Gesture Recognition Using MediaPipe and Keras for Smart Interfaces

ANIKET PATEL[1], HARSHIT KATIYAR[2], KUSHAL GUPTA[3]
[1, 2, 3]*School of Computing, Science and Engineering, Galgotias University, Greater Noida, India*

Abstract- It creates a significant distance between user and the computer, the use of a physical controller, such as a mouse or keyboard, hinders a natural interface in human-computer interaction. We have developed a reliable marker-less hand gesture recognition system in this paper that effectively tracks both dynamic and static hand gestures. When a gesture is detected, our system converts it into actions like opening websites and starting programs like PowerPoint and VLC Player. In a presentation, the gesture is used to move through the slides. Our findings demonstrate that minimal hardware requirements and can be used to create an intuitive HCI. Despite being a highly regarded technique for human-computer interaction, hand gesture recognition still faces many obstacles. The issue of individual user style variation, which has a substantial impact on system performance, is addressed in this paper. An efficient, flexible graphical user interface that supports user-defined hand gestures is presented here in this project. In contrast, previous research only permitted the manual addition of personalized hand gestures in extremely limited application contexts. Our system uses data from a specific user to train a camera-based hand gesture recognition. model for that user, allowing hand gestures to be personalized. In contrast to earlier recognition models that require large training datasets, we use a lightweight Multilayer Perceptron architecture based on contrastive learning, which reduces the size of the data required and the training timeframes. While a user study gathers and examines some preliminary user feedback on the system in use, experimental results show that the recognition model converges quickly and achieves satisfactory accuracy.

I. INTRODUCTION

Hand gestures are movements of the palms and fingers used to express ideas and thoughts. With hand gestures, users can use their hands to engage with the machine without the need for additional input devices.

They are becoming more and more common in varie ty of Human-Computer Interaction (HCI) applications because to advancements in both general computer t echnology and specialized hardware equipment. In order for humans and machines to communicate effectively, the machine must be able to recognize human gestures through a process known as hand gesture recognition (HGR). With a broad range of application domains, such as virtual gaming, virtual reality, robot control, and natural user interfaces, it's one of the main areas of present research.

Moreover, another important use of HGR is in technology-assisted human communication systems, such as those that recognize sign language. An experiment where users described an abstract chai r and a simple chair using their hand gestures reveale d that people preferred to create a hand gesture langu age with two hands in a virtual environment. The aforementioned studies examine user behavior in order to create better, standardized hand gesture vocabularies; however, they fail to address the issue that users from various backgrounds, such as cultural ones, may have very different definitions of the same hand gesture.

The suggested camera-based HGR, another area that has already garnered a lot of research interest, serves as the technical foundation of our approach. It is a multidisciplinary field of study that blends aspects of artificial intelligence, pattern recognition, and image processing.

Hand gesture identification from a single image or t he succession of photos from a single camera is still difficult, though, because it is usually difficult to infer 3D information from 2D data, and because possible links across sequences may make the task e ven more difficult.Ferreira et al. presented a novel model based on the Contrastive Transformer to show that learning rich representations from point sequences leads to strong classification between vector embeddings. The motivation behind our work can be summed up as follows: although a number of studies have examined user hand gesture behavior in an effort to create standardized hand gesture vocabularies, neither the users' varied cultural backgrounds nor their unique signing styles were taken into consideration.

However, improving single-camera-based hand gesture recognition will have immediate practical implications because in reality, there is typically only one screen and one camera in front of a user. We suggest that a user interface that uses a single camera and requires little data collection, and has user-defined hand gesture textual inputs, will allow each user to use it in a way that is expressive to them—that is, to create their own original expressions.

II. PROPOSED HAND GESTURE RECOGNITION SYSTEM

The back end and front end make up the entire system.Three modules make up theback end system: the interface, detection and camera modules.

The following is a summary of them:

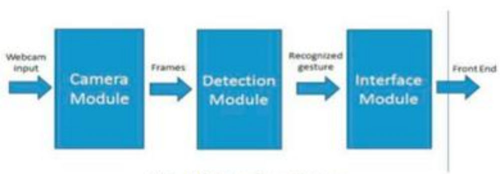


Fig. 1 Back end architecture

3.1 Camera module :

The camera module is in charge of connecting to various image detectors, capturing input, and sending the image to the detection module for processing in the form of frames.

3.2 Detection module :

The task of processing the photos falls the detecting module.Prior to contour extraction, the output of the camera module is subjected to a number of image processingtechniques, such as color conversion noise reduction, and thresholding. The gesture is recognized by looking for convexity faultsin the image, if there are any.

3.3 Interface module:

The interface module is in charge of translating the hand gestures that are detected into the corresponding actions. The relevant application is then notified of these actions. There are three windows on the front end. The first window displays the video input that was recorded by the camera along with the name of the gesture that was identified. The contours present in the input images are shown in the second window. The image's smooth thresholded version is shown in the third window. The benefit of including the threshold and contour window in the graphical user interface is that it alerts the user to background irregularities that could impact the system's input, allowing them to modify their laptop or desktop webcam to avoid them.

III. RECOGNITION HAND GESTURE

Using gesture modeling and analysis, the input visual gesture images are identified as meaningful gestures during the recognition phase. The correct choice of gesture parameters for features has an impact on the recognition process and, consequently, the classification accuracy. For instance, gesture recognition cannot use edge detection or contour operators since they may result in incorrect classification. Neural networks have been used extensively for hand gesture recognition and hand shape extraction.

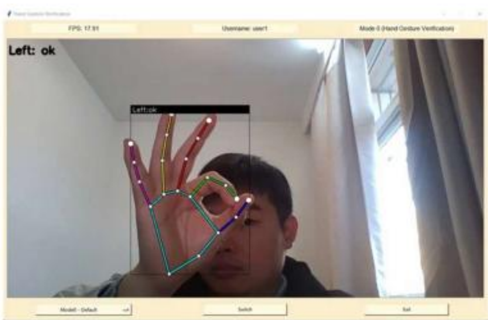
Excel has demonstrated its ability to function as a recognition tool for the Hidden Markov Model (HMM). The fuzzy c-mean clustering algorithm was used to control robot motion after clustering algorithms took a chance in the recognition process.

IV. USER INTERFACES FOR USER-DEFINED HAND GESTURE RECOGNITION

Numerous user-defined hand gesture user interfaces have already been developed, but the majority of them only offer a few features, like allowing the user to choose between two pre-established hand gestures.

Furthermore, we point out that the user experience may suffer if lengthy times are required to gather the required data, so it is crucial to consider this when creating unique hand gestures. It would take hours for users to define numerous hand gestures, which would be bad for the user experience. Our proposed user interface aims to significantly improve timings. Lastly, we point out that various user interfaces have been suggested for particular application areas.

However, assuming that only a small subset of t he vocabulary—that is, a limited number of words— would be needed in practice, we provide users with in putable methods to build a vocabulary correspon ding to their own hand gestures, allowing us to use a lightweight neural network for recognition trained with contrastive learning, which further reduces th e amount of time a user must spend defining a han d gesture through the interface. The majority of previ ous research that we have looked at is based on fixed vocabularies, and usually, users are given alarge num ber of hand gesture words to learn and use.



V. APPLICATIONS

Hand gesture recognition is crucial for several domains, including human-computer interaction, robotics, virtual andaugmented reality,in and sign language recognition. Future movies' characters make use of hand gestures to show how interaction systems can affect how we interact with computers if the ones we have now do not permit such flexibility.

6.1 Virtual and Augmented Reality Technologies :
Virtual reality (VR) is a digital technology that uses headsets to create sounds, images, and emotions to simulate a human presence in a virtual environment. While virtual reality (VR) totally replaces the real world with an artificial one, augmented reality (AR) enhances it.

6.2 Recognition of Sign Language :
Using nonverbal cues like gestures, sign language serves as a primary language for people with speech and hearing impairments. Sign language, Fingerspelling, and Sign Language are the three primary components of sign languages. Using gestures to spell words is known as fingerspelling. People who are hard of hearing frequently need the help of a qualified interpreter because of their limited comprehension of sign language.

But employing an interpreter isn't always practical and can be expensive. A sign language recognition system that uses visual cues could offer a natural, comfortable, and affordable means of communicating with people who are deaf or hard of hearing and have speech impairments.

7.1 Advantages :
Every user has a personal account and is able to create a customized vocabulary and recognition model. Unlike the user interface allows them to enter any words or phrases and generate a unique hand gesture representation of that textual input, unlike current user in the customized interfaces that offer a restricted range of possibilities to accomplish some limited customization in the project. Following Mediapipe feature vector extraction, contrastive learning with data augmentation is used to achieve high accuracy and rapid convergence.

7.2 Challenges :
As discussed in, hand gesture recognition systems face numerous difficulties, including: Variation in lighting conditions: any alteration in lighting has a negative impact on the area of skin that has been removed from the hand.

Rotation issue: this issue occurs when the hand area in the scene rotates in any direction.

Background problem: This is the term used to describe a complex background in which the hand objects are surrounded by other objects, some of which may have skin-like colors that cause misclassification issues.

Scale issue: this issue arises when the gesture image's hand positions vary in size.

Translation issue: erroneous feature representation is also caused by variations in hand positions across images.

CONCLUSION

The review highlights the significance and difficulties of hand gesture recognition in a number of domains, including robotics, virtual reality, gaming, sign language recognition, and human-computer interaction. It examines various methods, including vision-based, sensor-based, and data glove-based approaches, as well as their benefits and drawbacks. For the purpose of recording and interpreting hand gestures, precise hand modeling and feature extraction are essential. Classifying and identifying gestures using extracted features requires machine learning algorithms. Real-time performance, noise, complicated backgrounds, and changing lighting are some of the difficulties in hand gesture recognition. In order to increase the robustness, accuracy, and usability of hand gesture recognition systems, the review recognizes the necessity for additional study and development.

We are able to develop a reliable gesture recognition system without the need for markers, which made it more affordable and user-friendly. Our goal with this gesture recognition system is to offer gestures that cover nearly every facet of HCI, including system functions, application launches, and opening well-known websites. We hope to increase the accuracy even more in the future and add more gestures to enable more features.

Lastly, we hope to expand our domain scenarios and integrate our tracking mechanism into a range of hardware, such as mobile devices and digital TVs.

Additionally, we want to make this mechanism accessible to a wider range of users, including those with disabilities.

REFERENCES

[1] Granit Luzhnica, Elizabeth Lex, Viktoria Pammer. A Sliding Window Approach to Natural Hand Gesture Recognition using a Custom Data Glove. In: 3D User Interfaces (3DUI); 2016 IEEE Symposium on ; 2016 Mar 19 ; New York: IEEE; 2016 ; p.81-90.

[2] Ji-Hwan Kim,Nguyen Duc Thang,Tae- Seong Kim. 3-D hand Motion Tracking and Gesture Recognition Using a Data Glove. In; Industrial Electronics; 2009 IEEE International Symposium on ; 2009 July 5; New York : IEEE;2009 ; p.1013- 1018.

[3] Hung CH, Bai YW, Wu HY. Home outlet and LED array lamp controlled by a smartphone with a hand gesture recognition. In: Consumer Electronics (ICCE); 2016 IEEE International Conference on ; 2016 Jan 7; New York : IEEE;2016 ; p.5-6.

[4] Hung CH, Bai YW, Wu HY. Home appliance control by a hand gesture recognition belt in LED array lamp case. In: Consumer Electronics (GCCE); 2015 IEEE 4th Global Conference on ; 2015 Oct 27; New York : IEEE;2015; p. 599-600.

[5] She Y, Wang Q, Jia Y, Gu T, He Q, Yang B. A real-time hand gesture recognition approach based on motion features of feature points. In: Computational Science and Engineering (CSE); 2014 IEEE 17th International Conference on; 2014 Dec 19; New York: IEEE;2014;p.1096-1102.

[6] Lee DH, Hong KS. A Hand gesture recognition system based on difference image entropy. In: Advanced Information Management and Service (IMS), 2010 6th International Conference on; 2010 Nov 30; Seoul; New York: IEEE; 2010 ; p. 410- 413.

[7] Dulayatrakul J, Prasertsakul P, Kondo T, Nilkhamhang I. Robust implementation of hand gesture recognition for remote human-machine interaction. In:Information Technology and Electrical Engineering (ICITEE); 2015 7th International Conference on ;2015 Oct 29;p. 247-252.

[8] Tsai TH, Huang CC, Zhang KL. Embedded virtual mouse system by using hand gesture recognition. In: Consumer Electronics-Taiwan (ICCE-TW); 2015 IEEE International Conference on; 2015 Jun 6; Taiwan, Taipei; New York: IEEE; 2015; p. 352-353.

[9] Hussain I, Talukdar AK, Sarma KK. Hand gesture recognition system with real- time palm tracking. In: India Conference (INDICON);2014 Annual IEEE ;2014 Dec 11; India, Pune; New York: IEEE; 2014; p. 1-6.

[10] Huong TN, Huu TV, Le Xuan T. Static hand gesture recognition for vietnamese sign language (VSL) using principle components analysis. In: Communications, Management and Telecommunications (ComManTel); 2015 International Conference on; 2015 Dec 28; p. 138-141.

[11] Chen Y, Luo B, Chen YL, Liang G, Wu X. A real-time dynamic hand gesture recognition system using kinect sensor. In: Robotics and Biomimetics (ROBIO); 2015 IEEE International Conference on ; 2015 Dec 6; New York : IEEE;2015; p. 2026- 2030.

[12] C. Wang, Z. Liu and S. C. Chan. Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera. IEEE Transactions on Multimedia 2015; 17(1): 29-39.

[13] Chen WL, Wu CH, Lin CH. Depth- based hand gesture recognition using hand movements and defects. In: Next- Generation Electronics (ISNE); 2015 International Symposium on; 2015 May 4 ; Taiwan, Taipei ; New York : IEEE;2015 ;p. 1-4.

[14] Wong WS, Hsu SC, Huang CL. Virtual touchpad: Hand gesture recognition for smartphone with depth camera. In: Consumer Electronics Taiwan (ICCE- TW);2015 IEEE International Conference on; 2015 Jun 6; Taiwan, Taipei; New York: IEEE;2015 ; p. 214-215.

[15] Ishiyama H, Kurabayashi S. Monochrome glove: A robust real-time hand gesture recognition method by using a fabric glove with design of structured markers. In: Virtual Reality (VR); 2016 IEEE ;2016 Mar 19; Greenville, SC; New York : IEEE;2016;p. 187-188.