iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Digital Handwriting Recognition Using Hand Tracking by Using Media Pipe and OPENCV Libraries

Mr. M Nagaraju[1], MD Sohail[2], B Rahul Teja[3], N S Siddharth[4]
[1]*Assistant Professor, Information Technology Department, SreenidhiInstitute of Science and Technology, Yamnampet, Hyderabad*
[2, 3, 4]*Bachelor of Technology IV[th] year,IT Department, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad*

Abstract: ***While in taking online classes or in day-to-day activities writing is more important task, as we are living in digital age which means most of our tasks are been shifted to digital trends such as like paying bills, booking tickets, emotion tracking etc., among this writing is at most important task weather you share ideas plan for a particular task taking notes, in this data age stats show humans more often prefer to write and use about 100 pages per average. It is quite our instinct to understand and represent any information which is written in handwriting, but when it comes to digital world though they are gesture based, touch based methods they couldn't match the natural hand-write representation. So, we come with solution for this problem for some extent by using hand tracking and hand landmark position representation which is offered by media pipe library form google. This technique uses image processing (OpenCV) to capture video from webcam which will be given to our python application which uses media pipe library to recognize hand and represents it as 0-20 landmark positions. Which will be used to find index finger for writing and use two fingers for removing or erasing pre-written information. This type of application is very useful in a day to day basic activates for students, teachers and makers in order to easily plan and organize projects, notes etc., Though this technique neither match a real writing experience nor the writing tablet for computers and smartphones which can translate our writing into digital notebook, it does not require any additional equipment and can be easily used to if practiced. This type of application can improvise the understanding by allowing users to explain in a whiteboard like environment in online education.***

Keywords: Open Cv, Media Pipe, Hand Tracking, Landmarks.

I. INTRODUCTION

A. Motivation

Hand detecting and gesture tracking have become commonplace, presenting a wide range of break-throughs and constraints. As a result of the growing interest in computer vision, accessible technology that can enable new advancements in AI is rapidly developing and improving. We discuss, implement, and review the problems and upcoming opportunities in the realm of human user interaction and VR in this documentation. In light of the COVID-19 outbreak, the objective for this research is to limit human interaction and dependency on gadgets to operate systems. These findings could lead to more research and, in the long run, help people use virtual worlds more effectively. Bluetooth and wireless technologies are becoming increasingly accessible, thanks to recent breakthroughs in virtual reality and their implementation in our daily lives. This research offers a visual AI system that uses hand motions and hand tip acquisition to perform mouse, keyboard, and stylus activities using computer vision. Instead of using typical headsets or external devices, the proposed system uses a web camera or built-in camera to track finger and hand movements to process the computer. This solution may be removed indefinitely due to its simplicity and effectiveness. the utilisation of additional hardware, batterydurability, and, in the end, user convenience

B. Problem Statement

Build a Digital Handwriting Recognition using Hand Tacking by using media pipe and OpenCV libraries. The hand tracking is done by hand landmark position representation which is offered by media pipe library form google. We use image processing techniques which are offered by OpenCv to capture video from webcam which will be given to our python application which uses media pipe library to recognize hand and represents it as 0-30 landmark positions. Which will be used to find index finger for writing and use two fingers for removing or erasing pre- written information.

C. Project Objectives

Using hand detection module we need to develop a program by which following aresatisfied

1) By using single, one should be able to draw or write
2) By using two fingers, one should be able to erase the content on the screen

D. Project Report Organization

This book contains six chapters. The first chapter contains motivation, problem statement and project objectives. The second chapter includes the Literature survey which includes existing work and limitations of existing work. The third chapter includes specifications, software and hardware requirements needed for the project. The fourth chapter contains UML diagrams, Technology Description and Proposed methods.

The fifth chapter includes Implementation which contains the technologies used for developing the application and code snippets. The fifth chapter also contains test cases and screenshots of the applications. The sixth chapter investigates the future enhancements and conclusion of the project.

II. LITERATURE SURVEY

A. Existing Work

This method uses touch data from the camera to generate photos. The vision-based method places a strong emphasis on touch-captured images and draws attention to the most prominent and recognised feature. At the outset of the vision- based approach, colour belts were utilised.

The standard colour that had to be applied to the fingers was the method's biggest drawback. Then, instead of using colourful ribbons, use your hands. This is a difficult difficulty because real-time performance necessitates a background, continuous lighting, personal frames, and a camera. Furthermore, such systems must be designed to meet certain criteria, such as accuracy and resilience.

Theoretical analysis is the most difficult to learn because it is dependent on how humans perceive information about their surroundings. Several other approaches have been attempted so far. Create a three-dimensional model of a human hand as the initial phase. The palm form and combined angles parameters are derived by comparing the model to hand photos acquired by one or more cameras.

The touch phase is then created using these parameters. The second method is to use the camera to snap a picture and extract particular traits, which are then used as input in the partition to divide the data.

1) *ClayAIR:* Its hand tracking solutions aim at higher performance, quicker implementation time and higher accuracy. It can predict 22 3D key point coordinates. Using regression algorithms trained on 1.4 million images including real and synthetic images. It is being used by some leading tech giants like Enovo, Nreal, Qualcomm to bring virtual reality to the digital world.
2) *SOTA Hardware:* Data Gloves: Data gloves are pure VR devices in the sense that it can detect activity of the joints and on the other hand the feedback enables the user to feel the virtual targets in a pseudo-physical sense. They are especially famous in the VR field since they are highly accurate and the inference time is less. Additionally, they are a great way to collect data of hand-landmarks for machine learning models. But since photoelectric sensors and position trackers are costly, the production and maintenance of these gloves is also high.
3) *Inertial Sensors:* The Nintendo Wii was the commercial release of inertial sensors. They are composed of an actuator and a sensor which help to collect and obtain information about gestures. Built with an accelerometer and an infrared sensor, they can capture the user's wrist and arm gestures.
4) *KCF:* KCF algorithm or Kernelized correlation filter algorithm is mainly focused around creating large number of training examples by shifting the target area in a circular manner. It was widely used or object tracking and is the base of many recent tracking algorithms. Unfortunately the algorithm doesn't perform well in case of scale variations i.e changes in the size of target objects. Additionally it is not easy to train it for detecting multiple landmarks

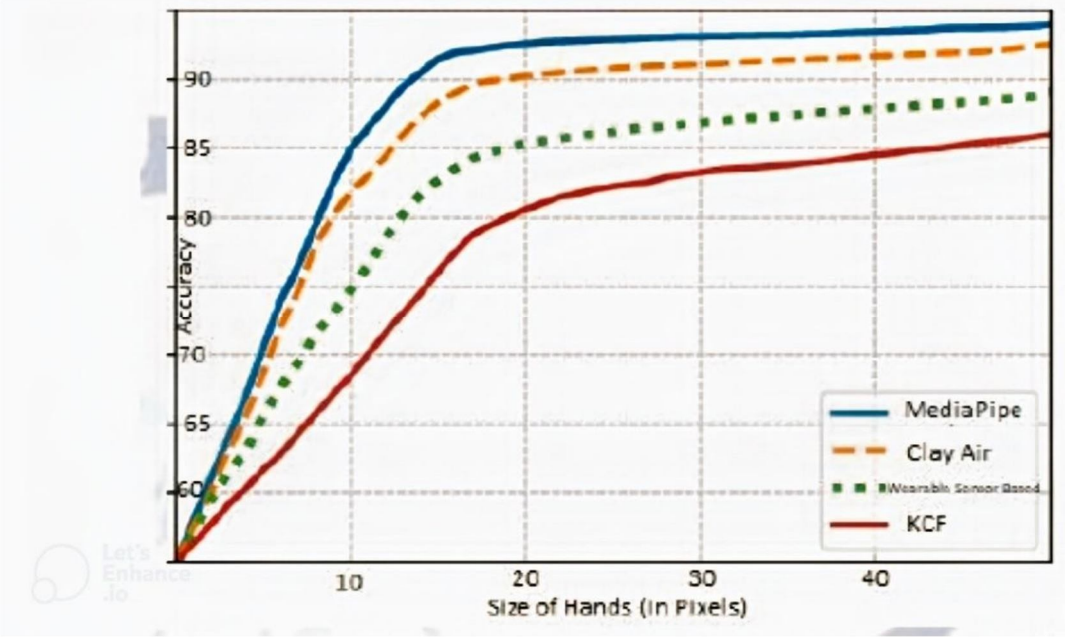


Figure 1 comparision of existing works

B. Limitations Of Existing Work

Bluetooth and wireless technologies are becoming increasingly accessible, thanks to recent breakthroughs in virtual reality and their implementation in our daily lives. This research offers a visual AI system that uses hand motions and handtip acquisition to perform mouse, keyboard, and stylus activities using computer vision. Instead of using typical headsets or external devices, the proposed system uses a web camera or built-in camera to track finger and hand movements to process the computer. This solution may be removed indefinitely due to its simplicity and effectiveness. the utilisation of additional hardware, battery durability, and, in the end, user convenience

The Python programming language and the OpenCV computer library were used to create the AI mouse programme. The proposed visual AI mouse system makes use of the Media Pipe module to detect palms and titles, as well as the Pynput, Autopy, PyGames, and PyAutoGUI libraries to browse the monitor and perform actions like left-click, right-click, and scroll. The suggested model's findings demonstrate a huge amount of precision, and the introduced design can operate efficiently in practical applications with only a CPU and no GPU.

III. PROPOSED SYSTEM

Recognition of hand shape and motion can help improve user experience across a wide range of technological disciplines and platforms. It can be used to understand sign language and control hand movements, as well as to enable the overlay of digital content and information on top of the physical world in augmented reality. Because hands regularly occlude themselves or each other (for example, finger/palm occlusions and hand shaking), and because they lack high contrast patterns, robust real-time hand perception is a difficult computer vision challenge. Artificial Intelligence (Al, a wide term describing a collection of advanced methodologies, tools, and algorithms for automating the execution of diverse tasks) has infiltrated virtually every corporate activity over the years. Hand tracking is one of the most prominent AI solutions; it is used to estimate the position and orientation of a person's hand given an image of them.

Your brain is programmed to accomplish all of this naturally and instantaneously as a human being. In fact, humans are overly adept at recognising faces, resulting in the appearance of faces in common items. Because computers areincapable of such high-level generalisation, we must teach them each step of the process separately. We need to create a pipeline in which each phase of face recognition is solved separately and the result of the current step is passed on to the next. To put it another way, we'll chain together a number of machine learning algorithms.

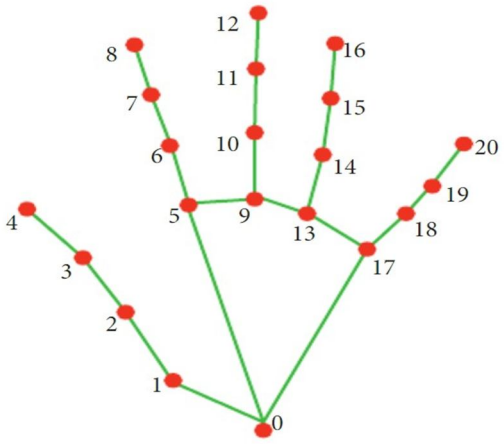
Because most solutions rely on key points and heat maps, we must first collect pose alignment data for each position. We can investigate several test cases in which the entire hand is displayed and key spots for the various hand portions can be detected. We can employ occlusion mimicking augmentation to ensure that the hand tracker can perform in high occlusions, which are distinct test cases than typical ones. There are 30000 real-world photos in the training data set, each with 21 3D coordinates.

Although it may appear that employing gradients rather than pixels is a randomoption, there is a valid reason behind it. If we look at pixels closely, we can notice that the pixel values of very dark and very light images of the same person are radically different. If you only consider the direction in which brightness varies, however, both completely black and extremely bright images will have the same accurate representation. This makes it much easy to resolve the issue. Keeping the gradient for each and every pixel, on the other hand, provides way too much information. The forest is overshadowed by the trees, and we lose sight of the woodland. To see the image's underlying pattern, it would be preferable if we couldsimply view the basic flow of lightness/darkness at a higher level.

The histogram of oriented gradients (HOG), a feature descriptor commonly used in computer vision, is used to detect objects. It works by measuring the number of times a gradient orientation appears in a specific area of an image. Edge orientation histogram, scale invariant feature transform descriptor, and shape contexts are all methods that are similar.

In our photograph, we isolated the hand. But now we have to cope with the fact that a computer sees a hand turned in different directions in a completely different way. To compensate for this, we'll try to warp each image so that the fingers are always in the sample location.

Following palm detection over the entire image, our next hand landmark model uses regression to accomplish exact keypoint localization of 21 3D hand-knuckle coordinates within the detected hand regions, i.e. direct coordinate prediction. Even with partially visible hands and self-occlusions, the model develops a consistent internal hand posture representation.

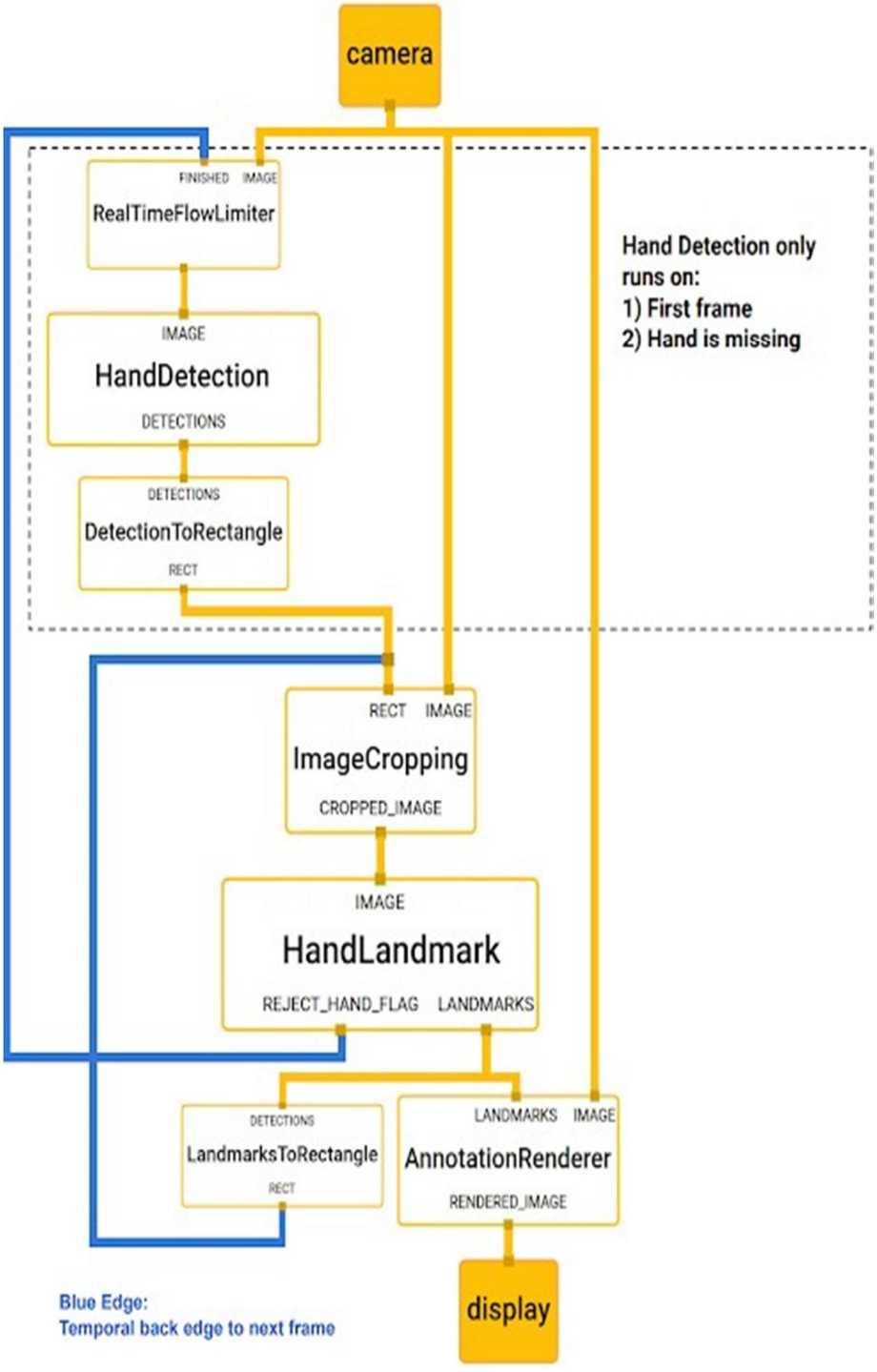


0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

IV. SYSTEM ARCHITECTURE

Our hand tracking solution employs a device adaptable pipeline that mixes the following 2 designs: • hand tracker that finds hands on a full input image using an aligned hand bounding box. • palm marker design which generates using the palm detector's clipped data to create high-fidelity 2.5D landmarks palm bounding box. Giving an appropriately palm picture edited for the hand landmark model considerably decreases the need for data enhancement. (for example, circular movements, translations, & scaling) and lets the matrix to concentrate almost all of its energy on finding landmarks with high precision. We employ a bounding box created from the earlier frame's marker assumption as a source of ongoing frame in a practical detecting scenario, preventing requirement to use the tracker on each frame.

Rather, only the first frame or when the hand prediction indicates that the hand has been lost is the detector used.`



V. HAND LANDMARK MODEL

The palm marker design employs regression to achieve accurate marker placement of 21 2.5D points in the discovered palm areas after hand tracking across the full image. The design produces even with partially visible hands and self-occlusions, a continuous internal hand position depiction. The model produces 3 results (see Figure 3): 21 x, y, and relative depth palm landmarks

Palm flag that shows the chance of the presence of a palm in the input image. Left or right handedness is a classification of handedness.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

For the 21 landmarks, we apply the same structure as [14].. Only synthetic photos are used to determine the relative depth with respect to the wrist point, as described.The 2D points are learnt from both practical and artificial data containers. To compensate for tracking failure, we constructed a new model output, same as [8,] that calculates chance of finding a hand that is adequately aligned in the given field.If number(score) goes below certain level, the tracker is activated, and tracking is restarted. One more significant factor in excellent AR/VR hand interaction is handedness.

It is useful in situations where individual hand does its own set of tasks. Because of this, a binary classification head was constructed to detect whether the input hand was left or right. Our system is optimised for real-time mobile GPU inference, but we've also built lighter and heavier implementations to handle Higher accuracy and CPU inference on mobile devices without GPU support requirements for desktop application, accordingly.

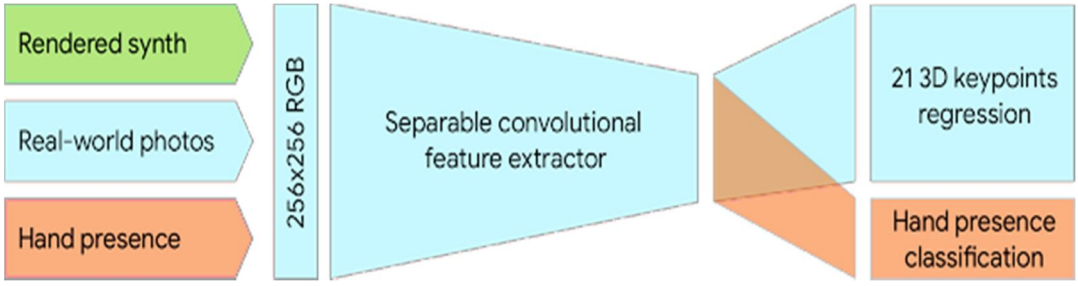


Figure 6 Hand landmark model

VI. DATASET AND ANNOTATION

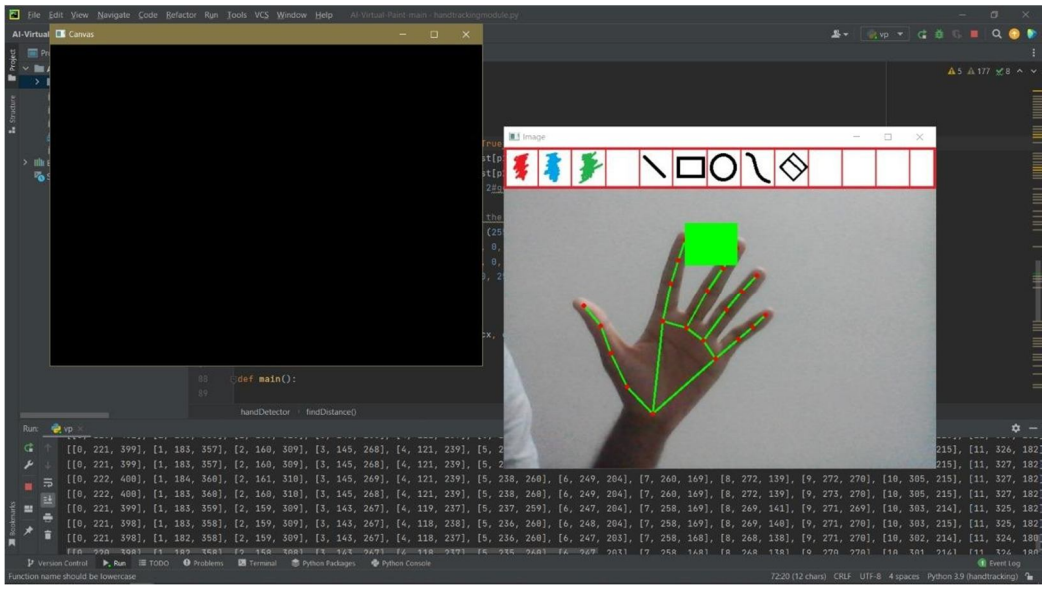
We generated the following datasets to collect ground truth data, each addressing adistinct part of difficulty faced:

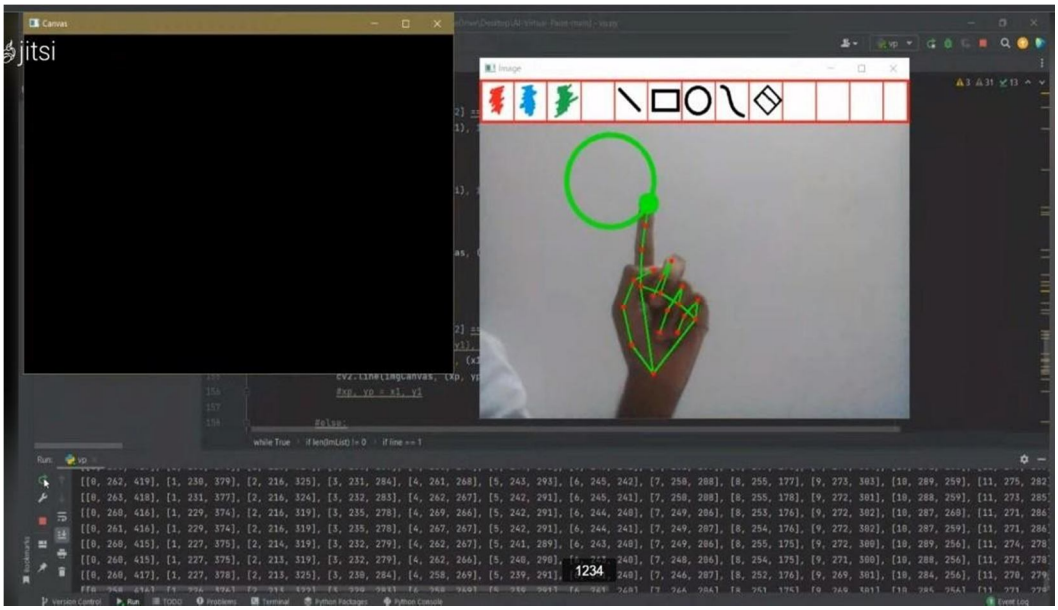
1) *The Raw Collection:* consists 6 thousand photos with a wide range of characteristics, such as geographical diversity, lighting circumstances, and palm look. This data container's drawback is that it does not have complicated hand articulation.
2) *In-house Gesture Dataset:* This data container consists of 10K photos that covers all conceivable hand movements from various angles. The dataset's shortcoming is that it was compiled from only 30 persons with little variety in their backgrounds. The in-the-wild and in-house datasets are excellent complements for improving robustness.
3) *Synthetic Dataset:* To further cover all probable hand poses and provide additional depth supervision, we render a high-quality synthetic hand model over various backgrounds and map it to the required 3D coordinates. To control the thickness of the fingers and palm, we use a commercial 3D hand model with 24 bones and 36 blendshapes. There are five different skin tones and textures included with the figurine. We created changing video sequences of hand postures and sampled 100K pictures from the videos. Three different cameras and a random high-dynamic-range lighting situation were used to produce each position.

The palm detector only uses the in-the-wild dataset because it is sufficient for hand localisation and has the largest variation in appearance. The hand landmark model, onthe other hand, is trained using all datasets. We employ projected groundtruth 3D joints for synthetic images and annotate realworld images with 21 landmarks. We select a subset of real-world images as positive examples for hand presence, omitting annotated hand regions as negative examples. We tag a subset of real-world images with handedness annotations to provide such data for handedness.

VII. RESULT

Human hand detection from video is important in a variety of applications, including quantitative sign language recognition and detecting handwriting. It can be used to support sign language translation, gesture recognition, and gesture control, for example. In augmented reality, it can also enable the overlay of digital content and information on top of the physical world. Media Pipe Pose is a machinelearning solution for high-fidelity hand pose tracking that uses our Blazepalmresearch, which also drives the ML Kit hand Detection API, to infer 21 3D landmarks on the entire hand from RGB video frames. The Open-CV library has a built-in solution for interacting with a streaming device, capturing a video stream, and generating video frames. The Open-CV Video Capture library can be used to accomplish this. This library is capable of reading video frames and displaying them in a window. BGR format frames were retrieved from Open-CV. As a result, we first convert it to RGB. We can use Media Pipe's hand on video frames to track hand posture once we get our video frames in RGB.





VIII. CONCLUSIONS & FUTURE SCOPE

The visual AI mouse system's main purpose is to allow users to manage mouse cursor functions with a palm sensation rather than by manipulating objects. Hand gestures and hand tips are detected and processed by the suggested system, which can be accessed via a webcam or a built-in camera.

We may deduce from the model's findings that the proposed artificial intelligence system has performed very well and is incredibly accurate when compared to current models, and that the model addresses the majority of the existing system's constraints.

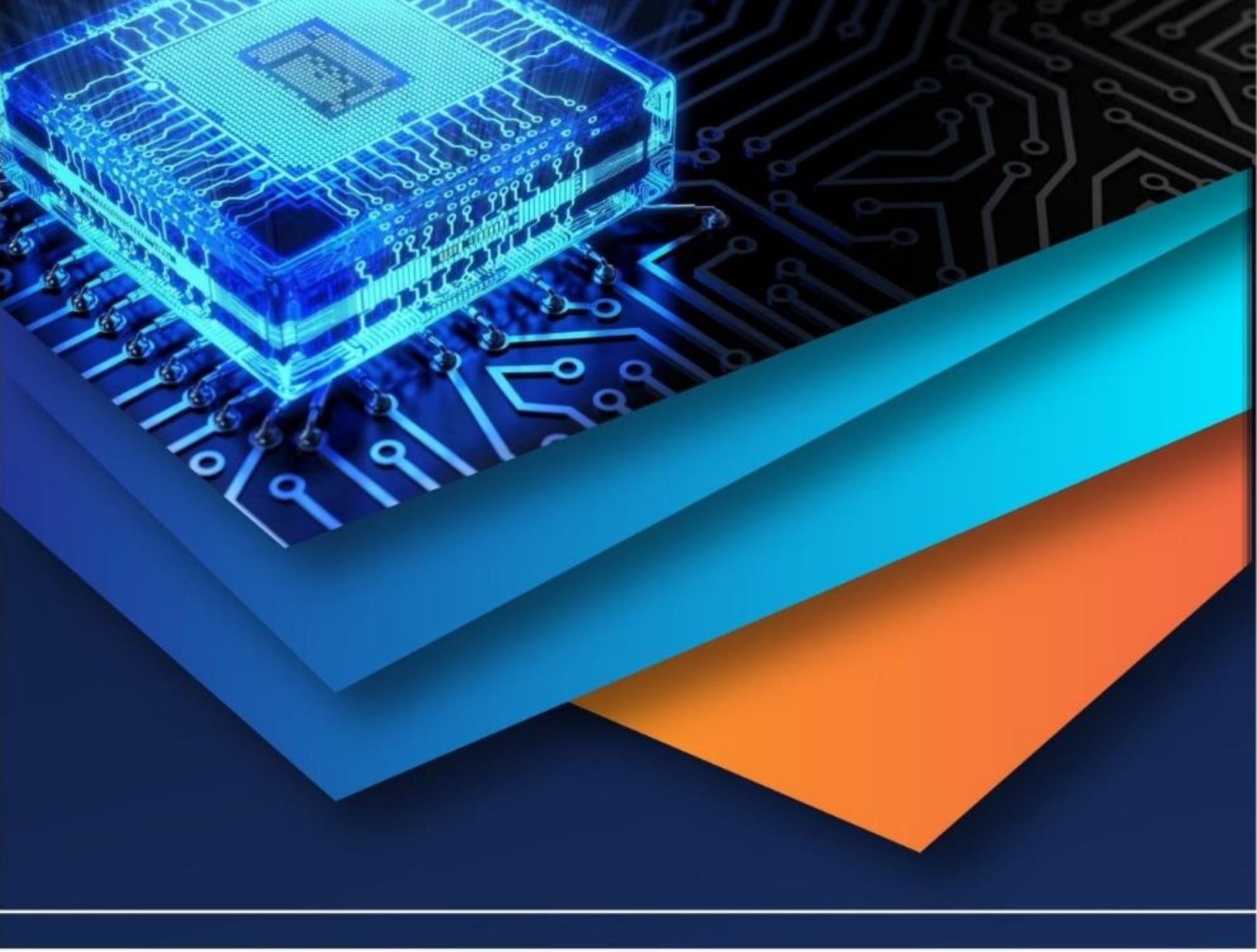
Because it is so exact, the introduced model can also be used in practical applications.

The model has some flaws, such as a slight loss of precision when using the right-click feature and the difficulty of selecting text by clicking and dragging.As a result, we'll work to overcome these limitations by developing a fingerprint acquisition process that will produce more accurate results in the future**.**

This project can be improvised by adding Neural Network for high accuracy.

REFERENCES

[1] https://arxiv.org/pdf/2006.10214.pdf
[2] https://www.researchgate.net/publication/342302340_MediaPipe_Hands_On-device_Real-time_Hand_Tracking
[3] https://kulinpatel.com/real-time-writing-with-fingers-on-web-camera-screen-opencv/
[4] https://www.researchgate.net/publication/357622313_Virtual_Control_Using_Hand_-Tracking
[5] Rao, A.K., Gordon, A.M., 2001. Contribution of tactile information to accuracy in pointing movements. Exp. Brain Res. 138, 438–445. https://doi.org/10.1007/s002210100717
[6] Masurovsky, A., Chojecki, P., Runde, D., Lafci, M., Przewozny, D., Gaebler, M.,2020. Controller-Free Hand Tracking for Grab-and Place Tasks in Immersive Virtual Reality: Design Elements and Their Empirical Study. Multimodal Technol.Interact.4,91.https://doi.org/10.3390/mti4040091.
[7] Lira, M., Egito, J.H., Dall'Agnol, P.A., Amodio, D.M., Gonçalves, Ó.F., Boggio, P.S., 2017. The influence of skin colour on the experience of ownership in the rubber hand illusion. Sci. Rep. 7, 15745.https://doi.org/10.1038/s41598-017- 16137-3.
[8] Danckert, J., Goodale, M.A., 2001. Superior performance for visually guided pointing in the lower visual field. Exp. Brain Res. 137, 303–308. https://doi.org/10.1007/s002210000653.
[9] Carlton, B., 2021. HaptX Launches True-Contact Haptic Gloves For VR And Robotics. VRScout. URL https://vrscout.com/news/haptx-truecontact-haptic-gloves-vr/ (accessed 3.10.21).
[10] Brenton, H., Gillies, M., Ballin, D., Chatting, D., 2005. D.: The uncanny valley: does it exist, in: In: 19th British HCI Group Annual Conference: Workshop on Human-Animated Character Interaction.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY