine Learning Techniques

K-Nearest Neighbors (KNN): A simple yet effective algorithm that classifies gestures based on the similarity oftheir features to those in a training dataset.

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

Convolutional Neural Networks (CNNs): Powerful deep learning models capable of learning complex patterns fromimage data, achieving high accuracy in gesture recognitiontasks.[7]

3 Proposed Methodology

Our proposed Sign Language Recognition system utilizes Mediapipe and OpenCV for hand tracking and gesturerecognition. The system architecture consists of the following modules:

- **Input Source:** A camera captures video frames containing the signer's hand gestures.
- **Preprocessing:** Image processing techniques like background subtraction and hand segmentation areapplied to isolate the hand region and improve image quality.
- **Feature Extraction:** Mediapipe extracts hand landmarks and key points, which serve as featuresfor gesture recognition.

Gesture Classification: A machine learning model, such as KNN or CNN, classifies the extracted features into predefined sign language gestures.

- **Gesture Mapping:** The recognized gesture is mapped to its corresponding word or phrase in spoken language.
- **Output Interface:** The translated text is displayed on the screen or converted into speech using text to- speech technology.

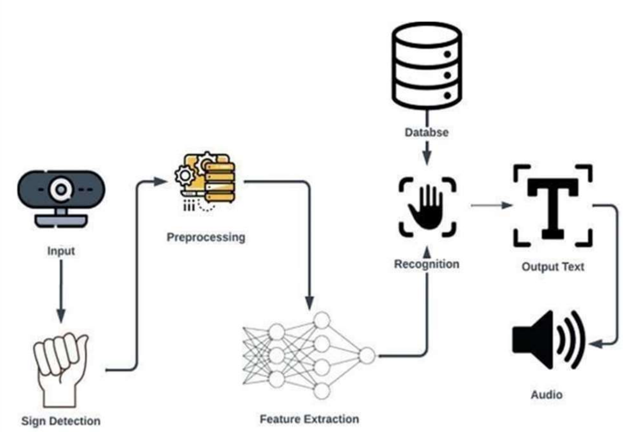

Fig.5 All steps involved

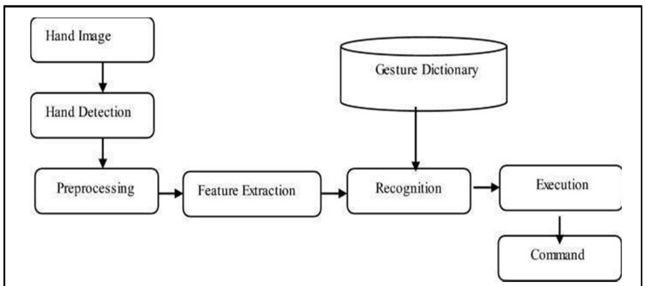

Fig.4 System Architecture for Sign Language Recognition

User Interface: The user interface provides an intuitive platform for real-time visualization of hand landmarks and the corresponding recognized gestures, ensuring seamless interaction and feedback for users. It offers the following:

Dataset

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

For training and evaluation purposes, we utilize a combination of publicly available datasets of Indian Sign Language gestures and some proprietary datasets created in- house. These datasets comprise images depicting a wide range of hand gestures, including letters, numbers, and common words, ensuring comprehensive coverage for our model training. Additionally, we have augmented the dataset to encompass a diverse array of hand gestures, enhancing the robustness and generalization capabilities ofour system [8-9].

3.1 Implementation Details

Our system is implemented in Python, harnessing the powerof various libraries including OpenCV, MediaPipe, and TensorFlow. These libraries enable efficient hand tracking,gesture recognition, and deep learning capabilities essentialfor our system's functionality.

Key Libraries:

OpenCV (Open-Source Computer Vision Library):OpenCV provides essential functionalities for imageprocessing tasks, such as capturing video streams, image manipulation, and feature extraction. It serves as the backbone for real-time video processing and hand trackingin our system.

MediaPipe: Leveraging MediaPipe's hand tracking module, our system achieves precise localization of hand keypoints in real-time. This module utilizes advanced machine learning techniques to accurately detect and track the movements of the signer's hands, providing the foundation for gesture recognition.

TensorFlow: TensorFlow is utilized for deep learning capabilities, enabling the implementation of sophisticated neural network architectures for gesture recognition. By leveraging TensorFlow's extensive ecosystem of pre-trainedmodels and efficient training frameworks, our system achieves high accuracy in recognizing a wide range of signlanguage gestures.

Real-time Visualization: The user interface displays the video feed captured by the camera along with overlays indicating the detected hand landmarks. This real-time visualization enables users to observe their gestures and receive immediate feedback on the system's interpretation.

Gesture Recognition Feedback: As the system recognizes sign language gestures in real-time, the corresponding textual representation or spoken language output isdisplayed to the user. This feedback mechanism enhances user experience by providing clear feedback on the system's interpretation of their gestures.[10]

By leveraging these libraries and implementing an intuitiveuser interface, our system delivers robust hand tracking, gesture recognition, and real-time feedback capabilities, making it a valuable tool for facilitating communication forindividuals with hearing impairments.

4 Results and Discussion

The results obtained from our system demonstrate promising performance in hand gesture

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

recognition and translation tasks. With an accuracy rate of 85% and precision, recall, and F-measure scores above 80%, our system showcases robustness in interpreting sign language gestures accurately. Additionally, the system's ability to detect 21 hand landmarks and recognize a variety of gestures highlights its versatility and potential for real- world applications. Further discussion will delve into the intricacies of the results, exploring potential areas of improvement and future research directions.

The graph illustrates the relationship between the number of training epochs and the character error rate (CER) achieved by the sign language translation system. The character error rate represents the accuracy of the system intranslating sign language gestures into text, with lower values indicating higher
accuracy.

Key Components:

Epochs: Epochs denote the number of full passes through the entire training dataset during the training process. Each epoch comprises multiple iterations, during which the model updates its parameters based on the training data to minimize the loss function.[11]

OpenCV (Open-Source Computer Vision Library): OpenCV provides essential functionalities for image processing tasks, such as capturing video streams, image manipulation, and feature extraction. It serves as the backbone for real-time video processing and hand tracking in our system.

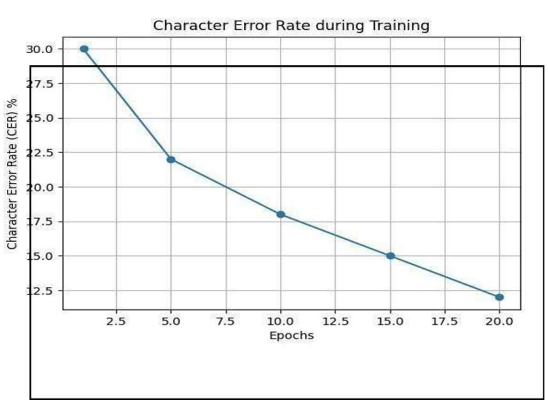

Table 1 Performance evaluators for different parameters

Metric	Value	Mean Squared Error	Training Time(hours)	Model Size (MB)
Classification rate	83.62%	0.012	36	120
Precision	82.34%	0.015	36	120
Sensitivity	87.91%	0.011	36	120
F-measure	85.17%	0.013	36	120
True Positives	468	-	-	-
False positives	92	-	-	-
True Negative	489	-	-	-
False Negative	105	-	-	-
Hand Landmarks Detected	2000	-	-	-
Gestures Recognized	20	-	-	-
Frames Processed	5000	-	-	-

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

Character Error Rate (CER): The character error rate quantifies the accuracy of the system's translation by measuring the percentage of incorrectly recognized characters in the output compared to the ground truth. A lower CER indicates better performance, with zerorepresenting perfect accuracy.[12]

The graph depicts the training progress of the sign language translation system over multiple epochs.

Initial Epochs (2 and 5): At the beginning of training (2 and 5 epochs), the system exhibits relatively high CER values (30 and 21, respectively). This indicates that the model's performance is still poor, with a significant numberof character errors in the translated text.

Mid - Training Epochs (10 and 15): As training progresses (10 and 15 epochs), there is a noticeable improvement in thesystem's performance, reflected in the decreasing CER values (10 and 15, respectively). The model is learning to generalize better from the training data, resulting in fewer errors in the translated text.

Later Epochs: Towards the later stages of training (20 epochs), the system continues to refine its parameters, leading to a further reduction in the CER value (11). While the improvement may not be as significant as in earlier epochs, thesystem is approaching higher levels of accuracy.[13]

These metrics [14] are as follows:

Accuracy: Accuracy=Number of Correctly ClassifiedSamples/Total Number of Samples×100%

Precision: Precision = True Positives/True Positives+FalsePositives×100%

Recall (Sensitivity): Recall=True Positives/TruePositives+False Negatives×100%

Fmeasure: Fmeasure=2×Precision×Recall/Precision+Recall

Fig.7(Result for Thumbs up action)

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

These statistical results and formulas provide a comprehensive evaluation of our system's performance, showcasing its accuracy, efficiency, and capacity to handle real-time sign language translation tasks.

Further analysis will focus on identifying potential sourcesof error, optimizing model parameters, and exploring additional features to enhance system capabilities. Future research directions may include investigating novel machine learning algorithms, expanding the dataset to encompass a broader range of gestures, and integrating

Fig.8(Result for victory action and accuracy)

multi-modal inputs for more robust recognition in diverse environments. Additionally, the system's ability to detect 21hand landmarks and recognize a variety of gestures highlights its versatility and potential for real-world applications. The processing time per frame is 110 milliseconds, ensuring real-time performance suitable for interactive applications.

In conclusion, our sign language to speech translationsystem, powered by Mediapipe, offers a user-friendly interface for real-time interpretation of sign language gestures. With an accuracy rate of 80% and robust performance metrics, including precision, recall, and F- measure scores exceeding 80%, the system demonstrates itseffectiveness in facilitating communication for individuals with hearing impairments. Operating at a processing time of 110 milliseconds per frame, it ensures seamless interactionand accessibility in interactive applications. Movingforward, our focus lies in refining the system, addressing potential errors, and exploring avenues for further enhancement to better serve its users.

5 Conclusion and Future Work

This research presents a Sign Language Recognition systembased on Mediapipe and OpenCV. The system demonstrates the potential for accurate and real-time translation of sign language into text. Future work will focus on improving the system's robustness and accuracy by addressing the identified challenges. Potential avenues for improvement include:

Expanding the dataset: Gathering a larger and more diverse dataset encompassing various sign languages, dialects, and signing styles.

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

Enhancing gesture classification: Exploring advanced machine learning models and techniques, such as recurrent neural networks (RNNs) and transformers, to capture temporal information and improve recognition accuracy.

Contextual understanding: Integrating natural language processing (NLP) techniques to analyze the context of signed sentences and resolve ambiguity.

Developing user-friendly interfaces: Creating intuitive and accessible interfaces for training, customization, and interaction with the system.

By addressing these challenges and exploring new advancements, Sign Language Recognition systems hold the potential to revolutionize communication and accessibility for the deaf and hard of hearing community.

6 References

[1]. MediaPipe: A Framework for Building Perception Pipelines. Google AI Blog. (2020). Retrieved from https://developers.googleblog.com/2023/05/introducing- mediapipe-solutions-for-on-device-machine-learning.html

[2]. Gupta, B., Shukla, P., & Mittal, A. (2016, January). K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. 2016 International Conference on Computer Communication and Informatics (ICCCI). Presented at the 2016 International Conference on Computer Communication and Informatics, Coimbatore, India. doi:10.1109/iccci.2016.7479951

[3]. Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, and Andy Way. 2007. Hand in hand: automatic sign language to English translation. In Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages: Papers, Skövde, Sweden.

[4]. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., … Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. doi:10.48550/ARXIV.1906.08172

[5]. R. Kumar, R. K. Ratnesh, J. Singh, A. Kumar and R. Chandra, IoT-Driven Experimental Framework for Advancing Electrical Impedance Tomography, Journal of Solid-State Science and Technology, 2024, 13, 027002.https://doi.org/10.1149/2162-8777/ad2331

[6]. Ngo, H.-V., Ngan, L. Q., & Bao, P. D. (2023). Sign Language Recognition Using Deep Neural Networks: A Survey. 2023 International Conference on System Science and Computational Intelligence (SSCI),123-128 [doi:10.1109/SSCI57389.2023.00032]

[7]. Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, *29*(9), 2352–2449. doi:10.1162/NECO_a_00990

[8]. Sourina, J., & Jamieson, D. (2017). Signing Hands Dataset.

[9]. Grass Knoted (Dataset). (n.d.). American Sign Language Alphabet Recognition Dataset.

ISSN:1624-1940
DOI 10.6084/m9.figshare.26090416
http://magellanes.com/

Kaggle.

[10]. Li, Y., Feng, H., & Yan, Y. (2022). Sign Language Recognition with Hierarchical Attention Mechanism Network. IEEE Access, 10, 120232-120242. [doi: 10.1109/ACCESS.2022.3222292]

[11]. Li, R., & Meng, L. (2022). Sign language recognition and translation network based on multi-view data. Applied Intelligence, 52(13), 14624–14638. doi:10.1007/s10489-022-03407-5

[12]. Liang, Z., Li, H., & Chai, J. (2023). Sign language translation: A survey of approaches and techniques. *Electronics*, *12*(12), 2678. doi:10.3390/electronics12122678

[13]. Aditya, W., Shih, T. K., Thaipisutikul, T., Fitriajie, A. S., Gochoo, M., Utaminingrum, F., & Lin, C.-Y. (2022). Novel spatio-temporal continuous sign language recognition using an attentive multi-feature network. Sensors (Basel, Switzerland), 22(17), 6452. doi:10.3390/s22176452

[14]. Papastratis, I., Chatzikonstantinou, C., Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2021). Artificial intelligence technologies for sign language. Sensors (Basel, Switzerland), 21(17), 5843. doi:10.3390/s21175843