## 1. Abstract

The use of a physical controller like mouse, keyboard for human computer interaction hinders natural interface as there is a strong barrier between the user and computer. In this paper, we have designed a hand gesture recognition system which can efficiently track both static and dynamic hand gestures. The proposed system translates the detected gesture into actions such as using a whiteboard on the video screen itself and other such features to help the user. The dynamic gesture is used to shuffle through the slides in a presentation. The result obtained shows that an intuitive HCI can be achieved with minimum hardware requirements.

## 2. Introduction

Gesture recognition is an important part of computer science with the aim of understanding human gestures through algorithms. Computer vision-based gesture recognition enables people to communicate more naturally with machines. Interactive presentation systems use advanced Human-Computer Interaction (HCI) techniques to provide a more convenient and user-friendly interface for giving commands to the machine, such as page up/down controls, open menu, exit app, etc. Compared with traditional mouse and keyboard control, the experience is significantly improved with these techniques. The advantage is that it is less affected by the environment.

Users can interact with the computer at any time, and it has less constraint on users, enabling computers to accurately and timely understand users' instructions. The instructions do not require any mechanical assistance making the whole process efficient in overall manner. Gestures are timely, vivid, intuitive, flexible and visual in the process of human-computer interaction. They can soundlessly complete interaction and successfully break the gap between reality and the virtual world. The introduction of gesture recognition into the online teaching world would open gates of new opportunities for students and teachers making the whole experience more interactive and effective.

Input units have been a major part of computing machines, however, technology has much more to offer. The proposed project believes in the idea of making the process of giving inputs to machines more time efficient as well as an effective process that will decrease the gap between the input and output cycle providing a more user-friendly interface that is also capable of customization for special needs of the situation as well as the user.

## 3. Literature survey

Early gesture recognition directly detected the position of each joint of the human hand with Wearable devices, and transmitted the information to the computer through wired transmission, accurately storing the user's hand motion information. It doesn't mean that these devices are the most convenient for most users. Since ancient times, gestures have been used for communication and interaction. Before the invention of language, people could easily express the idea by gestures.

Data Gloves and other equipment are expensive and inconvenient to use. Data Glove was displaced by the optical marking method because it used IR light to detect the position and hand movement of the human hand. It also has a good effect, but still requires more complicated equipment. Although higher accuracy can be obtained by the help of external devices, it is expensive and affects the user's actions to some extent. The gesture recognition based on computer vision refers to the processing of the video data collected by the camera through the gesture recognition algorithm, which achieves the purpose of gesture recognition and has become a research hotspot in recent years.

In the past decades the social life and computer interactions has been a huge breakthrough and complying to find something more specific to each functionality as the interactive components/ devices(smartphones, games, infotal inment systems) has been upgrading and complexed, has been an important issue. In these times gesture control techniques has become a new development trend in every possible computer-based interactions such as games for instance; The VR headsets and other such devices such as data gloves and motion sensors; Through these equipment or such techniques user can interact with the machines based on the movements of these devices more naturally than using a traditional devices such as joysticks or keyboards or a mouse.

Domestic computer vision-based gesture recognition has also achieved certain achievements. The proposed method of gesture recognition using apparent changes in image transformation, and used variational parameter model of image motion to identify 120 gestures. There has been another method proposed to solve dynamic gesture recognition through self-learning sparse representation. This method directly processes the original image without feature extraction and real-time processing and other such methods using fingers detection through SVMs to narrow down the possible number of inputs.

## 4. Existing System

In the era of modern technology, gesture recognition is spreading at a fast pace and is being implemented in various systems using different techniques. The majority of the early method involved wearable devices that consisted of motion sensors and would detect the movement of the hand and body pose respectively and send the data to the machine as input. Data gloves, stylus, joysticks, IMUS, etc are widely used devices for the purpose of gesture recognition.

The accuracy obtained by such devices is higher however, these devices create inconvenience for the user as they are rather expensive, bulky, not as user-friendly and doesn't support portability.

In a world where everyone is shifting towards the metaverse, there is a demand for easy human-computer interaction. Gesture recognition has already become very prominent in the gaming world, it would only rather make sense to implement the technique in other categories including day-to-day tasks such as online teaching, presentation, medical purposes, security, etc.

***Fig 1*** *: Bodysuit for controlling gesture*

## 5. Proposed System

Data gloves are a popular method used to perform similar activities however the proposed projects discourage the use of the same to minimise usage of multiple units. The removal of data gloves from the project is supported by the well known framework proposed by Google named MediaPipe. MediaPipe is a cross platform pipeline framework to built custom machine learning solutions for live and streaming media. The framework was open sourced by Google and is currently in the alpha stage.

The process of online teaching, office meetings, presentations, etc will become a lot easier if the minimum input required would just be the user and some room light. The proposed project aims in minimising the hassles listed above by introducing gesture and body recognition as a major method of providing commands to the computer.

The common problem of writing on online boards faced by users is solved by simply providing the option of writing directly on the screen by using the user's mere finger as a pointer and palm as an eraser. This makes the whole teaching experience more time sufficient, user friendly, interactive and productive

The gesture recognition method also makes multitasking easier. Eg- a user can drag an image on the screen, while resizing it accordingly and can also write or erase on the screen/ image at the same time. This process is not limited to images and can be used simultaneously with pdf files, slideshows, word files, etc.

This method removes the responsibility from the user to own the appropriate input units and use them correctly to give commands. By using the gesture recognition methods, the user can give commands in time to the machine and doesn't really have to worry about learning to use the various input units and wait for them to ask the machine to perform a certain task.

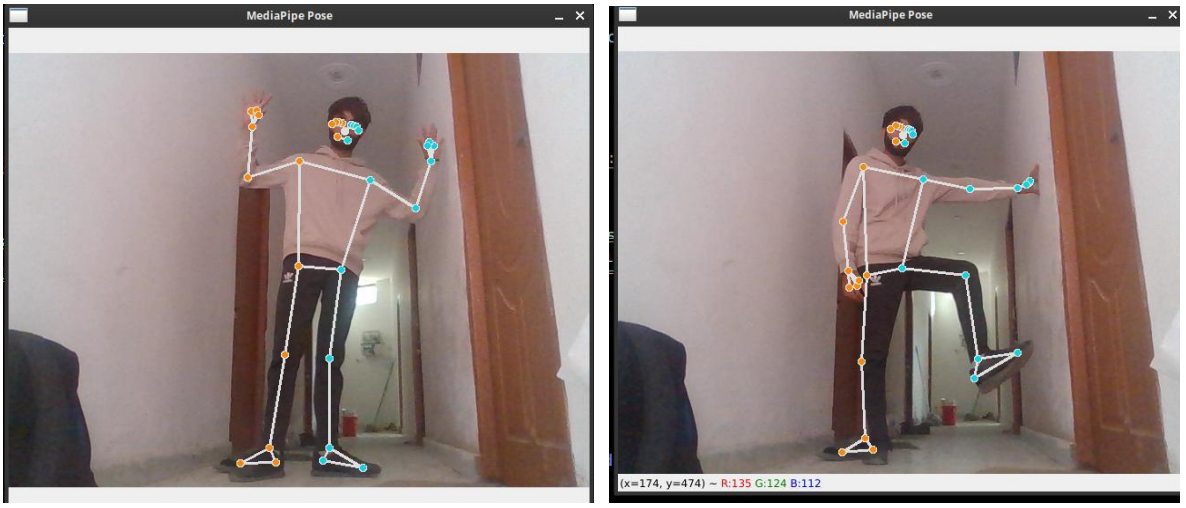
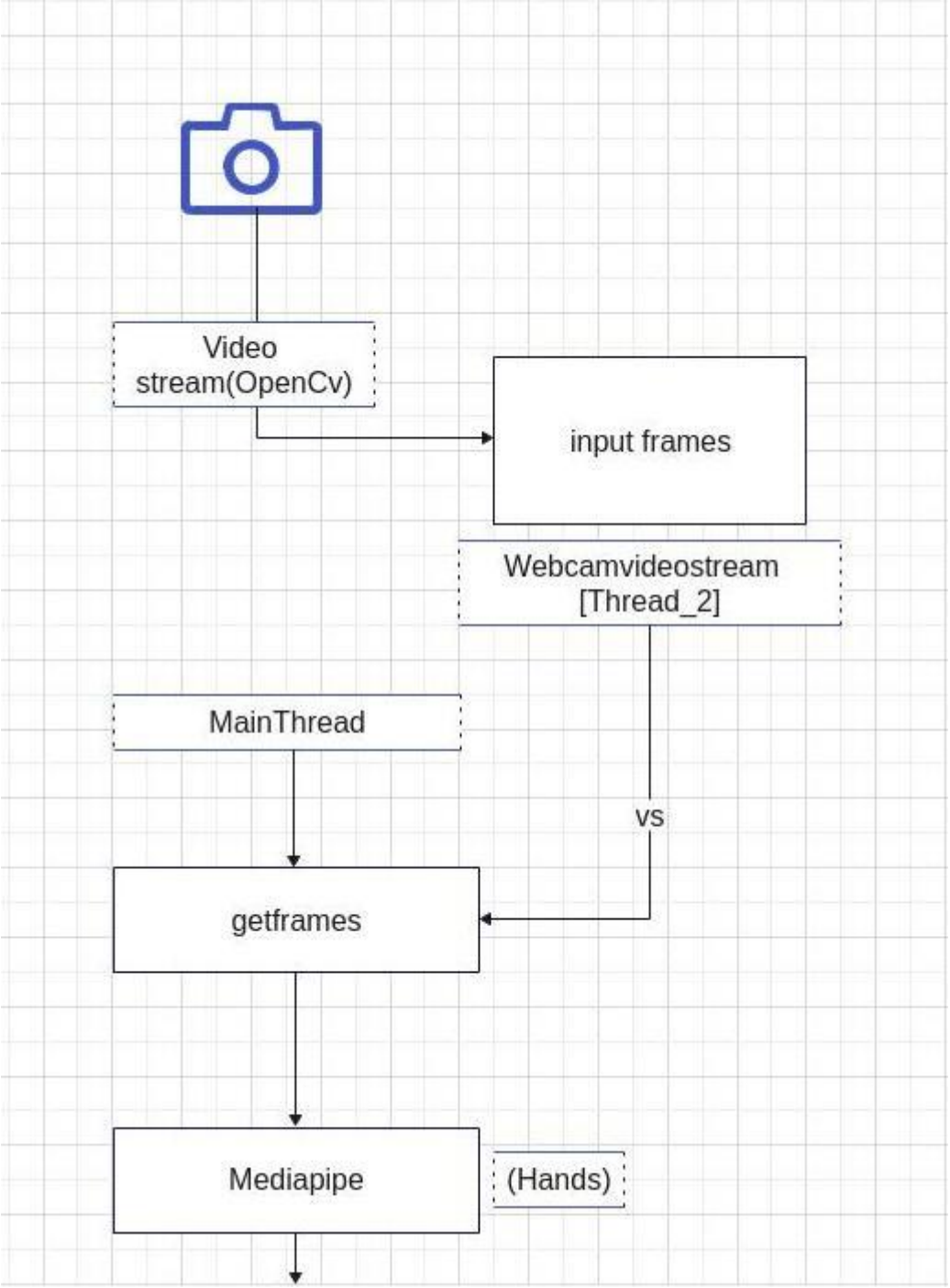



***Fig 2:*** *Body-pose detection using mediapipe*

6. **Design**

The interface of the proposed project is designed in a way to provide maximum user friendly features. It is highly customizable, portable and easier to control. Common problems such as, lagging of video being streamed, unavailability of input devices, a long list of system requirements, etc has been minimised in the following project. The efficiency and accuracy mentioned is achieved by the combination of MediaPipe and OpenCv.



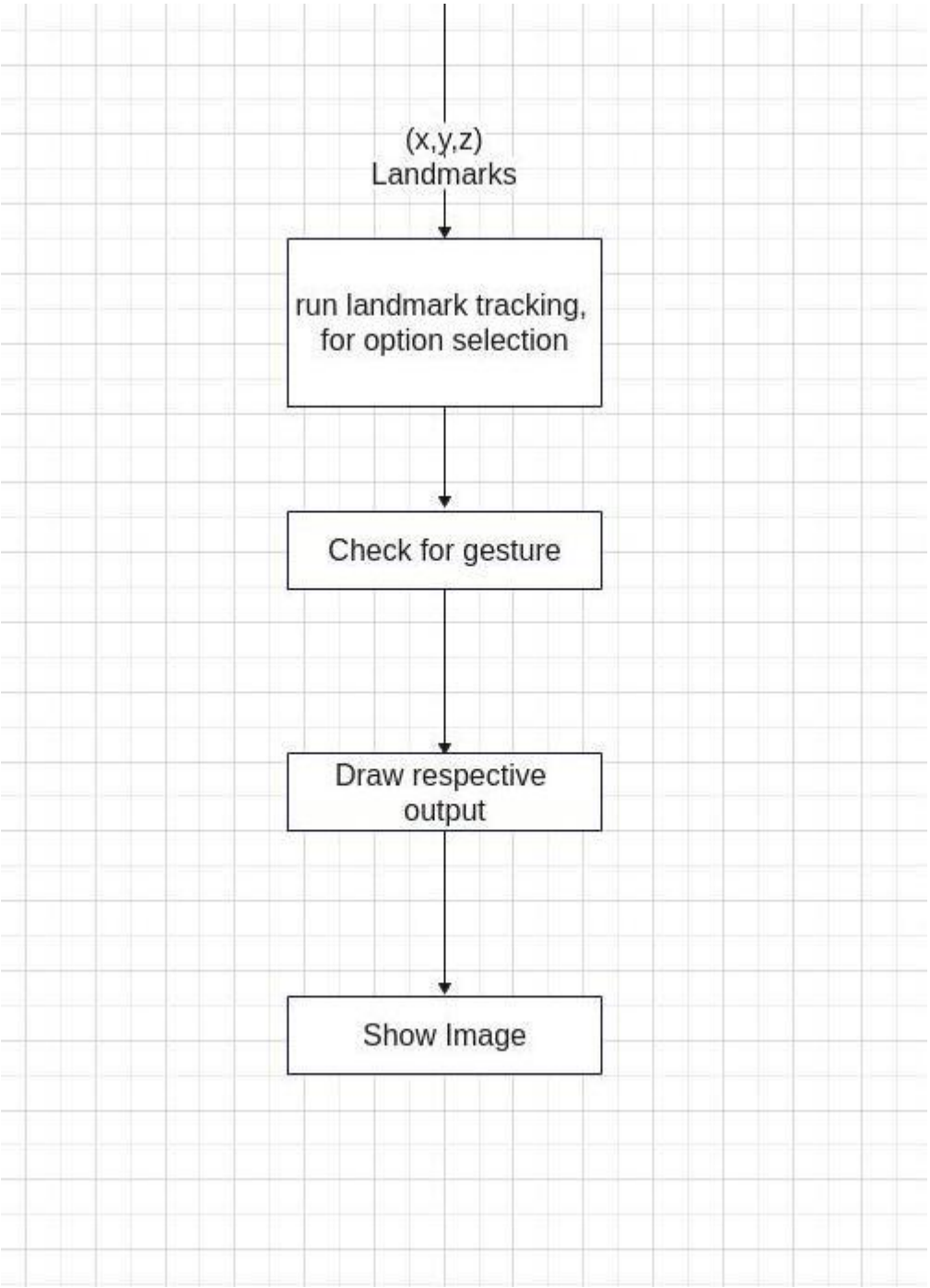



***Fig 3:*** *Design Diagram of proposed system*

**The user interface :**

The menu consists of 4 Options including:

- Pointer
- Painter
- Colour option
- Eraser


***Fig 4:*** *Screenshots of user guide application.*

## 7. Methodology

The model of this project can be describe into different phases:

***1. Extracting video frames using OpenCV***

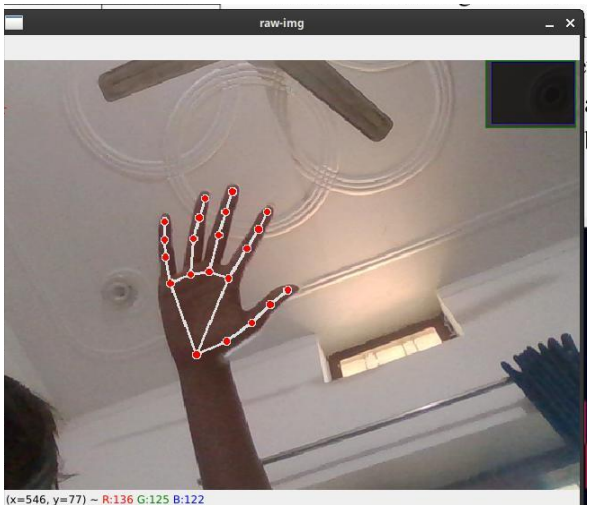
First importing OpenCV, it will access the camera and start taking the video frame by frame. Now the variable is storing the first frame and sending it to the thread made from Mediapipe so the it can read the frame using stream.read() function.
Here is how reading frame works:

***2. Reading the Frame using Mediapipe***

While reading the first frame it captured, mediapipe solution will detect the hand as this is a Well-trained framework. After that, it will make the landmarks on the palm, elbow, and shoulders and provide a defined name and number to that landmarks.

***3. Commands using conditions***

Now after storing the updated frame, conditions can be use to give suitable commands to the programme. These conditions and landmarks can be used to provide various outputs according to the user.




| Landmark key point | x | y | z |
|---|---|---|---|
| 0 | 0.19527593 | 0.6772005 | -7.258559e-05 |
| 1 | 0.263733 | 0.63610333 | -0.039326552 |
| 2 | 0.3196355 | 0.5412712 | -0.058143675 |
| 3 | 0.3613177 | 0.4677803 | -0.075389124 |
| 4 | 0.39756835 | 0.43665695 | -0.093960665 |
| 5 | 0.26121178 | 0.3753401 | -0.030742211 |
| 6 | 0.28375435 | 0.26732442 | -0.061761864 |
| 7 | 0.29418302 | 0.19642864 | -0.08401911 |
| 8 | 0.30149087 | 0.13136405 | -0.1029892 |
| 9 | 0.21288626 | 0.3534055 | -0.032817334 |
| 10 | 0.21505088 | 0.22275102 | -0.058613252 |
| 11 | 0.2152167 | 0.1385001 | -0.08102116 |
| 12 | 0.21389098 | 0.06872013 | -0.09661316 |
| 13 | 0.16952133 | 0.36720178 | -0.04239379 |
| 14 | 0.15782069 | 0.2474725 | -0.07075888 |
| 15 | 0.15325233 | 0.16784605 | -0.09313752 |
| 16 | 0.15079859 | 0.102125764 | -0.108897485 |
| 17 | 0.12903559 | 0.41147107 | -0.05464202 |
| 18 | 0.09621665 | 0.3332698 | -0.08286363 |
| 19 | 0.07500376 | 0.28210545 | -0.10173174 |
| 20 | 0.056006864 | 0.2304765 | -0.11720086 |

***Fig5:*** *Hand landmark coordinates*

## 8. Results

The interface designed allows user to have a more natural experience since, it minimises the devices being used and takes commands through human hand movements.

The whole experience becomes less time consuming as well as less complicated making the whole project efficient comparatively.

The features are easy to access and the user does not require much knowledge and practice to use the interface.

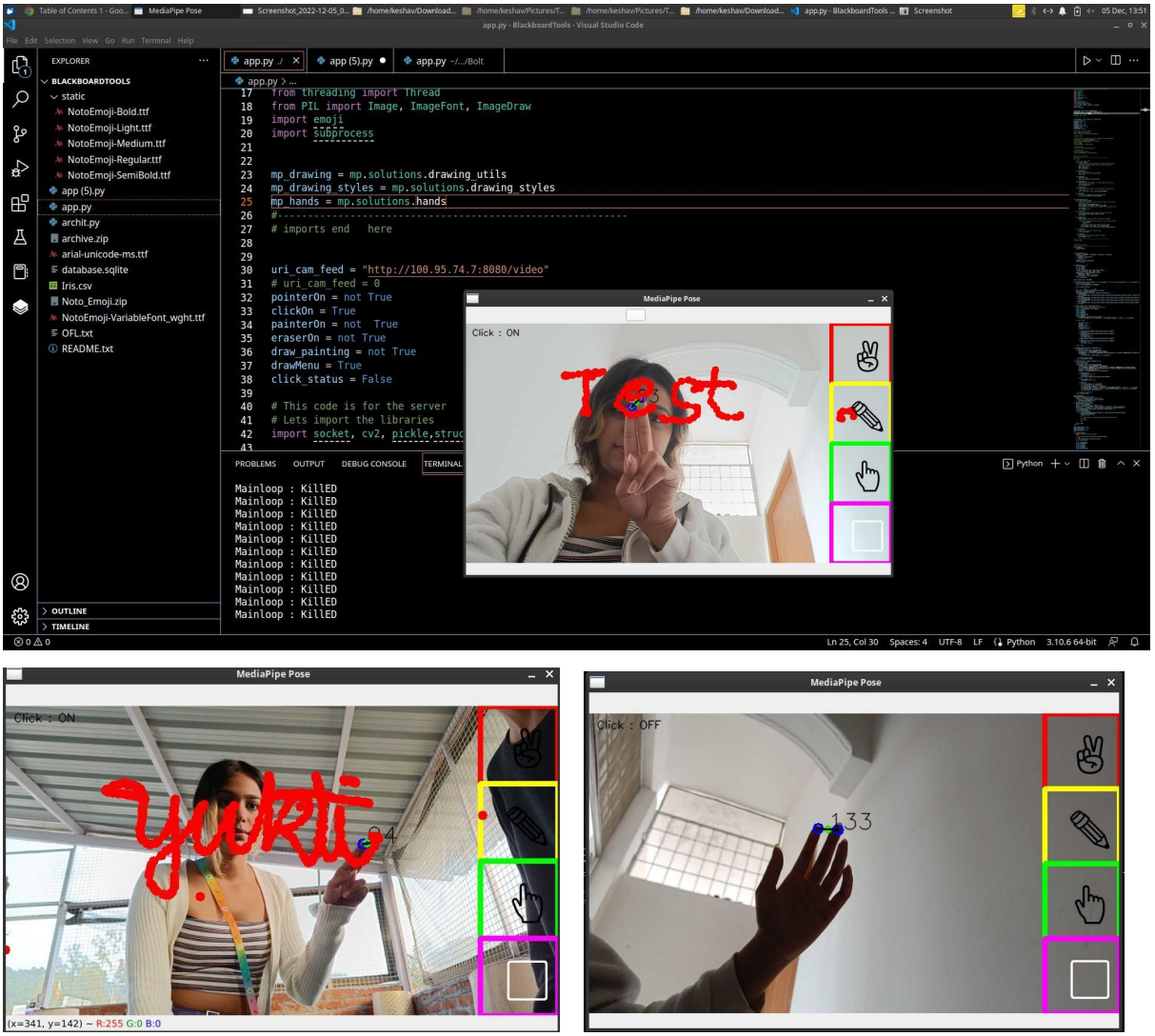



**Fig6:** *Output achieved by the proposed project.*

## 9. Comparative Analysis

In the emerging world of technology there is an emerging need for new computer-human-interaction methods to make day-to-day tasks easier as well as time efficient. The use of traditional input devices is very widely spread but there are other efficient ways in which humans can interact with the computers and use them to achieve their goals.

Gesture recognition is a methodology which is not so common since it is used prominently in the gaming world. However, the process can be used to achieve a lot more than just virtual reality experience in the games.

The project that the team is proposing deals with generalising gesture recognition to complete daily tasks such as online teaching,etc.

The project aims to increase ease for users to access the interface as well as increase the whole system efficiency without using bulky input devices and multiple sensors.

## 10. Conclusion

• Gesture recognition technology is the turning point of the world in technological advancement. It can allow seamless non-touchable control of computerised devices to create a highly interactive, yet fully immersive and flexible hybrid reality. The inclusion of this technology in multiple applications across various sectors is further revolutionising human-computer communication. That said, gesture recognition is no novice's game.

• It's a fully integrated, highly advanced technology that requires specialised skills of individuals with relevant experience that can guarantee favourable results. This software is a development with the resources, talents, and expertise of our proficient and dedicated team members who can recognize and understand human requirements and successfully deliver to the expectations. This software can help people realise that their imagination can be a reality

## 11. Future Scope

Controlling Computer Devices via Gesture Recognition System has various uses which also include using multi-tasking task completion. This technique can be used with various projects and in various ways according to one's imagination and needs as this method is flexible and highly customizable.

The practice can bring significant changes to online teaching methods by providing teachers tools that are easily accessible and students a more real-time experience. Some features accessed by gestures :

- Markers that can draw on the screen using a finger
- Erasers to clear the screen
- Sharing various types of file formats on the screen
- Menu accessed by hand gestures
- Online board to write on with hand gestures, etc.

Using body gestures to provide commands to the computing machine reduces the gap between reality and the machine world which further makes the experience of the user more extraordinary and real. Because of the usability of gesture recognition, this idea and its algorithms can be used in many other areas like game consoles, tablet PCs, etc., Virtual Realities, and so on. Moreover, the major parameters of this technology such as customization and minimization of input devices can be used to create interfaces for specially-abled people, older populations or younger children who are new to technology.

## 12. References

[1] Ustunug A, Cevikcan, Industry 4.0: Managing The Digital Transformation, Springer Series in Advanced Manufacturing, Switzerland. 2018. DOI: https://doi.org/10.1007/978-3-319-57870-5.

[2] Hamed Al-Saedi A.K, Hassin Al-Asadi A, Survey of Hand Gesture Recognition System. IOP Conferences Series: Journal of Physics: Conferences Series 1294 042003. 2019. DOI: https://doi.org/10.1088/1742-6596/4/042003

[3] S.Rautaray S, Agrawal A. Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey. Springer Artificial Intelligence Review. 2012. DOI: https://doi.org/10.1007/s10462-012-9356-9.

[4] Lugaresi C, Tang J, Nash H, McClanahan C, et al. MediaPipe: A Framework for Building Perception Pipelines. Google Research. 2019. https://arxiv.org/abs/2006.10214.

[5] C.Chua, H. Guan, Y.Ho, Model-Based 3D Hand Posture Estimation From a Single 2D Image. Image and VisionComputing vol.20, 2002, pp. 191-202.

[6] M.Panwar, Hand Gesture Recognition Based on Shape Parameters, In International Conferences: Computing Communication and Application (ICCCA), 2012.

[7] Marco Maisto, An Accurate Algorithm for Identification of Fingertips Using an RGB-D Camera, IEEE Journal on Emerging and Selected Topics in Circuits and System, 2013. pp. 272-283.

[8] Wu Xiayou, An Intelligent Interactive System Based on Hand Gesture Recognition Algorithm and Kinect, In 5th International Symposium on Computational Intelligence and Design.2012

[9] Lugaresi C, Tang J, Nash H et.al, MediaPipe: A Framework for Perceiving and Processing Reality. Google Research. 2019.

[10] Abadi M, Barham P, Chen J et.al, Tensorflow: A System for Large-Scale Machine Learning, In 12th USENIX Symposium on Operating System Design and Implementation (OSDI), USA, 2016, https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

[11] Matveev D, OpenCV Graph API. Intel Corporation. 2018.

[12] Zhag F, Bazarevsky, Vakunov A et.al, MediaPipe Hands: On – Device Real Time Hand Tracking, Google Research. USA. 2020. https://arxiv.org/pdf/2006.10214.pdf.

[13] MediaPipe: On-Device, Real Time Hand Tracking, In https://ai.googleblog.com/2019/08/on-devicereal-time-handtracking-with.html. 2019. Access 2021.
[13] MediaPipe: On-Device, Real Time Hand Tracking, In https://ai.googleblog.com/2019/08/on-devicereal-time-handtracking-with.html. 2019. Access 2021.
[14] Grishchenko I, Bazarevsky V, MediaPipe Holositic – Simultaneoue Face, Hand and Pose Prediction on Device, Google Research, USA, 2020, https://ai.googleblog.com/2020/12/mediapipeholistic-simultaneous-face.html, Access 2021.

[15] MediaPipe Github: https://google.github.io/mediapipe/solutions/hands. Access 2021.