ZKG INTERNATIONAL

ISSN: 2366-1313

DEAFBLIND PEOPLE COMMUNICATE USING SIGN LANGUAGE TO TEXT AND SPEECH

[1] Mr. A V Murali Krishna, [2] Pulimamidi Bhavya Sri, [3] Chinthakunta Roselyn, [4] Chappala Rohan

[1]Associate Professor, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

avmurali002@matrusri.edu.in

[2, 3, 4, B.E] Student, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

bhavyasri.3274@gmail.com, shinycrisolyte@gmail.com, cheppalarohan@gmail.com

ABSTRACT:

In this study, we delve into the fascinating realm of sign language comprehension, a crucial area within gesture recognition research. Our approach centers on utilizing camera vision-based methods for image analysis. Our overarching objective is to create a comprehensive sign language system capable of not only recognizing standard English phonetics but also interpreting a wide array of gestures, subsequently transcribing them into text and, ultimately, spoken sentences. To achieve this, we leverage convolutional neural networks (CNNs), a form of computing technology designed to mimic the workings of the human brain. Moreover, we integrate an audio built-in system to facilitate communication between the hearing impaired community and the general public. This system incorporates text-to speech (TTS) synthesis, a technique that automatically converts text into speech. By implementing TTS synthesis, our study enables the system to audibly articulate the letters and words it identifies, thereby enhancing accessibility and communication for individuals with hearing impairments.

Keywords: Gesture Recognition Systems, Convolutional Neural Network, Text to speech conversion, Histogram.

ISSN: 2366-1313

I INTRODUCTION

The focus of this project lies in Sign Language Recognition, a burgeoning field fueled by recent advancements in Computer Vision and Deep Learning. The primary aim is to leverage Neural Networks to predict various hand gestures from real-time images while enhancing the efficiency of recognition through computer vision techniques.

Prior research in Sign Language Recognition has yielded hardware-dependent systems utilizing specialized gloves, armbands, or wristbands, which are expensive and complex to use. Conversely, vision-based static gesture recognition software offers a more accessible and cost-effective solution, eliminating the need for additional hardware. Some systems even extend to dynamic gesture recognition using Motion History Images (MHI) to track real-time hand movements.

Different image processing techniques are employed for data preparation, ranging from direct resizing to binary conversion. While histograms and back-projection have been utilized for this purpose, they introduce computational overheads. To address these challenges, this study proposes leveraging native OpenCV functions for faster and more efficient gesture recognition, thereby reducing dependence on histograms and streamlining the process. The significance of Sign Language Recognition extends beyond mere technological advancements. It holds potential for revolutionizing human-computer interaction, particularly for individuals with disabilities. By interpreting pointing gestures, computers can facilitate natural communication, aiding those who are unable to verbally communicate. Moreover, gesture recognition complements spoken language, enabling seamless communication in diverse contexts.

OBJECTIVES

The primary objective of our project is to develop a model capable of recognizing fingerspelling-based hand gestures in Indian Sign Language. Specifically, we aim to:

• Train the model to accurately identify and interpret individual gestures.

• Enable the model to combine recognized gestures to form complete words.

• Facilitate seamless communication between D&M individuals and the broader community by converting sign language gestures into easily readable text.

ISSN: 2366-1313

By achieving these objectives, we aim to address the communication barriers faced by millions of D&M individuals, ultimately promoting inclusivity and accessibility in society.

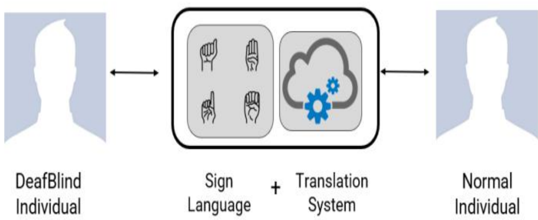


Fig 1. Sign language recognition and translation system model

II. LITERATURE SURVEY

In 2016, K.A Bhaskaran, A.G. Nair ET. al. developed an Equipment based Gesture Recognition System using smart gloves to track and recognize Hand Gestures.

The system utilized sensors attached to the gloves to track hand/finger movement and transduce finger flexion into electric signals for determining hand posture. Although accurate, this system lacked the ease of use factor of simple Vision-based systems and also incurred higher costs. The gloves were versatile, finding applications not only in gesture recognition but also in medical areas and video gaming.

In 2018, Nuwan Munasinghe developed a dynamic hand gesture recognition system capable of recognizing moving gestures instead of static ones.

The system generated Motion History Images from consecutive frames using an algorithm measuring structural similarity. Various OpenCV functions were employed for image processing, including Gaussian Background Subtraction, median filtering, and Otsu's thresholding. The resulting Motion History Images were classified by a Neural Network to predict the gestures.

In 2019, Hanwen Huang, Yanwen Chong et al. proposed a Vision-based Gesture Recognition system.

This system detected skin based on a specified formula and converted hand images into binary format. After preprocessing, hand contours were extracted using a 16-layer VGGNet. Gesture Recognition was performed using a Pyramidal-pooling module along with attention mechanisms. The system achieved a remarkable 98.41% accuracy over a test dataset of 900 images.

In 2020, researchers on Research gate-IEEE detailed a Hand Gesture Recognition system for Automated Speech Generation.

This system utilized an HSV model, Large Blob Detection, Flood Fill, and Contour

ISSN: 2366-1313

Extraction techniques for gesture recognition. The results showcased accurate, reliable, and efficient gesture processing and voice analysis.

In 2021, S. Divya presented "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning."

This study focused on achieving accurate translation between spoken English and ISL gestures, employing CNN, SVM, and RNN classifiers. Methodologically, the study involved image pre-processing, feature extraction, and the utilization of the Google Speech Recognition API and PYAudio for speech-to-text conversion. Training on 30,240 photos resulted in an impressive 88.89% test accuracy.

III SYSTEM ANALYSIS

EXISTING SYSTEM

Sign language to speech conversion systems aim to facilitate communication between deaf and hearing individuals. However, they encounter various limitations. Some struggle with adaptability, while others face computational complexities or training constraints. Few of them are given below.

Limitations of Existing system

1. Rule-based Systems:

Drawback: Limited Adaptability - Rule-based systems rely heavily on predefined rules and patterns. They might struggle with handling variations in sign language gestures that deviate from these predefined rules, leading to inaccuracies in translation.

2. Hidden Markov Models (HMM):

Drawback: Lack of Context Awareness - HMMs model transitions between states based solely on observed data. They may struggle to capture the semantic context of sign language gestures, resulting in less accurate translations, especially in cases where context plays a crucial role.

3. Support Vector Machines (SVM):

Drawback: Over fitting with Small Datasets - SVMs tend to perform well with highdimensional data but can over fit when trained on small datasets. In sign language translation, where datasets may be limited, over fitting can lead to poor generalization to unseen gestures, reducing the accuracy of translation.

4. Dynamic Time Warping (DTW):

Drawback: High Computational Complexity - DTW computes the optimal alignment between two sequences, which can be computationally intensive, especially for longer sign 7 language sequences. This high complexity may limit its scalability for real-time translation applications

5. Recurrent Neural Networks (RNN):

Drawback: Vanishing/Exploding Gradient Problem - RNNs can suffer from the vanishing or exploding gradient problem during training, especially with long sequences. This can impede the model's ability to capture long-term dependencies in sign language gestures accurately.

6. Long Short-Term Memory (LSTM) Networks:

Drawback: While LSTMs address the vanishing gradient problem of traditional RNNs, they may still struggle with capturing very long-term dependencies effectively. This limitation could affect the accuracy of converting complex sign language gestures into spoken words.

7. Random Forests:

Drawback: Random forests can be prone to over fitting, especially with small training datasets commonly available for sign language recognition. Over fitting may lead to poor performance on new sign gestures not seen during training**.**

PROPOSED SYSTEM

In summary, Existing sign language to speech conversion systems face a range of limitations, including adaptability issues, computational complexities, and over fitting challenges with small datasets. These drawbacks hinder their accuracy and scalability, impacting their effectiveness in facilitating communication between deaf and hearing individuals. However, the CNN algorithm offers promise in addressing these challenges by leveraging convolutional neural networks and histogram-based feature extraction to improve the translation accuracy and robustness of sign language interpretation systems.

Proposed system Advantages:

1. Improve the translation accuracy and robustness of sign language interpretation systems.

2. Flexibility and Simplicity

IV IMPLEMENTATION

Architecture:

ISSN: 2366-1313

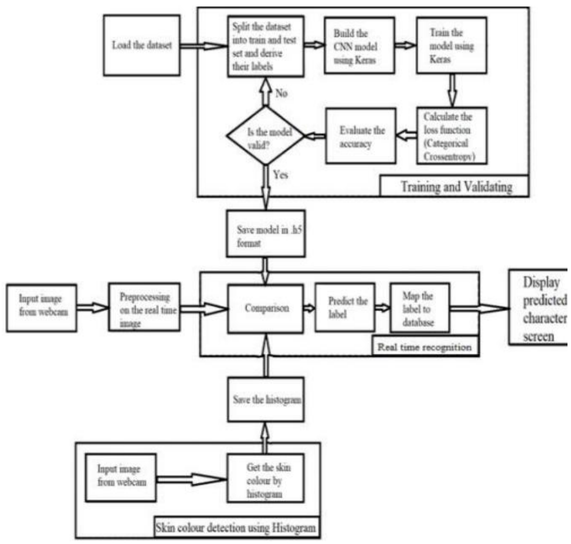


Fig-2. Architectures of the system model

1. System Architecture

In tackling the complexities of sign gesture recognition, our proposed methodology merges the versatile capabilities of OpenCV with the robustness of Convolutional Neural Networks (CNNs), all under the umbrella of Deep Learning. This amalgamation of technologies offers a promising avenue for overcoming the challenges inherent in accurately interpreting sign language gestures.

In our quest to bridge the gap between sign language gestures and textual representations, we leverage the pyttsx Python module to convert predicted text into speech. This enables our system to not only interpret sign language gestures but also vocalize the corresponding textual output, thereby fostering seamless communication between individuals with differing communication modalities. In essence, our proposed methodology represents a holistic approach to sign gesture recognition, leveraging cutting-edge technologies and meticulous data preparation techniques to develop a robust and accurate system capable of facilitating meaningful communication within the sign language community.

2. Modules

Hand Histogram Configuration

The hand histogram is a crucial feature that captures the user's skin tone. Alteration in lighting necessitates reconfiguration of the histogram. The setup involves the following stages:

• Preprocessing of Image Data: In this phase, images are transformed into a uniform format with a resolution of 75x75 pixels.

• Creation of Detection Squares: A total of fifty squares are displayed to the user. The user then positions their hand so that it completely covers the squares. By pressing

ISSN: 2366-1313

the "C" key, the system detects the skin tone within these squares. A structural element function "getStructuringElement()" with "MORPH_ELLIPSE" parameter assists in highlighting the skin-colored objects as white and the rest as black on the image.

• Saving the Histogram: Once satisfied, the user can save the hand histogram using the 'pickle' library by pressing the "S" key.

Real-time Image Preprocessing

• Frame Adaptation: Initially, video frames are captured via a webcam, flipped for proper orientation, and resized for consistency.

• Noise Reduction: Gaussian and Median blur techniques are used to reduce image noise.

• Histogram Comparison: The previously saved histogram file is used in a function, ensuring that only the user's hand is showcased as white against a black background.

• Conversion to Binary Image: The "thresh" function is utilized to convert the image into a binary format, distinguishing between two color levels for clarity.

Model Comparison and Probability Assessment

The real-time binary images are compared with a saved model in .h5 format using the "predict" function. This function yields the prediction probabilities and associated classes, with only classes having a prediction probability exceeding 90% being taken forward for display.

Database Interaction

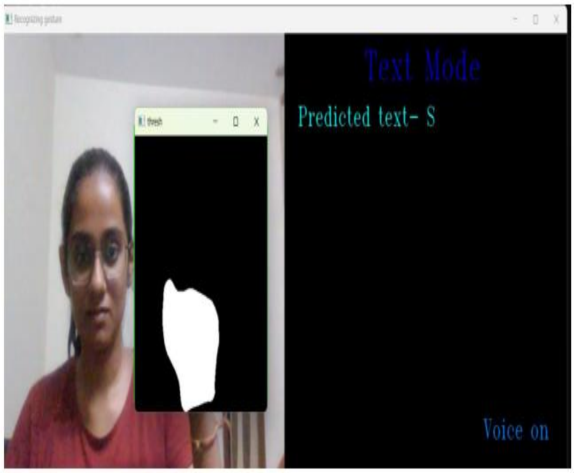
Each resulting prediction class is cross-referenced with a pre-existing database, 'gesture.db', which contains a table of gesture IDs and corresponding names, enabling the conversion of recognized gestures into actual characters for output.

Output Visualization

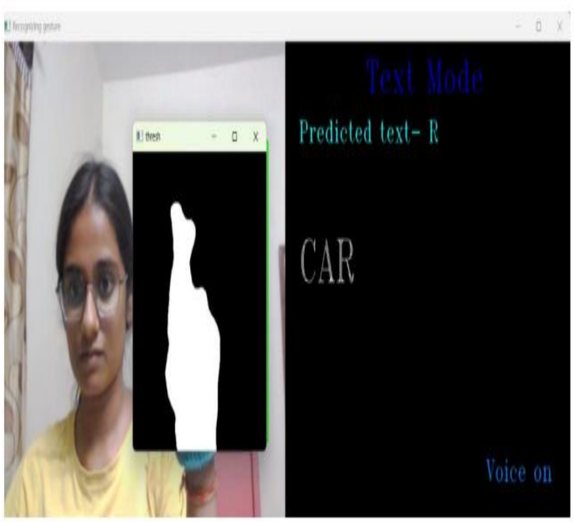
Finally, recognized characters are exhibited on the screen using the "imshow" function from the OpenCV library. This is how characters are clearly presented to the user on a digital blackboard within the interface

V RESULT AND DISCUSSION

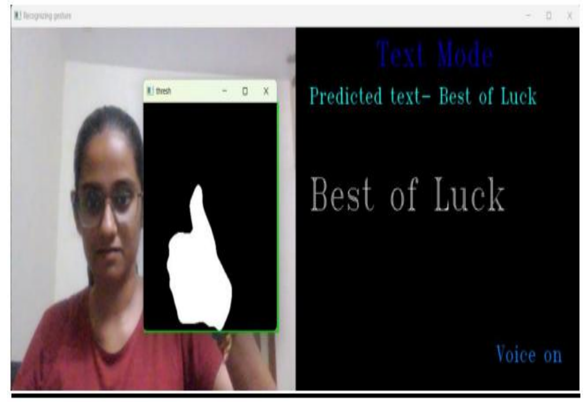
Recognition of Letters:



Recognition of word “CAR”:



Recognition of Gestures:

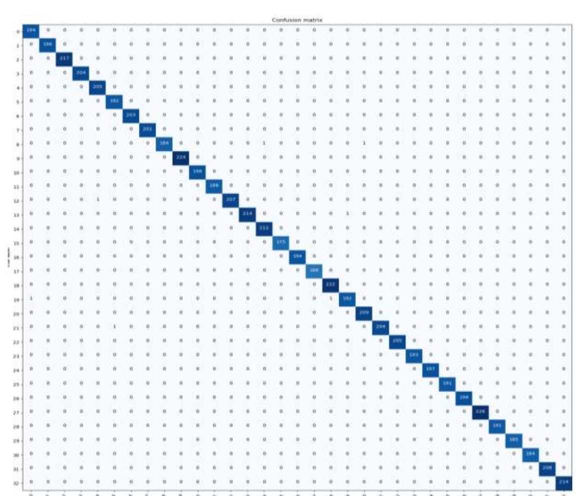


Classification Report:

Classification Report

	precision	recall	f1-score	support
0	0.99	1.00	1.00	194
1	1.00	1.00	1.00	196
2	1.00	1.00	1.00	217
3	1.00	1.00	1.00	204
4	1.00	1.00	1.00	205
5	0.99	1.00	1.00	192
6	1.00	1.00	1.00	203
7	1.00	1.00	1.00	201
8	1.00	0.99	0.99	186
9	1.00	1.00	1.00	224
10	1.00	1.00	1.00	198
11	1.00	1.00	1.00	188
12	1.00	1.00	1.00	208
13	1.00	1.00	1.00	215
14	1.00	1.00	1.00	212
15	1.00	1.00	1.00	175
16	1.00	1.00	1.00	194
17	1.00	1.00	1.00	166
18	1.00	1.00	1.00	222
19	0.99	0.99	0.99	194
20	1.00	1.00	1.00	209
21	1.00	1.00	1.00	205
22	1.00	1.00	1.00	205
23	1.00	1.00	1.00	193
24	1.00	1.00	1.00	197
25	1.00	1.00	1.00	191
26	1.00	1.00	1.00	198
27	1.00	1.00	1.00	226
28	1.00	1.00	1.00	191
29	1.00	1.00	1.00	185
30	1.00	1.00	1.00	184
31	1.00	1.00	1.00	208
32	1.00	1.00	1.00	214
accuracy			1.00	6600
macro avg	1.00	1.00	1.00	6600
weighted avg	1.00	1.00	1.00	6600

The confusion matrix for 33 classes



ISSN: 2366-1313

VI CONCLUSION

Final Remarks Individuals with hearing impairments stand to benefit substantially from the findings of this study. Thanks to the aforementioned efforts, such users will experience enhanced and precise on-screen representation of the English alphabet. The evidence garnered from this study advocates for a spectrum of user-centric services, aiming to reduce the complexities associated with day-to-day communications. A primary ambition of this study was to refine the detection of gestural representations of letters A through Z. An example lies in the swift identification of the gesture signifying the letter "C," underscoring the study's success in not only accelerating recognition rates for certain letters but also maintaining consistency across the board. Marked improvements have been observed in the speed and accuracy of gesture recognition. The employment of Convolutional Neural Networks was central to these advancements. In essence, this research has effectively met all proposed targets. Optimized performance of the system was ensured, as each component functioned as predicted. Rigorous testing protocols were put in place for each system element throughout every developmental phase, with results meticulously verified to guarantee full operational integrity. This scholarly endeavor led to the conceptualization and subsequent implementation of a user-friendly desktop application.

FUTURE ENHANCEMENTS

This research serves as a foundational model for addressing prevalent communication issues. Future developments could expand the applicability of this work in several ways, including: Development of a Web Application: This would enable real-time video communication through web browser interfaces. Creation of an Android Application: This would allow for gesture-based communication on mobile platforms, incorporating a messenger service. Expansion to Sentence Recognition: Efforts could evolve towards recognizing gestures that convey everyday sentences, enhancing the system's utility for comprehensive communication. By exploring these avenues, it is envisaged that the technology will become more accessible, allowing for broader adoption and more seamless integration into the daily lives of the hearing-impaired community.

VII REFERENCES

[1] K. A. Bhaskaran, A. G. Nair, K. D. Ram, K. Ananthanarayanan and H. R. N. Vardhan,

"Smart gloves for hand gesture recognition: Sign language to speech conversion system," 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), Kollam, 2016, pp. 1-6.doi10.1109/RAHA.2016.7931887,URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7931887&isnu mber =7931858

[2] "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning" Dhivyasri S; Krishnaa Hari K B; Akash M; Sona M; Divyapriya S; Krishnaveni V, Published in: 2021 3rd International Conference on Signal Processing and Communication (ICPSC) a. https://ieeexplore.ieee.org/document/945169

[3] Hand-gesture recognition for automated speech generation,May 2016,DOI:10.1109/RTEICT.2016.7807817. Conference: 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). Sunny Patel(Indian Institute of Technology (Banaras Hindu University) Varanasi) Ujjayan Dhar, Suraj Gangwani ,Rohit Lad (Full Sail University) a. https://www.researchgate.net/publication/312325638_Handgesture_recognition_for_automated_speech_generation .

[4] A Survey on Hand Gesture Recognition for Indian Sign Language.Profile image of Mahasweta JoshiMahasweta Joshi,2016 a. https://www.academia.edu/98786079/A_Survey_on_Hand_Gesture_Recognition_for_Indian_Sign_Language 49

[5] Human Skin Detection Using RGB, HSV and YCbCr Color Models S. Kolkur1, D. Kalbande2, P.Shimpi, C.Bapat,and J. Jatakia- Department of Computer Engineering, Thadomal Shahani Engineering College, Bandra,Mumbai, India(Department of Computer Engineering, Sardar Patel Institute of Technology, Andheri,Mumbai, India){kolkur.seema@gmail.com;drkalbande@spit.ac.in; prajwalshimpi@gmail.com;chai.bapat@gmail.com;jatakiajanvi12@gmail. com}

[6] International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9(2018) pp.7154-7161© Research India Publications. http://www.ripublication.com.

[7] C. Zhu, J. Yang, Z. Shao and C. Liu, "Vision based hand gesture recognition using 3d shape context," in IEEE/CAA Journal of Automatica Sinica, doi: 10.1109/JAS.2019.1911534.

ISSN: 2366-1313

[8] EvilPort2. Sign language gestures, Oct 2017

AUTHORS

Mr. A V Murali Krishna, Associate Professor, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

Email: avmurali002@matrusri.edu.in

Miss. Pulimamidi Bhavya Sri, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

Email: bhavyasri.3274@gmail.com

Mr. Chinthakunta Roselyn, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

Email: shinycrisolyte@gmail.com

Mr. Chappala Rohan, Dept. of CSE, Matrusri Engineering College, Saidabad, Hyderabad.

Email: cheppalarohan@gmail.com