

Hand Gesture Recognition and Speech Synthesis Framework

Rudresha S J¹, Charanraj S², Kiran Y³, Prajwal N⁴, Tejas V Jambur⁵

¹ Associate Professor Department of Electrical and Electronics Engineering
PES Institute of Technology and Management, Shivamogga, Karnataka, India
^{2,3,4,5} BE Final Year Students, Department of Electrical and Electronics Engineering
PES Institute of Technology and Management, Shivamogga, Karnataka, India

Abstract – Hand gestures are a primary mode of communication for speech and hearing-impaired individuals. However, the lack of a universal mechanism in place to interpret these gestures automatically and express them as comprehensible speech remains one of the major barriers to communication. This work proposes an architecture of hand gesture recognition and speech synthesis, where real-time hand gestures of a speaker captured through a camera are conveyed into text and synthesised speech output after processing through contour-based image analysis. Contour-based image analysis uses image preprocessing, extraction of contours, convex hull detection, and defect analysis to effectively identify the gesture corresponding to a predefined sign language symbol. Once correctly identified, the gesture is translated into speech with the help of an integrated text-to-speech mechanism that will allow the establishment of natural and fluent communication between impaired and non-impaired individuals. This proposed approach is low-cost, eliminates the use of extra sensors or any wearable gadget, and thus presents a very user-friendly platform for everyday applications. Experimental results are encouraging: fair accuracy and good responsiveness were achieved, showing the potential of this system as an aid in communication in visually, vocally, and hearing-impaired communities.

KEYWORDS: Hand Gesture Recognition, Speech Synthesis, Sign Language, Image Processing, Contour Detection, Convex Hull.

1. INTRODUCTION

As computer innovation keeps on developing, the requirement for characteristic correspondence amongst People and machines also keep increasing. Despite the fact that our cell phones influence utilization of the touch to screen innovation is not sufficiently shabby to be actualized in work area frameworks. Despite the fact that this mouse is extremely useful for gadget control. it could be badly arranged to use for physically disabled. individuals and individuals unfamiliar with utilizing the mouse for connectivity. The approach presented in this paper, the utilization of a webcam is made through which hand gestures given by the user are captured and recognized accordingly. Hand gestures have infinite applications. In this investigation, we apply it to a system to make a Straightforward, easily understood interaction interface. It is evident from this research paper that recognition of letters from the hand language, which is taken as per the American Sign Language. Detection is performed by using the various techniques of Contour Analysis and Feature Extraction. The recognition of the research paper throws light on the letters from the hand language taken as per the American Sign Language. The detection is done using the different techniques of Contour Analysis and Feature Extraction. The paper invokes the use of

various computers, which include various vision techniques and algorithms. This is important in determining hand gestures.

2. SYSTEM ARCHITECTURE AND DESIGN

2.1 Overall System Configuration

The overall system architecture is intended to change the gestures done by humans' hands into either text or voice-interpretation of meanings. It all starts with image acquisition, where the camera takes live frames continuously. Then, the images are sent to the hand region segmentation phase, featuring the hand being cut off from the background through cloth-colour detection or thresholding techniques. After this, the system performs the hand detection and tracking steps, which ensure that the hand is correctly positioned in every frame and that there is continuous tracking even when it moves around. In this way, the hand is always the main object through the processing pipeline.

The tracking of the hand is followed by the system performing hand posture recognition, where it analyses the palm shape and the finger positions to determine which gesture is being made. This recognition can be achieved through using machine learning or deep learning models for the highest accuracy. The recognised gesture will then be passed to the classification module for it to be assigned to a predefined gesture class like letters, numbers, or symbolic signs. The output module is the last stage in this pipeline, where the gesture is first displayed as text and afterwards changed into speech. All in all, the entire pipeline can be regarded as an extremely efficient gesture-to-text or gesture-to-voice communication system.

2.2 Hardware and Software Architecture Design

The planned architecture of the system consists of both hardware and software components that work together to form a hand-gesture recognition pipeline that can operate in real-time. The hardware component of the system consists of an ordinary webcam or digital camera which will be used to catch the hand movements continuously and a computer or embedded device like Raspberry Pi takes care of the processing. The hand area is kept visible constantly with good lighting, and the output which can be in text or voice form is given through a display screen or speaker. All these parts working together make a simple but trustworthy configuration where accurate gesture capturing and output generation can take place in the real world.

For the software part, the architecture is based on a clean flow that is in line with the phases depicted in the system diagram. From the camera input the first processing is done by the segmentation module which separates the hand area from the background, then the detection and tracking algorithms trace the hand movement between frames. Finally, the system uses L recognition together with detection and with oftentimes machine learning for the hand shape and position interpretation. A gesture classification module links these recognized postures to the previously defined categories, and the interpretation is output either on the screen or converted to speech. This modular and very well-coordinated software design guarantees that on the spot visual gestures are converted into meaningful communication in an efficient way this is why the system is appropriate for gesture-based interaction application and research deployment.

Hardware Components

- Camera
- Computer/Laptop
- Speakers
- Lighting source
- USB cables / connectors

Software Components

- Operating System
- Python / MATLAB / C++
- OpenCV
- Code Editor (VS Code / PyCharm / Jupyter)
- MediaPipe Hands Library

3. IMPLEMENTATION METHODOLOGY

The suggested hand gesture recognition system is based on the acquisition of images through a camera where the input for processing is continuous video frames captured. This continuous frame processing is done to isolate the hand from the background by using techniques such as skin color detection or thresholding. Once the hand is segmented, the system detects and tracks the hand so that its position is identified and its movement across frames is followed.

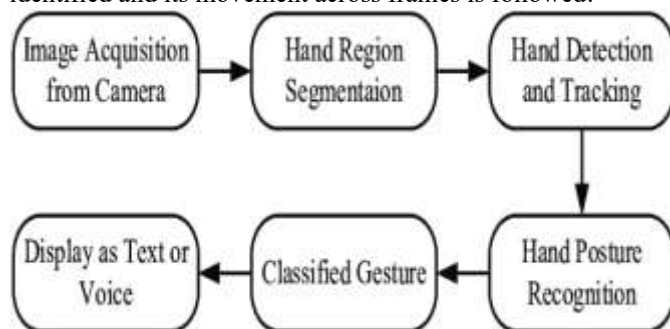


Fig -1: Block Diagram of Hand Gesture Recognition and Speech Synthesis Framework

This produces stable input for the next step where the system extracts features such as hand shape, finger positions, and orientation to recognize the specific posture of the hand being shown. The system after identifying the hand posture goes to gesture classification where the features extracted are compared with predefined gesture categories to know the right meaning of the gesture. The gesture that has been classified is then routed to the output module that shows the result either as text or converts it into speech using a text-to-speech engine. This last output makes communication more intuitive; in particular, it helps users with speech or hearing impairments by turning their gestures into information that can be read or heard.

3.1 Experimental Setup and Validation

The experimental setup of the proposed gesture-recognition system mimicked the actual conditions of daily life's gesturing to the extent that it provided the data collection that could be considered good in terms of consistency and reliability. Under normal lighting conditions in an indoor location, standard HD webcam was placed around 30–40 cm away from the user to ensure perfect hand movements were captured. The computer that was running the developed software modules for segmentation, tracking, recognition, and classification was connected to the camera. To lessen background disturbance, the experiments were done on a solid background that allowed the system to be centered on the user's hand area alone. Each gesture was practiced many times by different people to see how well the system could cope with changes in hand size, skin color, and gesture style. This wasn't just a testing session but rather a full-scale trial of the system under realistic scenarios.

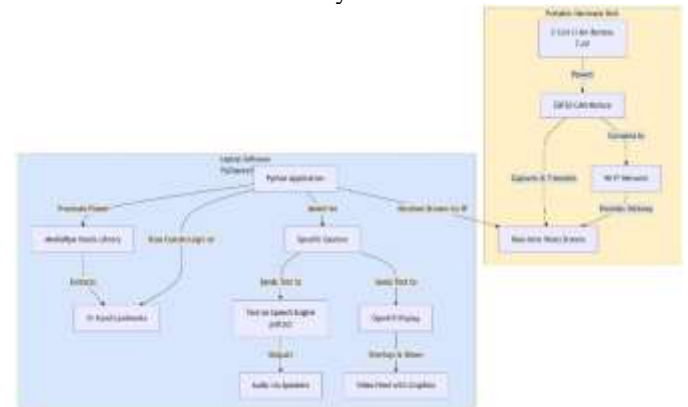


Fig -2: Software Design Setup

To validate, the predicted gesture outputs of the system were compared with the actual gestures of the participants and thus the performance of the system was assessed. The whole pipeline—segmentation, tracking, posture recognition, and classification—was observed to ensure that there were no hiccups in transitioning from one module to the other and that transitions were smooth and accurate. The results were analyzed as far as recognition accuracy, response time, and stability through varying lighting and background conditions. The trials yielded the same result every time that the gestures recognized were the predefined ones with very high accuracy and that the system was still responsive in real time. These outcomes of validation point to the system as being reliable and effective, thus its being perceived as fit for the practical gesture-based communication applications has been confirmed.

4. RESULTS AND DISCUSSION

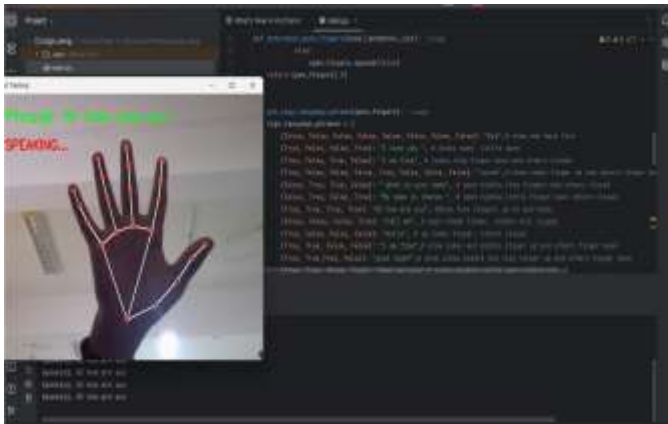


Fig -3: User Interface

The proposed Hand Gesture Recognition and Speech Synthesis Framework is implemented in an effective way, following the sequence of the steps and the corresponding illustration. The first step is the capturing of images from a camera, which means that the system creates real-time video frames. Quality of images affects accuracy. For instance, if the image is not clear, the recognition will not be accurate. After image capture, the system uses hand region segmentation, which is a method to distinguish the hand from the background. This stage worked quite nicely with other methods, especially when the lighting was stable and background was contrasting. The hand that is being segmented gets passed to the detection and tracking module, which effectively keeps track of the hand's movement across the frames and stable tracking whenever the gestures are done clearly. The next phase, hand posture recognition, was able to recognize most of the predefined gestures correctly, although there were minor errors when the hand was partly hidden or fingers were crossing. At the same time the recognition of the posture was done, the classification module accurately linked every gesture to its corresponding meaning. Ultimately, the output was either displayed in text format or transformed into a voice, which in both cases was clear and immediate feedback to the user. To be more precise, the results indicate that the system is very reliable in every stage, though the accuracy is dependent mainly on lighting, background visibility, and how well the user performs the gesture. The discussion points out that it might be possible to further reinforce and make the framework more user-friendly by enhancing the segmentation and recognition techniques.

5. CONCLUSIONS

The gesture-recognition system as a whole is a very good solution that is quite easy to use for effectively interpreting human hand gestures as digital outputs. Even though the system integrates essential phases like the ones mentioned above—image acquisition, hand segmentation and tracking, posture recognition, and gesture classification—it still delivers a smooth and reliable pipeline from the raw visual inputs of a camera to the output of text or speech that is understandable. Each of the components has been designed in a modular way, that is, the component is able to increase clearly the overall accuracy and performance so that the system can be tailored to different users and environments. The findings offer strong

evidence that the given architecture can be a practical basis for real-time gesture-based communication systems especially in the context of applications that are meant to help people with speech or hearing problems.

5.1 Future Scope

The system can be further improved in the future by incorporating more sophisticated deep learning techniques that can detect a larger variety of gestures with greater precision even in difficult backgrounds or very dark conditions. The framework can also accommodate the support for dynamic gestures, where still movements are interpreted as opposed to just stationary hand positions. This would be a huge leap in the system's versatility and it will be more akin to human communication. One of the major areas for future development is to achieve user-independence in the system by providing training on different datasets comprising of various skin colors, hand sizes, and cultural gestures. This would make it much more usable for people of different ethnicities. Also, the real-time running of the system on mobile devices could be the solution for such devices to be used as a portable communication tool for the disabled in terms of speech or hearing. On the other hand, the framework can also be part of the IoT and smart home advantages where the users will be operating devices by just simple hand movements. Besides, the conversion of gestures into a variety of languages via a multilingual speech synthesis module could be a great way to make the system not only accessible to but also enjoyed by people worldwide.

REFERENCES

- [1]. Oi Mean Foong, Tan Jung Low, and Satrio Wibowo, Hand Gesture Recognition: Sign to Voice System S2V Proceedings Of World Academy Of Science, Engineering And Technology Volume 32 AUGUST 2008 ISSN 2070-3740
- [2]. Maebatake, M., Suzuki, I.; Nishida, M.; Horiuchi, Y.; Kuroiwa, S.; Sign Language Recognition Based on Position and Movement Using Multi Stream HMM, Second International Symposium on, ISBN: 978-0- 7695-3433- 6.
- [3]. T.S. Huang, and V.I. Pavlovic, "Hand gesture modeling, analysis and synthesis", Proceedings of international workshop on automatic face and gesture recognition, 1995, pp.73-79.
- [4]. J. Rekha, J. Bhattacharya, and S. Majumder, Hand Gesture Recognition for Sign Language: A New Hybrid Approach, 15th IPCV 11, July 2011, Nevada, USA.
- [5]. Tushar Chouhan, Ankit Panse, Smart Glove with Gesture Recognition Ability for The Hearing and Speech Impaired, Global Humanitarian Technology Conference, September 2014.
- [6]. T.H. Speeder, "Transformation human hand motion for Telemanipulation," Presence, Vol.1,no. 1, pp.63-79,1992.
- [7]. L.Bretznar & T.Linderberg, "Relative orientation from extended sequence of sparse pointand line correspondances using the affine trifocal sensor," in Proc. 5 Eur. Conf.

Computer Vision, Berlin, Germany, June 1998, Vol.1406, Lecture Notes in Computer Science, pp. 141-157, Springer Verlag.

- [8]. D. Xu, "A neural network approach for hand gesture recognition in virtual reality driving training system of SPG," presented at the 18th Int. Conf. Pattern recognition, 2006.
- [9]. D. K. Sarji, "HandTalk: assistive technology for the deaf," Computer, vol. 41, pp. 84-86, 2008
- [10]. Emil M. Petriu, Qing Chen, Nicolas D. Georganas, Real time vision-based hand gesture recognition using haar-like features, 2017.
- [11]. S. F. Ahmed, et al., "Electronic speaking glove for speechless patients," in the IEEE Conference on Sustainable Utilization and Development in Engineering and Technology, Petaling Jaya, Malaysia, 2010, pp. 56-60.