Hand Gesture Recognition using CVZONE

Tran Tan Thien
thienttse171043@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

Nguyen Anh Hao
haonase171792@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

Ho Quang Hien
hienhqse173123@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

Ngo Duc Toan
toanndse173680@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

Vu Ngoc Khanh
khanhvnse170082@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

Huynh Le Trung Viet
viethltse170565@fpt.edu.vn
FPT University
Ho Chi Minh City, Viet Nam

ABSTRACT

This research investigates the use of the 'cvzone' library for hand gesture recognition, which is an important part of improving human-computer interaction. We create a system capable of interpreting and responding to a range of hand motions using computer vision techniques and machine learning. Our strategy, which real-time image processing, yields encouraging results in terms of accuracy and responsiveness. This technology's potential uses span from virtual reality and gaming to assistive devices, all of which promise a more intuitive and engaging user experience.

KEYWORDS

cvzone, hand gesture recognition, mediapipe hand landmark

ACM Reference Format:
Tran Tan Thien, Nguyen Anh Hao, Ho Quang Hien, Ngo Duc Toan, Vu Ngoc Khanh, and Huynh Le Trung Viet. 2024. Hand Gesture Recognition using CVZONE. In *Proceedings of International Conference on Intelligent Information Technology (ICIIT).* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

We are currently living in the Industry 4.0 era. The industrial revolution is gaining momentum. Artificial intelligence, automation, big data, and the internet of things (IoT) are always required in industries. Data centers are critical components of the modern digital world, acting as the foundation for the expanding demand for cloud computing and online services. With ever-increasing data volumes and the demand for real-time processing, data center energy usage has become a major concern. The global expense and environmental consequences of high energy consumption require novel ways to increase energy efficiency in these facilities.

In this research, we look at the development and testing of a 'cvzone'-based hand gesture identification system. Our key goals are as follows:

1. Research: 'cvzone' library overview. This section provides an overview of the 'cvzone' library, introducing its features, core capabilities, and integration with 'OpenCV' for computer vision tasks.

2. Construction: Build a model based on the functions of image processing, evaluation and recognition of hand gesture.

In conclusion, the purpose of this research is to investigate the possibilities and challenges of implementing hand motion detection with the 'cvzone' library. We hope to contribute to the advancement of natural and intuitive human-computer interactions by harnessing the power of computer vision and machine learning, ushering in a new era of technology accessibility and user engagement.

2 RELATED WORK

The field of hand gesture recognition has made significant advances through the use of Python libraries such as CVZone, e.g. This study [1] was conducted on a humanoid robot with an upper body implementing CVZone for real-time hand gesture recognition. Performance was evaluated at different distances and brightness levels, with the highest success rate observed at the shortest distance (50 cm) in medium light environments. In another survey [2], hand gesture recognition was discussed in the context of enhancing user interaction in artificial intelligence. This research also uses Python libraries, including OpenCV and cvzone2, for image capture, preprocessing, and image detection. These tools are used in conjunction with mapped action pairs to perform specific tasks. A real-time on-device hand gesture recognition (HGR) system was presented in another study [3]. The system is capable of detecting a predefined set of static gestures from an RGB camera, showing potential for real-time applications. Furthermore, this study [4] leveraged a large ASL dataset and advanced techniques to capture the intricate details and movements of ASL gestures accurately. This study highlights the potential of these techniques in understanding complex gestures. On the other hand, with a small dataset, there is a comparison of some methods [5]. Wide-DenseNet without transfer learning and PrototipicalNetworks with transfer learning were figured out to have the best result. In the context of Data-Free Class-Incremental Learning, this study [6] presents BOAT-MI, a Boundary-Aware Prototype Model Inversion, which delves deeper into the choice of the best sample for inversion thereby bringing significant improvements compared to traditional methods. Another technique for hand recognition problems is using a Temporal Convolution Network for Hand Action Recognition Framework [7]. The framework employs a simplified skeleton representation, standardization steps

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICIIT, Feb. 23–25, 2024, Ho Chi Minh City
© 2024 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

to enhance generalization, and a motion summarization module for efficient per-frame descriptor processing. With low-resolution thermal images, M.Vandersteegen et al. [8] introduced a hand gesture recognition model with a lightweight, low-latency algorithm, which combines a 2D Convolutional neural network (CNN) with a 1D Temporal Neural Network (TCN). This study addresses the challenges of low-cost processors. These studies provide a comprehensive overview of the current state-of-the-art techniques in hand gesture recognition using CVZone. However, there is still room for improvement and exploration which this study aims to achieve.

3 METHOD

Hand Recognition, an intricate field that has undergone decades of extensive research and development, necessitates the application of sophisticated algorithms and advanced techniques. With of the challenges within this problem, we introduce 'cvzone' as a streamlined solution for hand recognition, meticulously developed on the MediaPipe framework. Cvzone distinguishes itself by delivering exceptional performance, easy for use and rapid data processing, facilitated by its robust hand detection algorithm. It also boasts high practicality, with integrated functions for the preprocessing of raw data sourced from images and videoss. The foundation of cvzone on the Google MediaPipe platform ensures both operational stability and commendable performance. This innovative approach signifies a substantial leap forward in the realm of hand recognition, elevating the precision and efficiency of recognition procedures through the judicious amalgamation of contemporary technology and advanced algorithms.

In the context of Hand Recognition, CVZone has developed a solution based on the MediaPipe Hand Landmarker framework. This is a straightforward approach to detect and identify key points on the hand in images. This method operates on image data using Machine Learning, be it static data or a continuous stream, and subsequently provides information such as key landmarks on the hand based on image coordinates and information regarding whether it's the left or right hand for multiple detected hands. The model for identifying these hand landmarks consists of 21 coordinates representing the positions of hand joints. A connecting line is drawn between these key points, and the angles between them are meticulously calculated. This intricate processensures the precision of hand gesture detection in a realtime setting. To reduce complexity during operation, this method utilizes the region defined by the landmark points in the initial frame to determine the hand's position in subsequent frames. The hand recognition method is only activated when no hand is detected or when tracking is lost. This helps reduce the frequency of re-invoking the recognition model from the beginning, thereby improving the model's performance.

The cvzone method represents a significant advancement in the field of hand recognition, grounded in the technical prowess of the MediaPipe framework. Its streamlined and dependable nature positions it as a valuable tool for a wide range of scenarios. The method's continued evolution through focused research and development will be instrumental in addressing current limitations and expanding its utility.

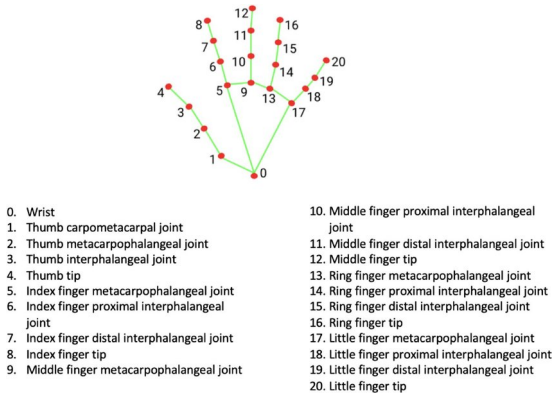

Figure 1: 21 Key Poins in Hand

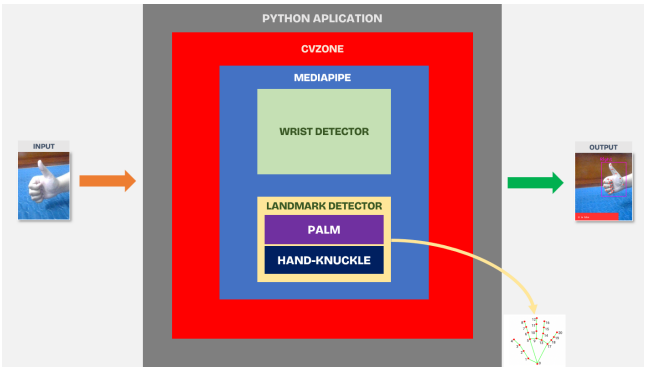

Figure 2: Model Work

The model takes in images or videos, including direct images from the camera containing the user's hand. Based on the input data, the model undergoes processing steps to identify the wrist and landmarks on the hand. The objective is to locate the points on the palm. Once these target points are identified, the model begins to match them with 21 pre-defined key points and connects these points to determine hand gestures. After processing as described, the complete output is generated as shown in the sample image.

```
import cv2
from cvzone.HandTrackingModule import HandDetector
detection = HandDetector(detectionCon=0.9, maxHands= 2)
video = cv2.VideoCapture(0)
```

Figure 3: Import and use Cvzone

4 EXPERIMENT

a. Data

We compiled our dataset by inviting multiple individuals to participate in model testing, utilizing a webcam for data collection. This effort resulted in an extensive dataset that comprises over 100 test images. This diverse dataset not only enhances the depth of our resources but also strengthens the model's robustness.

b. Model's hyperparameters

Our HandDetector Model has 5 parameters as follow:

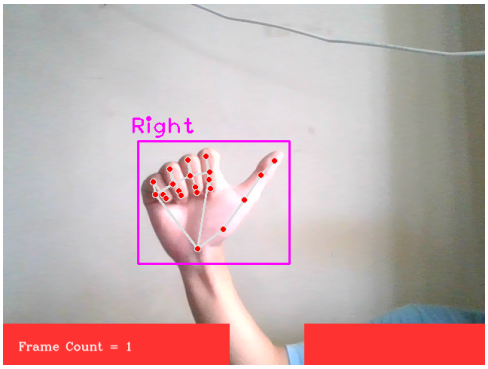

Figure 4: Model Testing

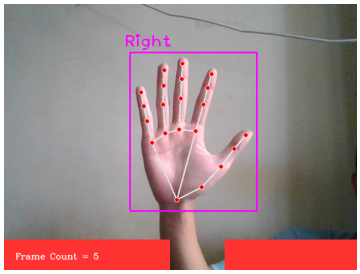

Figure 5: Example about Data

staticMode (Static Mode):

It controls whether the detector should perform hand detection on each image or on a sequence of images. When staticMode is set to True, the detector will perform hand detection on each image. This is slower, but it can be more accurate in cases where the hand position is changing rapidly. When staticMode is set to False, the detector will perform hand detection on a sequence of images. This is faster, but it can be less accurate in cases where the hand position is changing rapidly.

maxHands (Maximum Number of Hands):

It controls the maximum number of hands that the detector will attempt to detect in each image. When maxHands is set to 1, the detector will only attempt to detect one hand in each image. When maxHands is set to 2, the detector will attempt to detect two hands in each image.

modelComplexity (Model Complexity of Hand Landmark Model):

It controls the complexity of the hand landmark model used by the detector. The higher the modelComplexity value, the more accurate the landmark detection will be, but it will also be slower.

detectionCon (Minimum Detection Confidence Threshold):

It controls the minimum detection confidence threshold used by the detector. The higher the detectionCon value, the more confident the detector must be in a detection before it will be considered valid. right balance between reliable detections and the number of detections.

minTrackCon (Minimum Tracking Confidence Threshold): It controls the minimum tracking confidence threshold used by the detector. This threshold determines how confident the detector needs to be in a detection before it will be considered a valid track. A higher minTrackCon value means that the detector will be more likely to track hands that are moving slowly or smoothly, while a lower minTrackCon value means that the detector will be more likely to track hands that are moving quickly or erratically.

c. Testing process

After collecting a diverse set of 100 image samples, encompassing a range of characteristics, sizes, and angles, comprising two categories: Palm and Dorsal side of the hand, we conducted rigorous testing and analysis. This included statistical assessments, data-driven evaluations, and result reporting.

d. Evaluation metrics

The accuracy metric was chosen for three specific reasons:

Comprehensive Evaluation: Accuracy is an overarching metric that provides an overall assessment of the model's capability in classifying objects or events.

Representation of Precision: It offers insight into the model's ability to make both correct and incorrect classifications, making it a versatile metric for overall performance evaluation.

Model Comparison: Accuracy enables the comparison of performance across different models based on the ratio of correct predictions.

In this case, accuracy serves as an appropriate and well-rounded metric to assess the model's performance and compare it to other models.

e. Describing the result

The model has been tested on 100 images, including two types: Palm and Dorsal side of the hand. Here are the characteristics of these two types of images and the reasons for better performance with Palm images:

Palm Images:

Characteristics: Palm images typically show the front side of a hand, where the palm is exposed. These images often have clear and distinguishable features, such as fingers, palm lines, and nails.

Better Performance: The model performs better on Palm images due to the well-defined features present in these images. The distinct landmarks on the palm, fingers, and nails provide more salient points for the model to detect, making it easier to recognize and track the hand accurately.

Dorsal Side of the Hand Images:

Characteristics: Dorsal side images show the backside of the hand, which generally has fewer prominent features compared to the palm. They often lack distinctive landmarks found on the palm side.

Performance Challenge: The model may have a slightly harder time with Dorsal side images because of the relative lack of discernible features. The absence of palm lines, fingernails, and detailed hand contours can make it more challenging for the model to detect and track the hand correctly.

5 DISCUSSION

The model demonstrates optimal performance when recognizing the palm side of the hand under well-illuminated conditions, approaching near-perfection. During the testing phase, involving over 100 samples categorized into two primary labels: 'Dorsal side of the hand' and 'palm,' we obtained near-flawless results for palm recognition. However, when working with 'Dorsal side of the hand' samples, the success rate decreased to 79 percent.

Testing Case	No. of Tests	No. of Correct	Accurracy(%)
Palm	70	70	100
Dorsal side of the hand	70	55	79
AVG			89,5

Figure 6: The Accuracy of Data

The results indicate that the model performs effectively and reliably with real-time images from the camera. This is crucial for applications such as recognizing gestures of individuals with disabilities, controlling robotic arms, or interacting with computer keyboards. These are pressing and practical societal needs.

Figure 7: Enter Caption

The model exhibits several advantages and disadvantages. On the positive side, it performs impressively in well-lit conditions and demonstrates robust recognition of the palm side, achieving high accuracy in most cases. Furthermore, the model operates smoothly, is lightweight, and remains stable when all image conditions are met. However, there are some drawbacks. The model's performance significantly deteriorates in low-light environments, and it struggles with the recognition of the 'Dorsal side of the hand'. Moreover, it lacks mobility as it has not yet been implemented on web or smartphone platforms.

6 SUMMARY

We propose a straightforward solution to the intricate challenge of hand recognition. Historically, this problem has posed significant difficulties. However, our suggested approach introduces an efficient and rapid resolution. CVzone excels in projects requiring real-time image processing and hand gesture recognition, particularly in well-lit conditions. However, it faces notable limitations. CVzone's performance deteriorates significantly in low-light scenarios, and its lack of mobility restricts its use on web platforms and on resource-constrained devices such as Raspberry Pi computers.These limitations highlight the crucial areas that necessitate focused attention and development. Envisaging the future, CVzone holds the potential to serve as a proficient tool for a wide spectrum of projects, thanks to its streamlined and dependable nature.In the pursuit of expanding upon this concept, we recognize the need for further research and development. Addressing CVzone's low-light performance and optimizing its compatibility with web and resource-constrained platforms will be essential. Additionally, exploring applications beyond hand recognition, such as gesture-based control systems, and enhancing the user experience for individuals with disabilities are promising directions. This progress in hand recognition not only showcases the evolution of technology but also underscores its capacity to address real-world challenges and societal needs. In the journey ahead, we anticipate witnessing the continued growth and refinement of CVzone as it transforms into an even more versatile and accessible solution, contributing to a wide array of practical applications in the realms of technology and social inclusion.

7 ACKNOWLEDGEMENT

We would like to express our gratitude to the lecturer and students of the CPV301 class for their invaluable guidance and support throughout the process of writing this paper. We also extend our appreciation to those who provided data support, which was instrumental in conducting this research.

REFERENCES

[1] Muhammad Yeza Baihaqi, Vincent, and Joni Welman Simatupang. Real-time hand gesture recognition for humanoid robot control using python cvzone. pages 262–271, 2022.

[2] Chetana D. Patil, Amrita Sonare, Aliasgae Husain, Aniket Jha, and Ajay Phirke. Survey on: Hand gesture controlled using opencv and python. *International Journal of Creative Research Thoughts (IJCRT)*, 2022.

[3] George Sung, Kanstantsin Sokal, Esha Uboweja, Valentin Bazarevsky, Jonathan Baccash, Eduard Gabriel Bazavan, Chuo-Ling Chang, and Matthias Grundmann. On-device real-time hand gesture recognition. *arXiv preprint arXiv:2111.00038*, 2021.

[4] Rupesh Kumar, Ashutosh Bajpai, and Ayush Sinha. Mediapipe and cnns for real-time asl gesture recognition. *arXiv preprint arXiv:2305.05296*, 2023.

[5] Facundo Manuel Quiroga, Franco Ronchetti, Ulises Jeremías Cornejo Fandos, Gastón Gustavo Ríos, Pedro Alejandro Dal Bianco, Waldo Hasperué, and Laura Cristina Lanzarini. A comparison of small sample methods for handshape recognition. *Journal of Computer Science & Technology*, 23, 2023.

[6] Shubhra Aich, Jesus Ruiz-Santaquiteria, Zhenyu Lu, Prachi Garg, KJ Joseph, Alvaro Fernandez Garcia, Vineeth N Balasubramanian, Kenrick Kin, Chengde Wan, Necati Cihan Camgoz, et al. Data-free class-incremental hand gesture recognition. pages 20958–20967, 2023.

[7] Alberto Sabater, Iñigo Alonso, Luis Montesano, and Ana C Murillo. Domain and view-point agnostic hand action recognition. *IEEE Robotics and Automation Letters*, 6(4):7823–7830, 2021.

[8] Maarten Vandersteegen, Wouter Reusen, Kristof Van Beeck, and Toon Goedemé. Low-latency hand gesture recognition with a low-resolution thermal imager. pages 98–99, 2020.