# Gesture Recognition in the field of Human-Robot Interaction (HRI): A Case Study on the 'Air Canvas Board' Application

**Florisa Zanier**
Matr.: 3613450

**Venkata Gopi Krishna Miriyala**
Matr.: 3601156

**Sameer Tuteja**
Matr.: 3296444

## Abstract

Human-Robot Interaction, also known as HRI, is a field that is constantly developing and focuses on investigating how humans and robotic systems can collaborate efficiently. New opportunities have been made available in this field as a result of the development of technology that recognize gestures, which have made interactions more user-friendly and intuitive. The 'Air Canvas Board' application, which makes use of gesture detection for digital drawing and painting, is presented as a practical example of HRI in this paper.

**Keywords:** Human-Robot Interaction, Gesture Recognition, MediaPipe

## 1 Introduction

The field of HRI, has become an important topic of research in the areas of artificial intelligence and robotics in recent years. The purpose of this research is to investigate how people may interact with robots in a way that is both harmonious and effective, with the ultimate goal of incorporating robotic systems into everyday human surroundings. The development of artificial intelligence (AI), sensor technology, and processing capacity have all had a significant impact on the evolution of this discipline. 'Air Canvas' is an application that exemplifies this evolution by utilizing gesture recognition, which is a method of interaction that is both smart and intuitive, and it matches closely with normal human behavior. This study investigates the implementation of gesture-based HRI in the 'Air Canvas Board' and the implications of doing so. It highlights the role that gesture-based HRI plays as an innovative bridge between humans and digital systems (Cassell and Thorisson, 1999). The growing prominence of robotic systems in a variety of industries, including healthcare, education, and entertainment, has resulted in a demand for HRI that is more natural and intuitive in nature. Keyboard and mouse inputs, as well as touch displays, are examples of traditional interaction methods. However, these approaches frequently lack the flexibility and ease of use that are necessary for smooth human-robot collaboration. Because it provides a paradigm for interaction that is not only user-friendly but also flexible enough to accommodate a wide range of user requirements and circumstances, gesture recognition stands out as a potentially useful remedy to these restrictions.

## 2 Background

Throughout its history, human resource intelligence (HRI) has been limited by technological restrictions that prevent it from effectively comprehending human behaviors and intents. In the beginning, robotic systems were frequently restricted to doing tasks that had been pre-programmed and had very little interaction with human operators. On the other hand, the development of increasingly sophisticated artificial intelligence and machine learning algorithms has made it possible for robots to comprehend and react to a greater variety of human behaviors, such as voice, facial expressions, and gestures (McNeill, 1992). The technological developments that have been made in computer vision and sensor technologies have led to substantial advancements, in particular in the field of gesture detection. The ability of a system to recognize and understand human gestures as input commands is essential to the functioning of this type of interaction (Rifinski et al., 2021). The field is one that pulls from a variety of fields, such as computer science, linguistics, psychology, and human-computer interaction, among others. In the realm of non-verbal communication, gestures play a crucial role in how we convey our thoughts and feelings. The ability of gesture recognition to imitate human-to-human conversation is the essence of this communication method(**?**) The 'Air Canvas Board' application is a good example of how these technological breakthroughs in gesture recognition may be directly applied. Through the use of computer vision, it is able to monitor hand movements and convert them into digital commands that may be used for drawing and painting (Shneiderman et al., 2016). This approach represents a growing trend in HRI, which is shifting the focus toward making robotic systems more adaptive and responsive to the requirements and behaviors of humans.

## 3 The 'Air Canvas' Application: An In-Depth Look

The 'Air Canvas Board' makes use of MediaPipe, which is a framework for machine learning with the purpose of constructing pipelines for the processing of multimedia content, with a special emphasis on hand tracking solutions. Hand motions that are caught by a webcam are interpreted by the application, which then converts them into digital strokes that are painted on a canvas.

### 3.1 Key Features

- Advanced Hand Tracking: Utilizing MediaPipe's hand tracking capabilities for real-time gesture recognition.
- Digital Drawing Tools: Enabling users to interact with digital elements like selecting colors and drawing.
- Auditory Feedback: Implementing pyttsx3 for real-time text-to-speech feedback, enhancing user engagement.
- Intuitive User Interface: Incorporating a user-friendly GUI designed with Tkinter.

### 3.2 Gesture Recognition as the Core of HRI

Gesture detection makes the 'Air Canvas Board' app a model of advanced HRI. This technology is essential for converting human hand movements into meaningful digital interactions. This system uses a complicated set of algorithms to recognize motions that correlate to application commands. Technology: Cutting-edge algorithms evaluate camera data in real time to enable 'Air Canvas Board' gesture detection. These methods track hand motions using machine learning and geometric modeling. The 'Air Canvas Board' prioritizes gesture input over keyboard-and-mouse installations. This change makes technology more adaptive to human behavior rather than constraining people to use standard input devices. Interaction: The 'Air Canvas Board' uses natural motions for input, making it easy to use. Users may sketch, select, and navigate the application's UI using basic hand movements, just like they would with real things. Versatility and adaptability: This HRI method is flexible and adjustable. The gesture detection system can be refined or developed to recognize more movements, making it appropriate for VR, AR, gaming, and educational uses beyond digital drawing (Mitra and Acharya, 2007). New users benefit from gesture-based interaction's lower learning curve. Gesture control, unlike specialized input devices or complicated software interfaces, uses natural human motions, making it more accessible to more people, even those new to computers. Despite its benefits, gesture detection in the 'Air Canvas Board' has drawbacks. Lighting, background noise, and camera quality affect gesture detection accuracy. The system must also reliably understand gestures from a broad user base with different hand sizes and movement patterns. Finally, gesture detection in the 'Air Canvas Board' shows HRI's progress. It shows how digital interactions will be effortlessly interwoven into human behavior, enabling more intuitive and accessible technological interfaces.

### 3.3 MedialPipe Architecture

The MediaPipe Hand Landmarker task enables the identification and localization of the specific anatomical landmarks of the hands inside an image. It is utilized to identify crucial landmarks on hands and provide visual enhancements to them. This assignment utilizes a machine learning (ML) model to process image data, either as static data or as a continuous stream. The output includes the coordinates of hand landmarks in both image and world coordinates, as well as the handedness (left or right hand) of multiple detected hands. The following features are included (goo, ):

- Image processing involves performing operations such as rotating, scaling, normalizing, and converting color spaces on an input image.
- Score threshold: Use prediction scores to filter results.

The Hand Landmarker can process input data of the following types: still images, decoded video frames, and live video feed. The tool produces the following outcomes: the dominant hand's preference, the landmarks of the identified hands in image coordinates, and the landmarks of the detected hands in world coordinates. IT utilizes a model bundle that includes two pre-packaged models: a model for detecting palms and a model for detecting hand landmarks. To execute this task, you require a model bundle that includes both of these models. The hand landmark model bundle identifies the precise locations of 21 knuckle coordinates on the hand among the detected hand areas. The model underwent training using around 30,000 real-life photos, alongside many artificially generated hand models superimposed on different backdrops. The hand landmarker model bundle comprises a model for

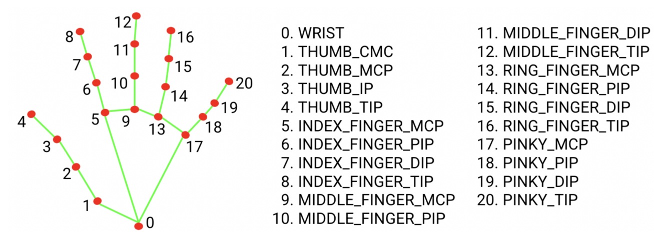


Figure 1: keypoint localization within the detected hand regions(goo, ).

detecting palms and a model for detecting hand landmarks. The palm detection model locates hands in the input image, and the hand landmarks detection model recognizes particular landmarks on the cropped hand image that the palm detection model has identified. Due to the time-consuming nature of running the palm detection model, Hand Landmarker utilizes the bounding box established by the hand landmarks model in a single frame to identify the area of the hands in consecutive frames during video or live stream running mode. The Hand Landmarker module will only activate the palm detection model again if the hand landmarks model no longer detects the existence of hands or if it fails to track the hands within the frame. This minimizes the frequency at which Hand Landmarker activates the palm detection model.

### 3.4 Practical Implications

The 'Air Canvas' application, with its innovative use of gesture recognition technology, finds a wide range of applications across various domains, significantly enhancing the way users interact within these fields:

#### 3.4.1 Education: Facilitating Interactive Learning Environments

The application introduces a dynamic element to traditional educational settings. By allowing students to interact with digital content through natural hand movements, it fosters a more engaging and participatory learning environment. Teachers can use this technology to demonstrate complex concepts in a visually interactive manner, catering to various learning styles, especially for visual and kinesthetic learners. It also offers potential for developing interactive educational games and activities, making learning more fun and effective.

#### 3.4.2 Creative Arts: Offering a New Medium for Digital Artistry

In the realm of creative arts, the 'Air Canvas Board' serves as a bridge between traditional artistic methods and digital art forms. Artists can use gestures to create paintings and drawings, offering a new tool for artistic expression. This technology can be particularly appealing to digital illustrators and graphic designers, providing a more intuitive and natural way of interfacing with digital canvases. The application also opens up possibilities for collaborative art projects and performances, where artists can create and manipulate visual elements in real-time, offering a novel way to engage with audiences.

#### 3.4.3 Rehabilitation: Assisting in Motor Skill Recovery Through Engaging Activities

In therapeutic settings, the 'Air Canvas Board' can be a valuable tool for patients undergoing rehabilitation for motor skill recovery. The gesture-based interaction requires users to move their hands and arms, which can be an integral part of physical therapy routines. It offers a fun and engaging way for patients to perform their exercises, potentially increasing compliance and enhancing the recovery process (Gupta and O'Malley, 2006). The application can also be tailored to suit the specific needs of individual patients, allowing therapists to adjust the level of difficulty and range of motion required according to their progress. By tapping into these diverse areas, the 'Air Canvas Board' demonstrates the broad potential of gesture recognition technology in enhancing not only the user experience but also in contributing significantly to fields like education, the creative arts, and rehabilitation.

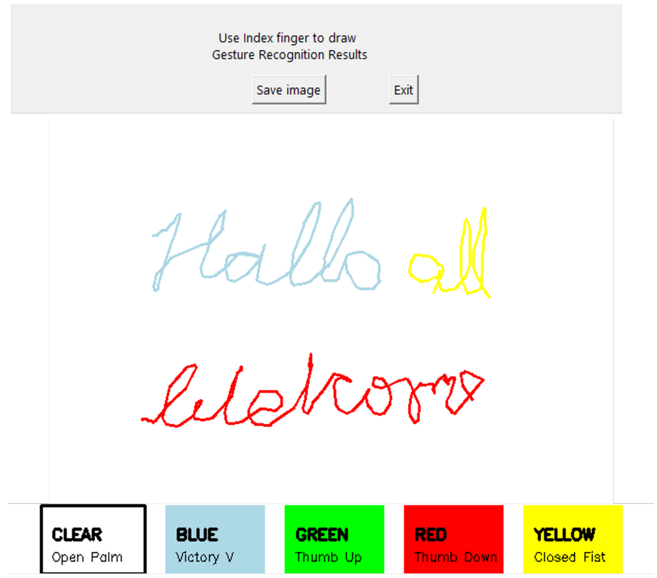


Figure 2: GUI from "Air Canvas Board" Application.

## 4 User Experience and Accessibility

The 'Air Canvas Board' that was designed with the user in mind brings to light the significance of accessibility in the realm of technological advancements. The application's user-friendly gesture control approach reduces the steepness of the learning curve, making it accessible to a wide range of users (Card, 2018).

## 5 Challenges and Future Directions

Despite the fact that 'Air Canvas Board' demonstrates the promise of gesture-based human-computer interaction (HRI), it also exposes problems such as dependence on technical capabilities and ambient circumstances. It is possible that future advances may concentrate on improvements to gesture detection algorithms and the expansion of the gestural language in order to accommodate interactions that are more complicated.

## 6 Conclusion

The application "Air Canvas Board" is an ideal illustration of the advancements that have been made in the field of HRI and offers a glimpse into the trajectory that human-computer interfaces may take in the future. As the technologies that are used for gesture detection continue to progress, they have the potential to build interaction paradigms that are more natural, intuitive, and accessible across a wide range of businesses.

## References

Stuart K Card. 2018. *The psychology of human-computer interaction*. Crc Press.

Justine Cassell and Kristinn R Thorisson. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4-5):519–538.

Hand landmarks detection guide. `https://developers.google.com/mediapipe/solutions/vision/hand_landmarker`. [Accessed 24-01-2024].

Abhishek Gupta and Marcia K O'Malley. 2006. Design of a haptic arm exoskeleton for training and rehabilitation. *IEEE/ASME Transactions on mechatronics*, 11(3):280–289.

David McNeill. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Sushmita Mitra and Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3):311–324.

Danielle Rifinski, Hadas Erel, Adi Feiner, Guy Hoffman, and Oren Zuckerman. 2021. Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction. *Human–Computer Interaction*, 36(4):333–359.

Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. *Designing the user interface: strategies for effective human-computer interaction*. Pearson.