|  |  |
| --- | --- |
|  |  |

**Real-Time Sign Language Interpreter using Mediapipe, Dynamic Time Warping and NLP**

Yash Mankar1, Vedanshu Thune1, Shrikant Khawshe1, Yogvid Wankhede1 and Madhuri Sahu1

1-5 Department of Artificial Intelligence, G.H. Raisoni College of Engineering, Nagpur, India

Email: {yashmankar1120, vedanshuthune, shrikantkhawshe727, yogvidwankhede, madhuridec9}@gmail.com

**Abstract.**

The "Sign Language Interpreter" project dedicates itself to helping individuals facing communication challenges by enabling those unable to express themselves through other means. This innovative study utilized hand gestures to generate sentences in time, thereby facilitating effective communication. Inspired by the language of signs our system employs algorithms for gesture recognition. It translates the nuances of hand movements into sentences ensuring accuracy and contextual understanding. The Sign Language Interpreter incorporates cutting-edge technologies to comprehend sign language intricacies.

**Index Terms*.*** Indian Sign-Language, Hand Gesture Recognition, Mediapipe, Dynamic Time Warping, Landmark Extraction, NLP

1. **Introduction**

The proposed study urges the implementation of a model that detects hand gestures to interpret conversations in the existing written and spoken languages. A real-time sign language interpreter using Mediapipe and Dynamic Time Warping is being developed. Sign language is a highly expressive and intricate form of communication that combines fluid hand movement, facial expression, head movement, and body language. Seamless integration of these elements makes it a complete and effective means of communication. However, this presents a formidable task in the field of computer vision, primarily because of the complexity of the data extraction and subsequent analysis. Furthermore, each individual possesses a unique body shape, volume, and distinctive gesture style, which complicate the recognition process. Moreover, the presence of uncertainties such as variations in viewpoint, illumination, shadows, self-occlusion, deformation, noise, and clothing further intensifies the complexity of this problem.

Our primary goal is to provide an accessible solution for individuals with speech impairments that enables effortless communication with the audience. In addition to converting sign language into spoken words, the proposed sign language interpreter offers features, such as an affordable and portable user interface, making it suitable for sign recognition and converting them to spoken words or phrases. The interpreter is designed to be user friendly so that individuals with proficiency levels can communicate seamlessly using sign language. This paper outlines the development process and underlying technologies used in creating the Sign Language Interpreter. The goal of this project is to contribute to technology that promotes better communication and inclusiveness, for people who have difficulty speaking by combining technical advancements such as Mediapipe, Dynamic Time Warping and Natural Language Processing.

At the center of our action lies the application of a comprehensive Indian Sign Language database as well as our custom dataset for training our model. This expansive database serves as the basis for developing our system. The addition of Indian Sign Language ensures that our model is culturally sensitive, understanding unique verbal nuances and expressions. Variance in the performance of human actions over time results in unique feature representations for each action in the different samples. This poses a challenge for widely used classifiers such as neural networks and support vector machines, which typically expect fixed-size input feature vectors and struggle with action-recognition tasks. This study introduced a classification model using Dynamic Time Warping (DTW) and a voting algorithm to address the challenge of unique feature representations in human actions. The system also uses natural language processing to convert text into speech in the spoken language, thereby enhancing its effectiveness. This groundbreaking design of assistive technology is a groundbreaking effort.

1. **Related Work**

## These existing systems aim to enable communication between individuals who generally use sign languages and common people. One of the main challenges in developing sign language recognition systems is the lack of sufficient data for training the machine learning models. To address this issue, various models have been proposed for collecting and recycling sign language data, such as depth cameras, marker-grounded systems, and wearable detectors. Another challenge is the higher variance in sign language gestures, which makes it difficult to train models to recognize specific signs with a high efficiency. In terms of speech conversion, some studies have used Text-to-speech (TTS) has been employed for speech conversion, which has limitations such as a lack of persuasive voice and restricted language support. More research is needed to develop reliable sign recognition models and improve the voice conversion quality.

In Sign Language Recognition, numerous former studies that have proposed different styles and ways to improve the accuracy of the recognition system. A comparison of crucial features from former exploration and a specific paper in Sign Language Recognition would involve looking at factors similar to the dataset used, the AI model armature, input and output considered, the evaluation criteria, and the results achieved. For illustration, a comparison of crucial features between former exploration and a specific paper in Sign Language Recognition could include the following:

1. The dataset for prior exploration may have used a different dataset than the specific study, which can affect the results achieved.
2. Input and Output for former exploration may have used different input and affair formats than those in the specific paper, which can affect the performance of the recognition system.
3. The evaluation Criteria in prior exploration may have used different evaluation criteria than in the specific study, which can affect the interpretation of the results.
4. The results achieved by former exploration may have been different results than the specific study, which can affect the overall conclusion of the exploration.

In conclusion, a comparison of crucial features between former exploration and a specific paper in subscribe Language Recognition would involve looking at factors similar to the dataset used, input and output, evaluation criteria, and results achieved. Understanding these parallels and differences.

1. **Proposed Methodology**

In this section, we explain the proposed methodology in detail. The proposed model of sign language recognition for speech interpretation using our methodology involves several key steps, culminating in the implementation of a real-time sign language interpreter utilizing dynamic time warping and Mediapipe technologies.Illustrated in Figure-1, our dynamic time warping classification model serves as the cornerstone of the system, aimed at generating corresponding words based on sign language recorded in a video recorded through the camera. Our sign language recognition model focuses mainly on hand landmarks, representing significant points of interest on the hand, which we meticulously track and utilize as input data. The prediction process in our model proceeds in the following manner:

1. Extract Landmarks
2. Calculate the DTW distance between the recorded and reference signs.
3. Predict the sign/gesture by analyzing which reference sign is likely to be recorded.
   1. **Problem domain**

The domain encompasses the development of a comprehensive system capable of real-time detection and interpretation of hand gestures from sign language, bridging the communication gap between individuals with and without speech impairments.

* 1. **Problem Definition**

Creating an effective sign language interpreter using Mediapipe and Dynamic Time Warping, capable of instantaneously translating sign language into spoken and written language. This involves overcoming the complexities inherent in sign language recognition owing to the non-rigid nature of the human body, individual variations in gestures, viewpoint uncertainties, and environmental factors such as illumination and occlusion.

* 1. **Problem Statement**

**C**reating an accessible solution for individuals with speech impairment, facilitate seamless communication with diverse audiences. Beyond converting sign language into spoken words, the proposed interpreter aims to offer affordability, portability, and user-friendliness, accommodating varying proficiency levels in sign language users. By integrating Mediapipe, Dynamic Time Warping, and Natural Language Processing (NLP), this research endeavors to contribute to inclusive technology, addressing the communication barriers faced by those with speech disabilities.

* 1. **Dataset**

## The Indian government hosts an extensive Indian Sign Language (ISL) dataset on its website, encompassing a rich repository of sign language gestures. This collection comprises diverse gestures, expressions, and linguistic nuances vital for comprehensive sign language interpretation. Developed to ensure cultural sensitivity, this dataset serves as a fundamental resource for training models aimed at enhancing communication accessibility for individuals reliant on ISL as their primary mode of expression. We have also included our sign language videos to train our model which include the various words as represented in Table I.

Table I: Custom training dataset

|  |  |  |
| --- | --- | --- |
| Index | Action Type | Training  Sample set |
| 1 | How | 5 |
| 2 | You | 5 |
| 3 | I | 5 |
| 4 | Good | 5 |
| 5 | What | 5 |
| 6 | Time | 5 |
| 7 | Where | 5 |
| 8 | Washroom | 5 |
| 9 | Please | 5 |
| 10 | Help | 5 |

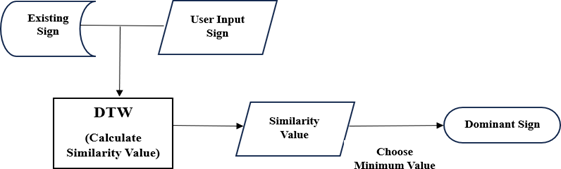


Figure 1. Dynamic Time Warping Classifier

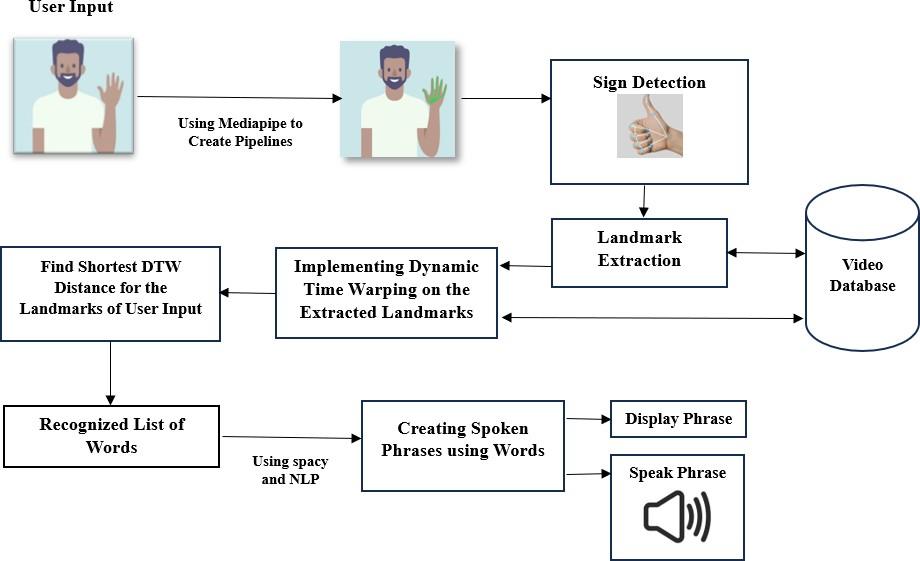


Figure 2. Proposed Model for Sign Language Interpretation using Dynamic Time Warping

1. **Implementation**

The proposed Dynamic Time Warping (DTW) classifies and contrasts time series data. The focus is on calculating the similarity between the two-time series. Here, we compare the existing and user input signs for classification. For this study, we considered a custom dataset for training our model. The contents of our model are listed in Table I. These are some of the basic sign language gestures that are widely used in communication. The number of training samples is also mentioned. The implemented system for sign language interpretation is illustrated in Figure 2.

* 1. **MediaPipe Detection/Holistic Model**

MediaPipe is an open-source computer vision framework that was developed by Google a few years ago. Within the suite of models offered by MediaPipe, the Holistic Model stands out, enabling the real-time tracking of Hand, Pose, and Face landmarks. In our current model, we leverage Hand positions and Pose landmarks, including the shoulder, elbow, and wrist positions, to inform our predictions.

* 1. **Extract landmarks**

Initially, we initiated the process by extracting landmarks from the video stream using the Holistic model. To ensure seamless integration with OpenCV, a crucial step involves converting the colors. OpenCV operates in BGR color space, whereas MediaPipe uses RGB colors. This color conversion is imperative for the accurate rendering and analysis of extracted landmarks.

* 1. **Draw landmarks**

The “drawing\_utils” sub-package within MediaPipe offers a convenient toolkit encompassing all the necessary instruments needed to effortlessly visualize landmarks on a singular image.

* 1. **Models**

The current interpretation of the project involves the execution of two distinct models. The "HandModel" class encapsulates detailed information about hand gestures within a single image. Meanwhile, the "SignModel" class encompasses and extends the "HandModel" data that are seamlessly applied across all frames within a video, ensuring comprehensive coverage and accurate analysis throughout the entire footage.

* 1. **Dynamic Time Warping (DTW)**

It is a widely employed algorithm that is specifically designed to compare time series data. It excels in identifying optimal alignments between the two time series through a warping technique, allowing the analysis of patterns rather than mere sequences. In our context, DTW is invaluable for recognizing similarities between embeddings of the same signs. This method enables us to identify the most similar signs based on DTW distances, providing a reliable means of sign recognition. The resulting approach provides a comprehensive list of signs sorted in the order of their proximity to the recorded sign. This sorting is achieved by the respective distances calculated using the Dynamic Time Warping technique. For a detailed understanding of the DTW calculation process, we must consider two matrices to find similarities between them. The following represents how to use Dynamic Time Warping to determine the similarity score:

**Input:** **T** = [t*i,d*]*MxD*and **S**= [s*i,d*]*MxD*

**Output:** similarity between two input action matrices

function matrixsimlilarity(T,S)

let [w*i,j*]*MxN*be warping matrix aligning two feature representations

set w*i,j* infinity for all i and j

for *i=1* to M do

for *j=1* to N do

let T*i* be the row vector of matrix T at row *ith*

let S*j* be the row vector of matrix S at row *jth*

distance = EuclidDistance (T*i*,S*j*) =

w*i,j* = distance + min( w*i-1,j-1* , w*i-1,j* , w*i,j-1* )

end

end

return w*M,N*

end function

* 1. **Sign Prediction**

After determining the distances between the recorded sign and all reference signs, sorting them allows us to form a batch comprising the signs most similar to our record. This batch serves as a crucial indicator, helping us determine if a sign repetition occurs frequently enough to bolster our prediction confidence. In our research, we chose a batch size of 2 and set the threshold to 0.8. This configuration signifies that if the same sign appears at least two times within a batch, we confidently generate the corresponding sign label as the output. However, if the sign does not meet this frequency criterion, the output is labelled as "Unknown sign." The choice of batch size and specific threshold values is contingent on the availability of videos for each sign within the dataset, ensuring optimal accuracy in our predictions.

* 1. **Phrase Generation**

Based on the keywords obtained from our sign recognition model, we passed a list of these words to form meaningful phrases or sentences with the help of the spacy library in NLP and provided sentences constructed both in the form of text and audio. This phase of the model ensures that the sign input provided by the user is converted to its corresponding text and speech format. The combination of Mediapipe, Dynamic Time Warping (DTW) and Natural Language Processing (NLP). The implementation of our model is described in several words and greetings in Figure 4. In addition, we describe the simulation and interpretation of certain sentences in Figure 5. The simulation led to the successful interpretation of sign language in speech using our model. During the simulation, both the text and speech forms of the associated sign are obtained accurately.

1. **RESULT**

In this section, we discuss the recognition ability of the proposed model in terms of accuracy and complexity. The accuracy and complexity of the two approaches used for recognition were compared. The first method used Long Short-Term Memory (LSTM) for recognition, and the other used Dynamic Time Warping. The model in which recognition was performed using LSTM required a large amount of data for each sign to work efficiently. To train an LSTM model, we must run iterative epochs, and the time required to complete these epochs depends on the size of the dataset and the computational power of the system. Thus, the LSTM model had worse time and space complexity. However, the model in which DTW was used for recognition required less data as it compared the similarity between the two signs. This model is more accurate and had less time and space complexities. Based on the three test cases conducted, the accuracy was obtained as mentioned in TABLE-2.



Figure 4. Using the combination of Mediapipe and Dynamic Time Warping to Interpret Distinct ISL Signs

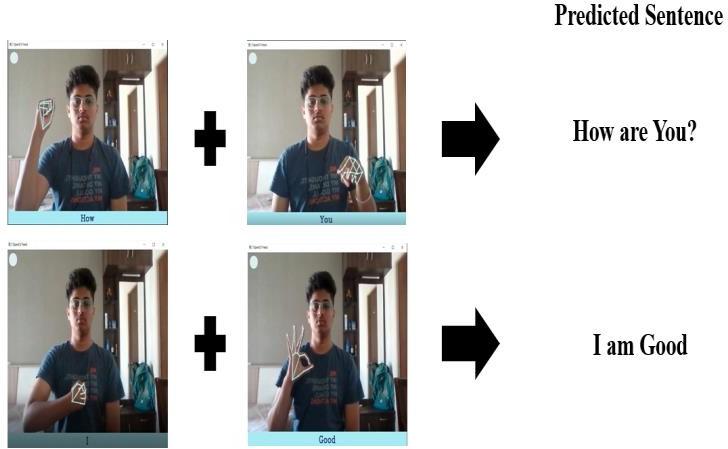
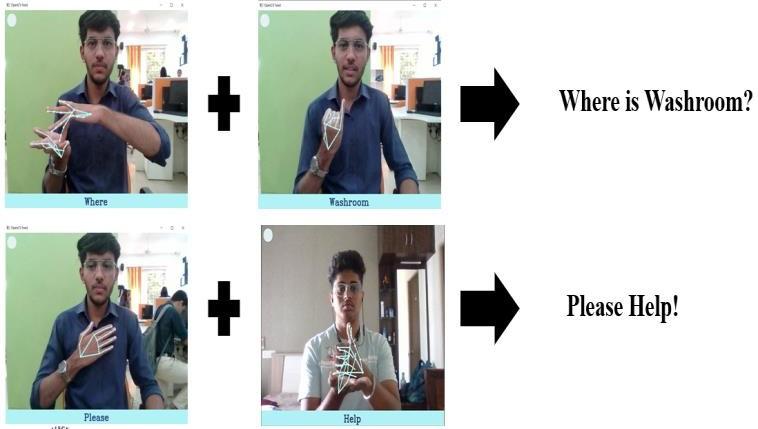


Figure 5(a)

Figure 5(b)

Figure 5. Using the proposed Sign to Sentence Interpretation using Dynamic Time Warping to Run the Simulation against our model and generate the Sentences based on the Indian Sign Language

Table 2: Model Accuracy Comparison

|  |  |  |
| --- | --- | --- |
| Test Case | LSTM  Accuracy | DTW Accuracy |
| Test 1 | 80% | 92% |
| Test 2 | 83% | 95% |
| Test 3 | 82% | 93% |
| Average | 81.6% | 93.3% |

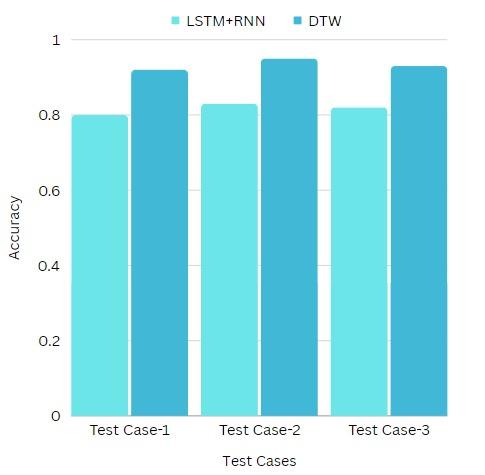


Figure 6. Comparison of Accuracy of LSTM+RNN and DTW

1. **Discussion**

We developed a state-of-the-art sign language to speech interpreter using Dynamic Time Warping, Mediapie and NLP for Indian Sign Language (ISL). It can be used to convert a given sign into its respective conversion in spoken language. The present sign language generation model has shown promising results based on DTW scores. However, evaluating its real-world performance, it requires input from expert sign language users. Assessing the coherence and synchronization of non-manual elements, such as body posture, head movements, and eye gaze, is pivotal. Although the model learns broad hand movements, it lacks the finesse of finger and facial movements. One potential remedy involves exploring alternative loss functions such as keypoint loss. In addition, the limitations in dataset size and diversity might constrain the model's ability to capture the intricate nuances of sign language. Investigating how different signer styles affect the model's output and devising adaptive strategies for varying signing styles is imperative for practical implementation.

1. **Conclusion**

This paper outlines a successful and comprehensive approach to sign language recognition using MediaPipe and Dynamic Time Warping. By leveraging the model's capabilities to track hand positions, pose landmarks, and facial landmarks in real time, this study enabled the prediction of sign language gestures. It also provides an effective and efficient approach to sign language recognition, making a valuable contribution to the fields of computer vision and gesture recognition. The use of Dynamic Time Warping for sign comparison allows for robust recognition even under variations in sign execution speed. Additionally, the reliance of the approach on feature vectors and time-series analysis minimizes the need for extensive training data, making it a practical solution for real-world applications in sign language recognition.

##### **References**

1. Yifan Zhang, Ling Long, Diwei Shi, Haowen He, and Xiaoyu Liu, “Research and Improvement of Chinese Sign Language Detection Algorithm Based on YOLOv5s”, 2nd International Conference on Networking, Communications and Information Technology (NetCIT)
2. Aakash Deep, Aashutosh Litoriya, Akshay Ingole, Vaibhav Asare, Shubham M Bhole, and Shantanu Pathak, “Realtime Sign Language Detection and Recognition”, 2nd Asian Conference on Innovation in Technology (ASIANCON)
3. Dhruvin Gandhi, Kushal Shah and Madhav Chandane, “Dynamic Sign Language Recognition and Emotion Detection using MediaPipe and Deep Learning”, 13th International Conference on Computing Communication and Networking Technologies (ICT)
4. Gouri Anilkumar, M S Fouzia, and G S Anisha, “Imperative Methodology to Detect the Palm signs (American Sign Language) using YOLOv5 and MediaPipe”, 2nd International Conference on Intelligent Technologies (CONIT)
5. Aishwarya D Shetty, Jyothi Shetty, Karthik K, Rakshitha, and Shabari Shedthi B, “Real-Time Translation of Sign Language for Speech Impaired”, 7th International Conference on Computing Methodologies and Communication (ICCMC)
6. Sakshi Sharma and Sukhvinder Singh, “Recognition of Indian Sign Language (ISL) Using Deep Learning Model”, Wireless Personal Communications -An International Journal -Springer
7. N Rajasekhar, M Geetha Yadav, Charitha Vedantam, Karthik Pellakuru, Chaitanya Navapete, “Sign Language Recognition using Machine Learning Algorithm”, International Conference on Sustainable Computing and Smart Systems (ICSCSS)
8. B. V. Chowdary, Ajay Purshotam Thota, Alwa Sreeja, Kotla Nithin Reddy, and Karanam Sai Chandana, “Sign Language Detection and Recognition using CNN”, International Conference on Sustainable Computing and Smart Systems (ICSCSS)
9. Shagun Katoch, Varsha Singh, and Uma Shanker Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN", International Journal of Trend in Scientific Research and Development (IJTSRD)
10. Atharv Ganpatye and Sunil Mane, “Motion Based Indian Sign Language Recognition using Deep Learning”, 2nd International Conference on Intelligent Technologies (CONIT)
11. Kotkondawar, R.R., Sutar, S.R., Kiwelekar, A.W., Wankhede, H.S., “Performability of Deep Recurrent Neural Networks for Molecular Sequence data”, International Journal of Computing and Digital Systems, 2023, 13(1), pp. 1317–1327
12. Laddha, M.D., Netak, L.D., Wankhede, H.S., “A Correlation Analysis Between Cognitive Process and Knowledge Dimension in Software Engineering by Using the Revised Bloom’s Taxonomy”, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2021, 12615 LNCS, pp. 34–41
13. Sahu, M., Dhawale, K., Bhagat, D., Wankhede, C., Gajbhiye, D., “Convex Hull Algorithm based Virtual Mouse”, 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, 2023-June, pp. 846–851
14. Sahu, M., Francis, N.A., Sahu, A., Ulhe, P., “Data Wrangling and Visualization for Credibility and Usability”, Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022, 2022, pp. 1834–1840
15. Nirale, P., Madankar, M., “Design of an IoT Based Ensemble Machine Learning Model for Fruit Classification and Quality Detection”, International Conference on Emerging Trends in Engineering and Technology, ICETET, 2022, 2022-April
16. Madankar, M., Chandak, M., Chavhan, N., “Information Retrieval System Based on Query Translation Approach for Cross-Languages”, Lecture Notes in Electrical Engineering, 2021, 700, pp. 1261–1269
17. Sonekar, S.V., Pal, M., Tote, M., Sawwashere, S., Zunke, S., “Computation termination and malicious node detection using finite state machine in mobile ad-hoc networks”, Proceedings of the 7th International Conference on Computing for Sustainable Global Development, INDIACom 2020, 2020, pp. 156–161, 9083710
18. Pal, M.M., Jasutkar, R.W., “Implementation of hand vein structure authentication-based system”, Proceedings - International Conference on Communication Systems and Network Technologies, CSNT 2012, 2012, pp. 114–118, 6200601
19. Agarwal, M., Kotecha, A., Deolalikar, A., ...Yadav, R.K., Thomas, A., “Deep Learning Approaches for Plant Disease Detection: A Comparative Review”, 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2023, 2023
20. Agarwal, M., Kalia, R., Bahel, V., Thomas, A., “AutoEval: A NLP Approach for Automatic Test Evaluation System”, 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies, GUCON 2021, 2021