Virtual Paint

Dr. Sharanabasappa Gandage[1], Poornachandra Swami[2], Rohit[3], Shankar Reddy Bengire[4]

[1]*Assistant Professor, Department of Computer Science, PDA College of Engineering, Kalaburagi, India, sharangandage@pdaengg.com*

[2]*Student, Department of Computer Science, PDA College of Engineering, Kalaburagi, India, poornachandraswami2000@gmail.com*

[3]*Student, Department of Computer Science, PDA College of Engineering, Kalaburagi, India, rohitrg000@gmail.com*

[4]*Student, Department of Computer Science, PDA College of Engineering, Kalaburagi, India, shankarreddysb03@gmail.com*

ABSTRACT

While most human-machine interactions today take place via direct contact methods like the mouse, keyboard, remote control, touch screen, etc., most human-human interactions take place via non-contact methods like sound and physical movement, which are more natural and intuitive. Many researchers have worked to train machines to recognize human-like nonverbal cues, such as speech, facial expressions, body language, and gestures, in the hopes that they will one day be able to communicate with greater fluidity and efficiency. The most fundamental aspect of human language is gesture, and gestures play crucial functions in human communication. They're the most straightforward approach of conveying information between people and machines. Sign language recognition, robotics, and other fields may all benefit from gesture recognition. Wearable sensor-based approaches and optical camera-based methods are the two simplest ways to classify the devices utilized in gesture recognition. The data glove is an example of a device employed in the wearable sensor-based approach that can accurately capture the motion characteristics of the user's hands, leading to improved recognition accuracy. Wearable sensors are costly to produce and have an adverse effect on the fluidity of user engagement. The use of optical cameras that record a sequence of pictures is at the heart of the optical camera-based technique for distant gesture recording. These visionbased techniques, which use optical cameras to identify motions, work by evaluating visual information collected from pictures. It is challenging to identify and track the hands accurately with optical cameras due to their sensitivity to lighting conditions and clutter backgrounds. Both the optical camera-based method and the wearable sensor-based method have their benefits and drawbacks; however, the project's primary objective was to create a simple and inexpensive approach to enhancing HCI.

Keywords- Virtual Paint, Optical Camera, HCI.

I. INTRODUCTION

As the fields of statistical language modeling and picture recognition have progressed, automatic image caption generation has arisen as a demanding and essential study subject. As well as helping the visually handicapped, the ability to automatically and cheaply identify the millions of photographs published to the Internet every day is another practical advantage of caption creation. The discipline also unifies cutting-edge models from other important subfields of AI, such as Natural Language Processing and Computer Vision.

Wearable sensor-based techniques and optical camera-based methods are the two most common approaches to gesture detection.

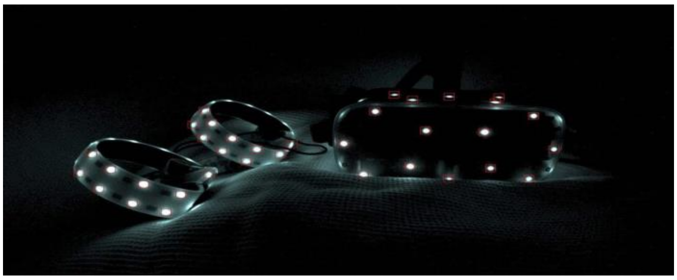

Figure 1: Oculus quest tracking system

Systems like the HTC Vive and the Oculus Rift already use approaches based on wearable sensors. These gadgets are essentially virtual reality game consoles. These systems generally use a signal-generating device and a signal-reading sensor as part of their tracking infrastructure. These may come in a wide variety of shapes and sizes, from optical to electromagnetic to acoustic to mechanical. Cameras of many types are used to record the light generated by optical devices. Small electrified coils that have an effect on others may be located in space via electromagnetic tracking. Ultrasonic sound waves are used in acoustic systems to determine the location and orientation of objects. Accelerometers and gyroscopes are often used in conjunction with articulated arms/limbs/joysticks/sensors that are attached to or housed inside headsets in order to provide mechanical tracking. The HMD and controls for devices like the Rift make use of a constellation of infrared LEDs. Two or more desktop sensors are then able to pick up on the unique illumination of each LED, translating it into precise coordinates. The headgear also has a magnetometer, gyroscope, and accelerometer, which work together to provide precise tracking in all three dimensions. However, the costs and complexity of these monitoring methods are significant drawbacks. It's costly to develop and integrate a wide variety of sensors that all function together and provide reliable data. The portable sensors alone may set you back as little as $200 (12ndpoin, Rs.16000) for a pair.

As the sole source of information, optical camera-based solutions are not as popular as Wearable sensor-based systems due to their lower accuracy of data delivered. The likelihood of realizing reliable information-providing optical camera-based systems has grown thanks to developments in machine learning, artificial intelligence, neural networks, and cutting-edge image processing methods. The limitations of systems based on wearable sensors are avoided by optical camera-based solutions. Since a camera is the only kind of sensor required, the system's price and complexity go down significantly, making it more accessible. In light of the above, we have decided to center our research on camera-based optical systems.

II. PRODUCT SURVEY

2.1 BMW iDrive

Figure 2: BMW gesture control

The notion of gestural input is not novel. Since the early 2010s, automakers have discussed using touchless infotainment controls in place of touchscreens, although few have actually done so.

BMW's newest feature, gesture control, uses hand motions detected by a 3D camera to operate some aspects of the car's infotainment system, iDrive. A call may be accepted or declined, the volume can be adjusted, navigation destinations can be entered, and the vehicle's 360-degree view can be rotated with a few simple gestures. BMW's innovative gesture-control technology was developed to facilitate risk-free iDrive system use. BMWs equipped with gesture control have sensors built into the roof paneling just below the rearview mirror. They interpret hand motions produced in front of the center console, above the gearshift lever, from a predetermined set. The technology works best when the driver is near to the screen and uses hand gestures.

2.2 Oculus rift / HTC vive

Figure 3: VR headset

Virtual reality (VR) has quickly become the most popular kind of online entertainment. PwC predicts that between 2021 and 2025, the market for virtual reality (VR) content will increase at a compound annual rate of 30 percent, surpassing the growth rates of OTT video, video games, and even conventional cinema. Virtual reality headsets provide consumers access to VR material in a fully realistic, three-dimensional setting. Virtual reality (VR) headsets combine a display screen, stereo sound, sensors, and suitable controllers into a single device worn on the head to provide an interactive and immersive audiovisual experience. Putting on a virtual reality headset blocks off the user's surroundings, replacing them with whatever is projected onto the screen. This may be anything from a 360-degree film to a VR game to a virtual office or conference room. Virtual reality is not a spectator sport like 2D video. The virtual environment responds to the user's actions in real time. Virtual reality (VR) headsets feature a variety of sensors, including a 6DoF system for head tracking, to help you do this.

A 6DoF system monitors head motion via gyroscopes, accelerometers, and other sensors, then adjusts the screen to match. The eyes' attention may be detected and translated into action by some VR headsets' eyetracking sensors. Each eye receives its own picture via the lens, simulating how we take in and interpret visual information in the actual world. To further simulate a 360-degree environment, virtual reality (VR) headsets gradually shift the displayed information in reaction to head tracking data, giving the impression that the visuals are moving side to side. Each eye receives its own picture via the lens, simulating how we take in and interpret visual information in the actual world. To further simulate a 360-degree environment, the display information in VR headsets is slightly shifted in reaction to head tracking data, giving the impression that the user is looking about.

2.3 Outcomes of product survey

- At the moment, gesture recognition technologies are only employed in specialized applications.
- Currently available gesture recognition systems are both highly accurate and compatible with a wide range of settings.
- In the present day, each sensor must be hardwired to the controlling module/PC in order for the gesture recognition system to function.
- The vertical field of view of the utilized sensor array is quite limited.
- The high production and integration costs of existing gesture recognition systems
- Base stations, which are positioned on the walls at regular intervals, are used by the systems to track the users' hands.
- There might be hiccups due to reflective surfaces in the room.

III. LITERATURE SURVEY

1. Sushmita Ray, "**A quick review of machine learning algorithms**": - An IEEE xplore publication. This paper introduces the reader to the fundamentals of machine learning and describes several of the most popular and widely-used algorithms in the field, including gradient descent, linear regression, multivariate regression analysis, decision trees, support vector machines, Bayesian learning, naive bayes, the K nearest algorithm, and K means clustering,

among others. All of the algorithms have been compared for their performance and learning speed, and their benefits and drawbacks have been explained.

2. Pramila P. Shinde and Dr. Seema Shah, "**A review of machine learning and deep learning applications**": - The IEEE xplore database has this paper. The importance of knowing about machine learning and deep learning is discussed in this article. We've spoken about how machine learning and deep learning have developed and been used over the last several decades.

3. J. Francis and A. B K, "**Significance of hand gestures recognition systems in vehicular automation – A survey**", In this research, we examine the state-of-the-art approaches to hand gesture detection and recognition, including their strengths and weaknesses, advantages and disadvantages, operating range and design issues, etc.

4. Radhika Bhatt, Nikita Fernandes, Archana Dhage "**Vision based hands gesture recognition for human computer interaction**" University of Mumbai, 2013: This study describes a methodology for creating a hand gesture detection system that can facilitate interaction between humans and computers in real time. "Vision-Based" systems rely only on a camera and Computer Vision (CV) technology, which includes image processing that can identify a variety of hand motions.

5. Jonas robin, Mehul R S, Rishabh Dubey, Nimish Datkhile, Jyoti Kolap "**Computer vision for hand gestures**", International conference on convergence of digital world, 2020: The model used in this study is utilized in environmental detection. The hands (gestures) are the intended items to be identified from the surroundings, and here the identification of the picture is done using the Convolution neural network (CNN) method, the gesture is anticipated, and the result is shown on the screen or audibly presented.

IV. FEASIBILITY STUDY

We analyzed the project's specifications and made sure it could be accomplished using the tools at our disposal by conducting a feasibility analysis. The feasibility study's principal function is to examine the practicability of the proposed project from a technical, operational, and financial perspective. The feasibility study's preliminary inquiry covers three areas.

- Technical feasibility
- Economic feasibility
- Social feasibility

4.1 Technical feasibility

A number of problems arose during technical feasibility analysis that must be resolved in order to realize the technical feasibility.

- Asking whether the necessary tools already exist to carry out a project.
- If this system is ever created, will it be upgradable?
- Does the system provide reliable results, easy access, and secure data?
- Is it easy to get the necessary hardware and software for the proposed system?

Because we want to use optical camera-based detection methods, we may take advantage of already-existing technology such as computer vision, object identification, machine learning algorithms, and artificial intelligence to accomplish very precise tracking. As existing models are refined with additional data over time, the suggested system may be simply updated with faster processors, higher resolution cameras, and more optimized tracking algorithms. The model's precision is limited compared to systems that use a large number of sensors, but with the aid of modern computer vision technology, the findings are quite accurate and trustworthy. The suggested system's intuitiveness stems from its graphical user interface. Because it doesn't need any specialist software or hardware to run, the program is both inexpensive and simple to deploy and update.

4.2 Economic feasibility

The term "economic feasibility" refers to an evaluation of a project's viability in terms of its costs and potential returns. Definition of Project Feasibility Analysis: an objective and reasoned analysis of a project's merits that helps with decision-making by outlining the project's benefits, drawbacks, opportunities, risks, and required resources, as well as estimating its likelihood of success.

Market research is a crucial aspect of economic viability. The project's concentration on gesture and contactless HCI makes it relevant to the market right now. Contactless HCI came into focus after the COVID-19 pandemic as a means to improve public health by reducing the spread of disease via the use of touch-based devices such as automated teller machines, ticket dispensers, and the like. It could be worth it if such services can be used without having to physically touch the screen.

It should still be a good investment for a company to adopt the suggested system if it is created and all the anticipated features are deployed. Since the system is constructed using existing technology and resources, it is cost-effective, requires no special hardware or software, and poses no security risks to the business.

4.3 Social feasibility

The goal of a social feasibility study is to investigate the potential social and economic effects of a project. This sort of research may, for instance, examine how the existing social structure of the region affects the availability of competent staff or the level of community support for the project.
The potential social benefits of the proposed initiative are encouraging. The general population may have to become used to using new machinery and procedures if widespread adoption is to occur. On the plus side, it has the potential to boost public health by decreasing people's exposure to germs on routinely handled objects.

V. SYSTEM ANALYSIS

5.1 System study and environment

By studying the features of the scene's structure, computer vision describes how a 3D scene may be reconstructed, interrupted, and understood from its 2D pictures. The focus is on simulating human eyesight using computational tools. We use the OpenCV framework and a computer to do computer vision tasks. OpenCV was chosen because of its flexibility and the wide variety of features it supports, including object identification, pattern recognition, and more. Alternative frameworks include TensorFlow, OpenGL, PyTorch, and OpenCL.

5.2 Computer vision

The goal of computer vision is to give computers the same ability to recognize and understand the world around them as human beings do by simulating the human visual system as closely as possible. Computer vision has limited functionality until recently.

Artificial intelligence, namely deep learning and neural networks, has made tremendous strides in recent years, allowing it to outperform humans in various object-related detection and categorization tasks.

The massive amounts of data we produce now are helping to train and improve computer vision, which is helping to fuel the field's rapid expansion.More than 3 billion photographs are uploaded on the internet every day, and with that comes the availability of the processing power needed to make sense of it all. New hardware and algorithms in the area of computer vision have led to increased accuracy rates for object detection. Today's technologies are more accurate than humans in immediately reacting to visual inputs, with an accuracy rate of 99 percent achieved in less than a decade.

From its inception in the 1950s through its first commercial use in the 1970s, when it could tell the difference between typed and handwritten text, the field of computer vision has come a long way. Computer vision, at its core, relies on the ability to recognize patterns. One way to teach a computer to understand visual data is to provide it with large amounts of labeled images—thousands, if not millions—and to run the computer through a series of software algorithms designed to detect hidden patterns in the data.

Thousands of photographs of Abraham Lincoln, for instance, would be analyzed by algorithms that take into account things like color, form, distance from other shapes, bordering objects, and so on. Therefore, a definition of "Abraham Lincoln" is established. Each pixel's brightness is represented by a single 8-bit value, with a range of 0 to 255, as seen in the example picture 4 depicting the grayscale image buffer that saves our image of Abraham Lincoln.

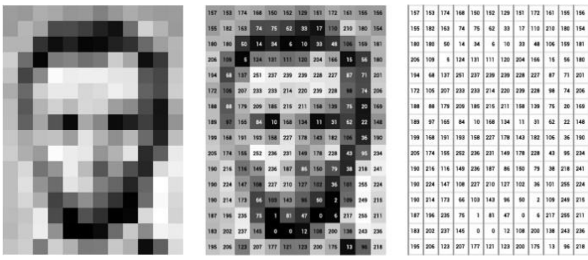

Figure 4: Abraham Lincoln Pixel data diagram

5.3 OpenCV

Open-source computer vision is what makes OpenCV possible. Simply said, it's a library for handling images. It is a large open-source library used for image processing and other computer vision-related applications, as well as in domains driven by AI and ML. Therefore, it is becoming more important in modern systems for use in real-time applications. One can analyze photos and videos using OpenCV to recognize objects, people, and even human handwriting.

OpenCV was created by Intel, then maintained by Willow Garage and Itseez before being bought by Intel. The first release of OpenCV was 1.0. This open-source, cross-platform library is distributed under the BSD license and is available without cost for any purpose, whether it educational or commercial. The application programming interfaces (API) for the languages C++, C, Python, Java, and MATLAB (a proprietary multi-paradigm programming language that provides a solid numeric computing environment) are all available in the online documentation. OpenCV works with a wide variety of platforms, including Windows, Linux, Mac OS, iOS, and Android. OpenCV was originally developed with real-time applications and computing efficiency in mind. For real-time computations, OpenCV has included GPU acceleration since 2011. Python's array processing capabilities allow for analysis with OpenCV once it is combined with additional libraries like NumPy. Using vector space and performing mathematical operations on the characteristics of interest in a picture is necessary for pattern recognition.

5.4 Mediapipe

MediaPipe is a machine learning pipeline framework that may be used to handle audio, video, and other forms of time-series data. The Raspberry Pi and the Jetson Nano are only two examples of the embedded devices that are compatible with this framework.

The MediaPipe Toolkit includes both the Infrastructure and the Fixes. The following structure diagram illustrates how the MediaPipe Toolkit works.

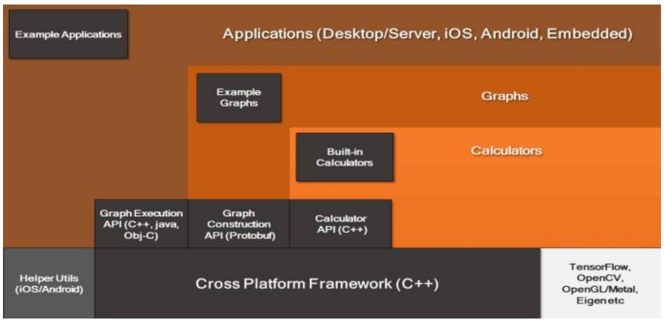

Figure 5: Mediapipe Toolkit

5.5 Mediapipe Hands

MediaPipe Hands is a sophisticated tool for monitoring hands and fingers. Using ML, it can deduce 21 unique 3D landmarks of a hand from a single image. Our solution delivers real-time performance on a cell phone and even scalable to many hands, whereas the existing state-of-the-art relies mostly on powerful desktop systems for inference. By making this hand perception capability available to the broader research and development community, we want to inspire the creation of novel use cases, which in turn will pave the way for exciting new applications and lines of inquiry..

5.6 Existing systems

There are a few patents and trademarks that are tangentially connected to our work that we found when searching for preexisting systems on the websites of the Indian patent office and the USPTO (United States Patent and Trademark Office).

The aforementioned trademarks and patents make use of various technologies, with some even using hardware like raspberry pi. All of the outcomes share the fact that they were developed with a particular goal in mind (in this case, the control of a robotic arm). The purpose of our suggested project is only to showcase the potential of gestures utilizing computer vision, which can be readily adapted to a wide variety of solutions, including automated teller machines, kiosks, and even systems like sign language recognition and controls for Internet of Things devices.

5.7 Proposed System

For the purpose of applying digital paint, we suggest a hand recognition system. The camera is used to record the user's movements, and each frame is then analyzed using Opencv and mediapipe. When mediapipe analyzes a video, it begins by following the hand's palm before marking each of the 21 key locations. In order to detect the gesture and provide input to the system or application, the seen landmarks are matched with gesture library. In contrast to the current system, our suggested system does not rely on sensors like gyroscopes, accelerometers, etc., nor does it need any specialized IR or UV cameras/sensors to identify gestures. This bodes well for the system's robust and consistent functioning. Since the purpose of the virtual paint application is only to demonstrate the efficiency of the system, the gesture detection system is flexible enough to be adapted to suit a wide variety of systems and purposes.

5.8 System requirements

5.8.1 Hardware requirements

- CPU/processor: Minimum dual core intel/AMD processor at 2Ghz
- RAM: 4GB minimum
- Storage: 500MB minimum
- Camera: Minimum 720p or 1080p camera

5.8.2 Software requirements

- Programming language: Python3
- IDE: PyCharm
- Operating system: OS independent

PAGES:25-38
5/21/23
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY VOLUME-1 ISSUE-2 MAY
Journal Recognised by Government of India

- Tools/modules/frameworks: OpenCV, NumPy, MediaPipe, Cvzone

VI. SYSTEM DESIGN

6.1 System architecture

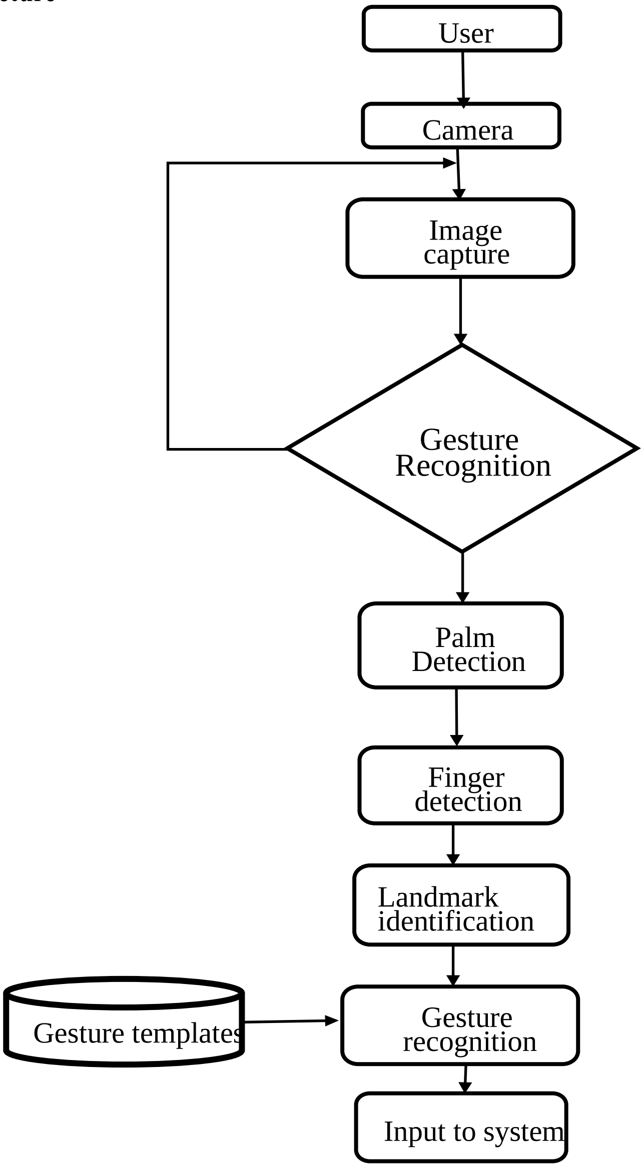

Figure 6: System architecture

6.2 UML Diagrams

Using the modeling notation governed by a set of grammar, linguistics, and pragmatic norms, the Unified Modeling Language enables the technologist to specify AN analytical model. In the Unified Modeling Language (UML), there are five diagrammatic viewpoints that each explain the system from a unique angle. The following is a set of diagrams that serves as an overview of every read. It depicts the interactions between a wide variety of structural components outlined in the user model and structural model read, and so portrays the dynamic of behavioral as parts of the system. Applicable Instances The diagrams show how useful the system is from the perspective of the reader. Throughout the process of needs induction and analysis, use cases are a representation of how useful the system will be. The system's behavior is the focus of use cases, which are written with an outside reader's needs in mind. A system's actors are non-system entities that participate in its dynamics. Examples of actors include the system administrator, the bank customer, etc.

6.2.1 Use case diagram

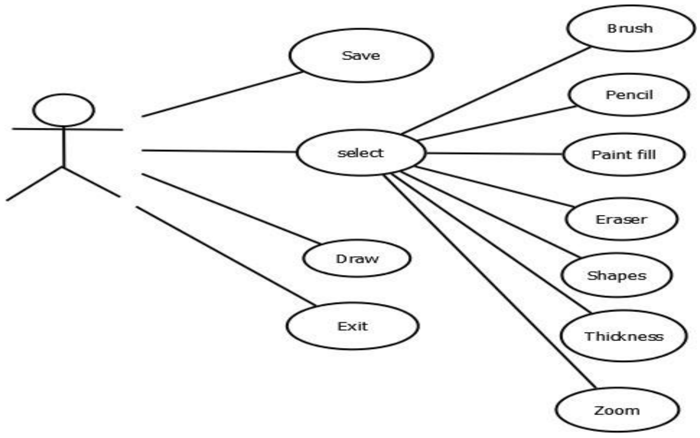

Figure 7: Use case diagram

6.2.2 Component diagram

The arrangement of a system's physical components may be shown using an element diagram. An element may represent a tangible component of the whole. It's shown here as a tabbed parallelogram.

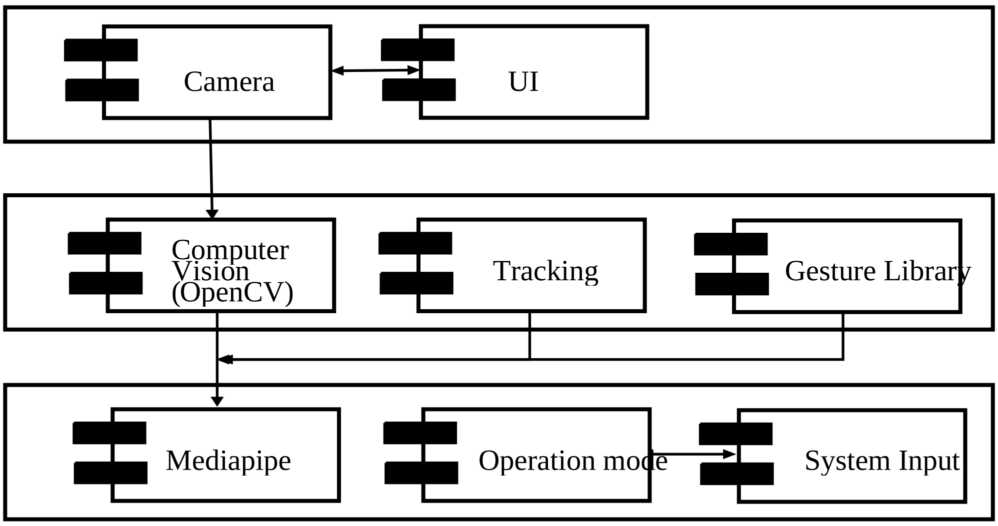

Figure 8: Component diagram

VII . SYSTEM TESTING

The goal of testing is to unearth flaws. The goal of testing is to expose every possible flaw or vulnerability in a product. It allows one to test the operation of individual parts, whole assemblies, and even final products. Software testing is the practice of putting a program through its paces to make sure it won't crash or otherwise behave badly during usage. There are several kinds of examinations. There are several kinds of tests, and each one caters to a different need.

Initial testing entails crafting a thorough strategy for verifying core features and operation across a wide range of hardware configurations. Quality is ensured by means of stringent controls.

This procedure ensures that the system requirements paper have been met and that the program is bug free. The following are some of the factors taken into account when designing the framework for the testing procedures.

7.1 Unit Testing

The goal of unit testing is to ensure that the underlying logic of a program is working as intended and that the inputs to the program result in expected results. It is important to verify the correctness of all code paths and decision trees. It's the process of verifying the functionality of the application's constituent parts.It's done after each component is finished but before they're integrated. This is an invasive kind of testing that requires specific information about the structure being tested. Unit tests are simple tests that check one business process, application, or system configuration at a time. Each branch of a business process should have its own set of unit tests to guarantee that it follows the published requirements and produces the desired outputs.

7.2 Functional Testing

The purpose of functional testing is to systematically verify, against documented business and technical requirements, system documentation, and user guides, that the functionality being tested is in fact there.

7.3 System Testing

System testing verifies that the totality of an integrated software system is up to snuff. It puts a setup to the test to guarantee consistent and reliable outcomes. The configuration-oriented system integration test is a kind of system test. System testing is based on process flows and descriptions, with an emphasis on pre-driven integration nodes.

7.4 Performance Testing

By timing how long it takes the system to compile, respond to users' requests, and retrieve results, the Performance test guarantees that output will be delivered on time.

7.5 Integration Testing

To simulate failures brought on by interface faults, software integration testing progressively integrates two or more software components on a single platform.

The goal of an integration test is to ensure error-free communication between different parts of a system or different levels of an organization's software.

7.6 Acceptance Testing

In order to be successful, user acceptability testing must include the end user extensively. In addition, this process guarantees that the system can perform as intended.

VIII. RESULTS

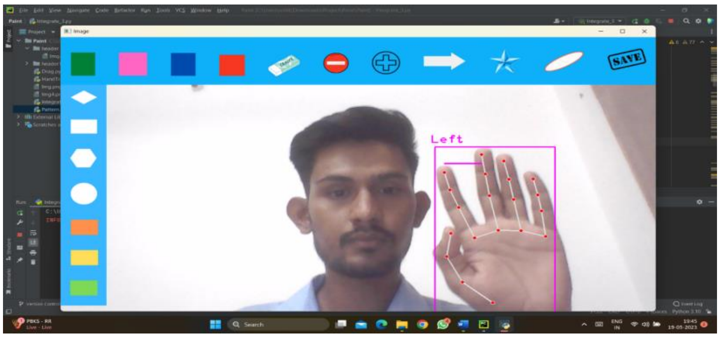

Fig-9: Working of handtracking

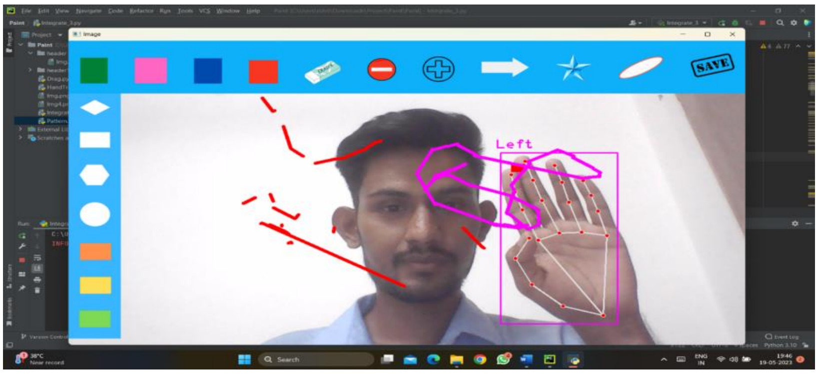

Fig-10: Working of brush

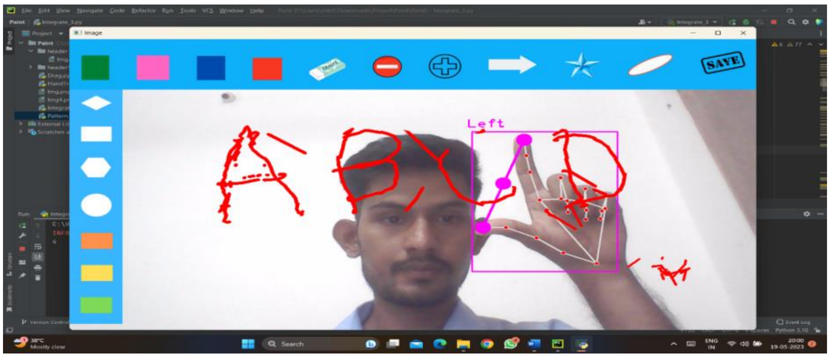

Fig-11: Working of thickness adjust

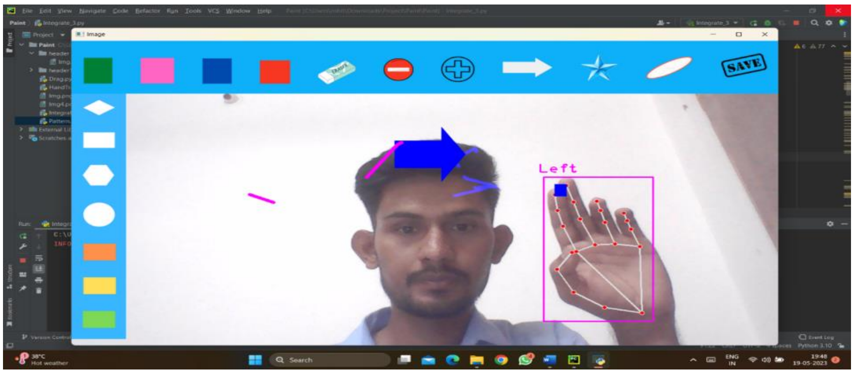

Fig-12: Working of shape insertion

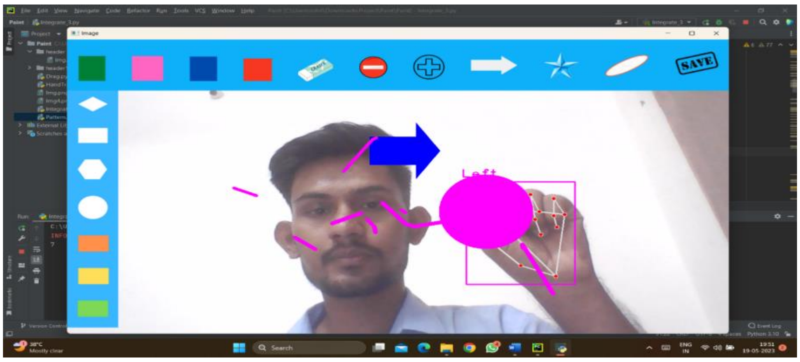

Fig-13: Working of drag and drop shape

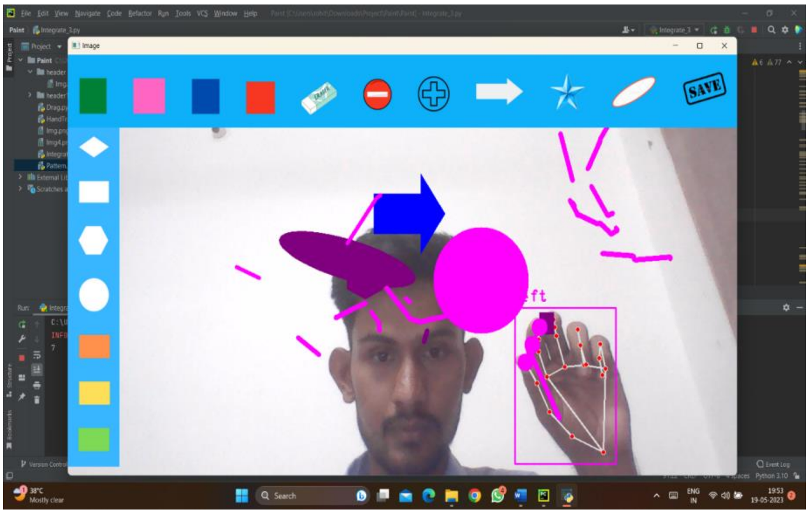

Fig-14: Working of change shape

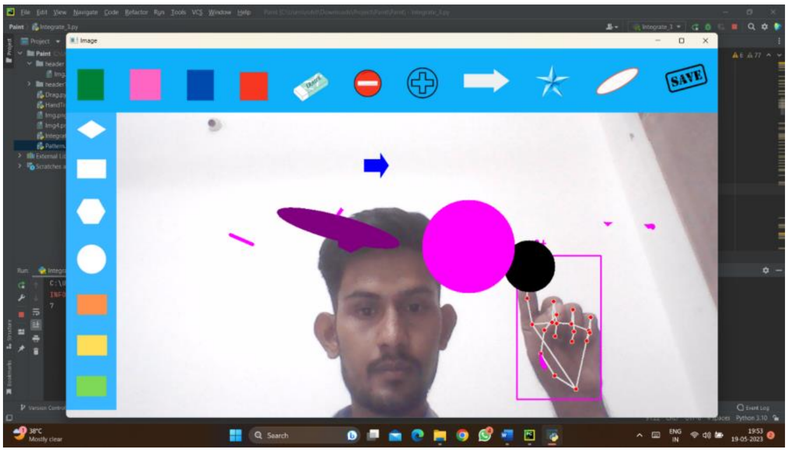

Fig-15: Working of eraser

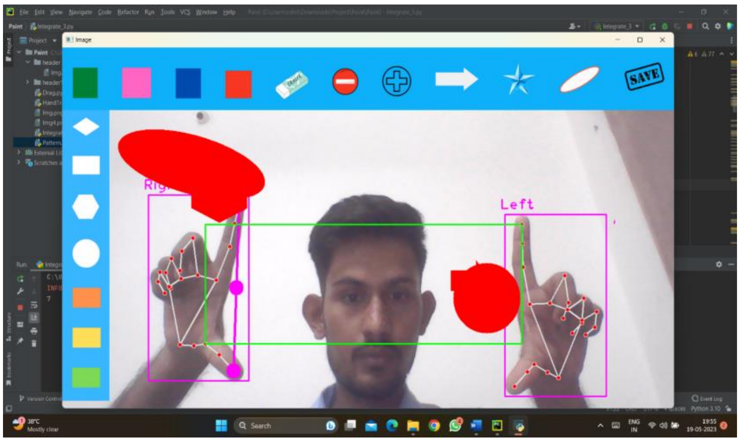

Fig-16: Working of crop tool

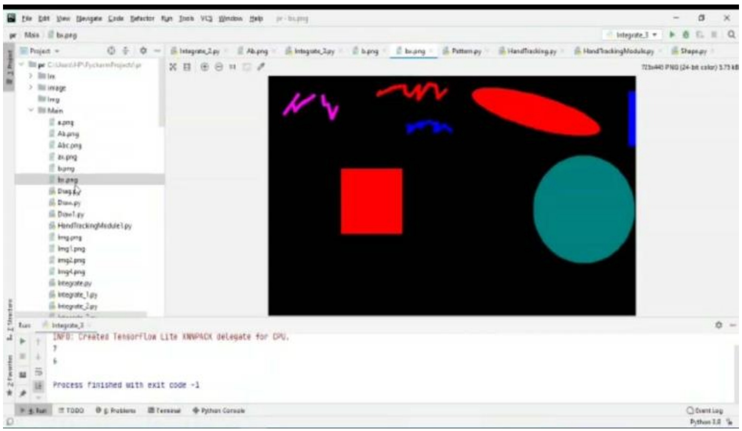

Fig-17: Saved cropped image

IX. CONCLUSION AND FUTURE SCOPE

We planned and executed a project to create a gesture-based user interface (HCI) application in an effort to demonstrate that this method is just as effective as the more conventional methods of HCI (using a mouse, keyboard, and touch screen). We used current technology to create a low-cost, user-friendly method of incorporating gesture-based controls into present and future gadgets and services. Machine learning this method of engaging with computers will get simpler as camera file technology improves at a quick pace. Gesture controls are a wonderful approach to allow the public to confidently connect with public services like ATMs at a time when the globe is seeking for ways to avoid interaction with public services. Future progress in gesture based HCI may benefit greatly from open source models like mediapipe.

REFERENCES

1 J. Francis and A. B K, "Significance of Hand Gesture Recognition Systems in Vehicular Automation-A Survey", International Journal of Computer Applications, vol. 99, no. 7, pp. 50-55, 2014.
2. R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady, “A Linguistic Feature Vector for the Visual Interpretation of Sign Language”, in Tomas Pajdla, Jiri Matas
(Eds), Proc. European Conference on Computer Vision, ECCV04, v. 1: 391-401, LNCS3022, Springer-Verlag, 2004.
3. Radhika Bhatt, Nikita Fernandes, Archana Dhage "Vision Based Hand Gesture Recognition for Human Computer Interaction" University Of Mumbai (IJESIT) Volume 2, Issue 3, May 2013.
4. Siam, Sayem & Sakel, Jahidul & Kabir, Md. . Human Computer Interaction Using Marker Based Hand Gesture Recognition., 2016
5. M. Lee and J. Bae, "Deep Learning Based Real-Time Recognition of Dynamic Finger Gestures Using a Data Glove," in IEEE Access, vol. 8, pp. 219923-219933, 2020, doi: 10.1109/ACCESS.2020.3039401.

6. S. Khan, M. E. Ali, S. Das and M. M. Rahman, "Real Time Hand Gesture Recognition by Skin Color Detection for American Sign Language," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 2019, pp. 1-6, doi: 10.1109/EICT48899.2019.9068809.
7. P. Ramasamy, G. Prabhu and R. Srinivasan, "An economical air writing system converting finger movements to text using web camera," 2016 International Conference on Recent Trends in Information Technology (ICRTIT), 2016, pp. 1-6, doi: 10.1109/ICRTIT.2016.7569563.
8. Prajakta Vidhate, Revati Khadse, Saina Rasal, “Virtual paint application by hand gesture recognition system”, 2019 International Journal of Technical Research and Applications, 2019, pp. 36-39.