International Journal of Innovative Research in Technology and Science

ISSN: 2321-1156 www.ijirts.org Volume 12 Issue 2, March 2024

Air Pose Canvas with Hand Pose Estimation using Machine Learning

Heena Ansari, Aryan Dakhore, Aditya Paunikar, Aryan Khobragade,
Nishant Ramteke, Pallavi Sambhare
Department of Information Technology
Kavikulguru Institute of Technology and Science, Maharashtra, India

Selection and peer review of this article are under the responsibility of the scientific committee of the International Conference on Current Trends in Engineering, Science, and Management (ICCSTEM-2024) at SAM Global University, Bhopal.

Abstract— The Air Pose Canvas with Hand Pose Estimation and Machine Learning is a transformative technology that re-imagines digital interaction by combining the power of machine learning, computer vision, and natural hand gestures. This innovative system enables users to create, draw, and interact with digital content using their hands as the primary input device. Hand pose estimation, powered by advanced machine learning models, accurately tracks hand movements and gestures in real-time, providing a seamless and intuitive user experience. This groundbreaking technology has broad applications across various domains, including artistic expression, virtual reality, education, healthcare, gaming, and accessibility. Users can sculpt, draw, and precisely manipulate digital content, making it a valuable tool for artists, educators, therapists, and individuals with disabilities. It enhances accessibility and inclusivity by offering a natural interface for individuals with diverse abilities. As technology advances, the “Air Pose Canvas” opens up exciting possibilities for immersive virtual and augmented reality experiences, collaborative workspaces, and intelligent assistance through artificial intelligence integration. It represents the future of human-computer interaction, where physical gestures become a powerful medium for creative expression and interaction with the digital world. This abstract explores the key components and significance of the “Air Pose Canvas” while highlighting its potential to reshape digital interaction across various applications and industries.

Keywords— Air Pose Canvas, Hand Pose Estimation, Machine Learning, Gesture Recognition, Digital Interaction.

I. INTRODUCTION

In an ever-evolving landscape of human-computer interaction, where innovation seeks to bridge the gap between the digital and physical worlds, Air Pose Canvas with Hand Pose Estimation using Machine Learning emerges as a transformative technology [1]. This groundbreaking system redefines how we engage with digital content, unleashing creative possibilities through gestures' intuitive language and machine learning's power. In technology, we constantly seek more natural, immersive, and accessible ways to interact with computers and digital environments. While effective, traditional input devices like keyboards and mice often lack the intuitive and expressive qualities of human gestures and touch. Air Pose Canvas represents a paradigm shift, enabling users to communicate with computers and digital canvases through body language. At its core, Air Pose Canvas relies on sophisticated machine learning algorithms, advanced sensors, and cameras to interpret and respond to the intricate movements and poses of the user's hands in real time. This fusion of cutting-edge computer vision and artificial intelligence technologies unlocks a new dimension of creative expression and interaction, blurring the physical and virtual boundaries. The motivation behind this technology is multifaceted. It aims to enhance user experiences, democratize digital creativity, foster inclusivity, and facilitate learning across diverse domains [4]. It promises to empower artists, designers, educators, professionals, and individuals of all backgrounds to harness the power of gesture-based interaction for many applications. This journey into the world of Air Pose Canvas begins with a deep dive into its architecture, exploring the intricate interplay of hardware and software components that enable the system's magic. From hand pose estimation algorithms to gesture recognition models, from sensor technologies to user interfaces, we explore a technology poised to redefine how we interact with the digital realm.

II. OBJECTIVE

The objective of developing "Air Pose Canvas with Hand Pose Estimation using Machine Learning" is multifaceted, guiding its development, implementation, and utilization. The primary goal is to develop and implement robust machine learning algorithms capable of accurately and reliably estimating hand poses in real-time, ensuring precise tracking of hand movements and gestures. Additionally, the aim is to create an intuitive and natural interaction experience for users, enabling them to interact with digital content effortlessly by mimicking real-world hand movements and gestures. Another objective is to ensure the system's real-time responsiveness, minimizing latency to provide immediate feedback on users' hand movements and gestures. Furthermore, the goal is to make the technology accessible to a wide range of users, including those with varying technical expertise and physical abilities. This involves ensuring the interface is user-friendly and customizable to cater to different user needs. Additionally, the objective is to enable users to express themselves creatively by supporting various creative activities such as drawing, painting, sculpting, and 3D modelling, allowing them to digitally bring their ideas to life.

III. APPLICATIONS

It has various applications across domains because it enables natural and intuitive interactions with digital content. Artists and designers can use Air Pose Canvas to create digital artworks, illustrations, and 3D models with the precision and expressiveness of hand movements. It can be employed in graphic design software, allowing designers to naturally sketch, draw, and manipulate digital elements. Teachers can use this technology in educational settings to make lessons more interactive and engaging, helping students better understand complex concepts. Medical schools can utilize it for anatomy lessons, allowing students to dissect and explore anatomical structures virtually. Game developers can integrate hand pose estimation into virtual (VR) and augmented reality (AR) games, enabling players to interact with in-game objects using natural hand gestures. It can enhance immersive experiences in virtual worlds and interactive storytelling applications. Architects and engineers can use Air Pose

Canvas to sketch building designs, manipulate 3D models, and visualize architectural plans in real time. It aids in collaborative design reviews and presentations. Surgeons and medical professionals can use this technology for surgical simulation and planning, allowing them to practice procedures and explore patient-specific anatomical models. It can be used for physical therapy exercises and rehabilitation programs.

IV. LITERATURE SURVEY

Sayli More et al. (2022) proposed an Air Canvas, which develops a concept or presents an idea in the real world. This is past the customary void (white), rectangular, and level dimensional material seen in customary data representation. Dayanand G. Savakar et al. (2022) proposed that Nand gestures, the key component of communication since the creation of epoch hand gestures, are the foundation of sign language, a visual form of communication. In order to identify the many hand gestures used for fingerspelling in sign language, this paper uses convolution neural networks on the datasets (CNN). Anuradha Kanade et al. (2020) define a paper focusing on the advanced study of gesture control-based robots. The user-centric depiction of the features has been provided using face detection, and an efficient classifier is erudite for categorizing between gestures. A number of hand gesture recognition technologies and applications for human-vehicle interaction are also proposed. Akhil Dixit et al. (2019) derive that a gesture can be recognized as a form of nonverbal communication in which certain bodily actions represent and communicate particular messages. The hand movements are easily deciphered to scrupulous, universal communications. This research presents a model system allowing differently capable people to communicate effectively with their counterparts. The overall description of the air canvas using the latest technology can be seen in the literature predicted by the authors, which provides a specific platform for users to communicate with others and learn natural language processing to enhance the technology and connect with the world.

V. PROPOSED WORK

The proposed work presents an integrated approach and system architecture that merges hardware components, software modules, and machine learning algorithms to facilitate natural hand gesture interactions with digital content. The methodology encompasses several key steps. Firstly, a substantial dataset of annotated images or videos encompassing various hand poses and gestures is compiled, serving as the foundation for training and validating the machine learning models. Subsequently, a deep learning model tailored for hand pose estimation is developed, leveraging architectures such as Convolutional Neural Networks (CNNs), PoseNet, or OpenPose. This model undergoes training on the collected dataset to accurately predict the 3D positions of pivotal hand joints or landmarks. In tandem, a gesture recognition model is implemented to interpret sequences of hand poses, discerning specific gestures or actions. This model is trained to associate these sequences with user-defined commands or actions. Cameras or depth sensors (e.g., RGB, depth, or infrared) are deployed and integrated into various devices such as smartphones, tablets, VR headsets, or standalone camera systems to capture user hand movements and gestures. Real-time image processing techniques isolate the user's hand from the background, generating a clean input stream for the machine learning models. Further processing includes filtering, smoothing, and normalizing hand pose data to enhance accuracy.

The trained hand pose estimation model is then applied to the preprocessed input data, continuously estimating the positions of the user's hand joints in 3D space. An interaction engine is implemented to map recognized gestures and hand movements to specific actions or interactions with digital content. This engine can simulate diverse interactions such as virtual drawing, painting, 3D manipulation, or others based on recognized gestures. Finally, the rendered digital content, including canvases, 3D models, or virtual environments, is displayed in real time based on the user's interactions on the user's device or within a VR/AR headset. The architecture of Air Pose Canvas comprises multiple components and stages, orchestrating seamless hand pose estimation and interaction with digital content, as delineated in Figure 1.

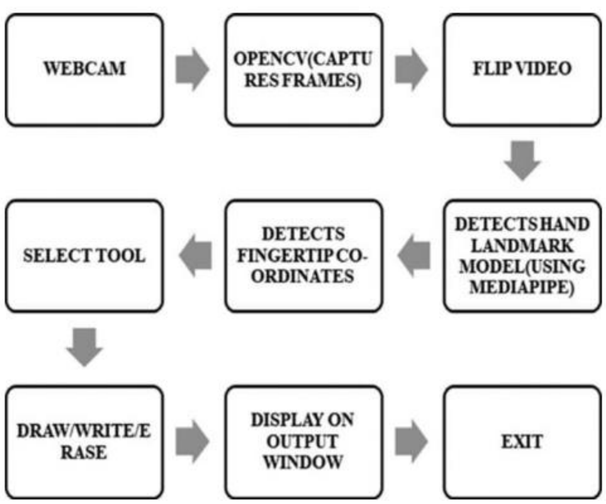

Figure 1. Represent System Architecture

Figure 1 delineates the system architecture, wherein users engage with the system via input devices equipped with sensors, such as depth-sensing cameras like Kinect, stereoscopic cameras, or specialized gesture recognition hardware. This module is tasked with processing input data from the sensors, estimating the positions and orientations of the user's hands in real-time, typically employing deep learning models like convolutional neural networks (CNNs) or pose estimation networks trained to discern hand gestures. Subsequently, the system interprets the estimated hand poses, identifying specific gestures or movements facilitated by gesture recognition algorithms. The interaction engine then processes these recognized gestures, translating them into actions within the digital environment, enabling interaction with digital content like drawing, object manipulation, or menu navigation. The digital canvas is the virtual space for user interaction, adaptable to different applications, including 2D drawing, 3D modelling, or immersive VR/AR environments. The user interface (UI) component complements this by providing a user-friendly interface for controlling and customizing the Air Pose Canvas system, integrating features such as menus, settings, and feedback mechanisms. In 3D environments, a rendering engine renders the digital content, encompassing objects, textures, lighting, and shadows, to ensure a realistic visual experience. Additionally, the system offers feedback to users, confirming actions, highlighting selections, or providing guidance through various means, such as visual, haptic, or auditory feedback.

VI. WORKING OF FINGERTIP MODEL

The functioning of the fingertip model is pivotal in accurately determining the positions of fingertips on a user's hand, thereby facilitating precise tracking of fine-grained movements and gestures. These movements are subsequently translated into interactions with the digital canvas. Initially, the fingertip model commences by gathering data, typically consisting of images or depth maps of the user's hand captured by sensors or cameras like depth sensors such as Kinect or RGB cameras. The collected data undergoes preprocessing steps to enhance its quality, including noise reduction, image enhancement, and resizing. The fingertip model is used for landmark detection once the hand is detected within the input data through hand detection algorithms. Trained using architectures like Convolutional Neural Networks (CNNs), this model identifies key landmarks on the hand, including fingertip positions.

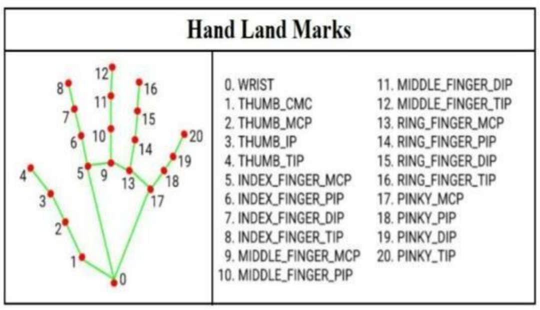

Figure 2. Represent Hand Land Marks

Moreover, it may also estimate the overall pose of the hand, encompassing orientation and finger joint angles, thereby enabling a more comprehensive understanding of the hand's configuration for improved tracking and gesture recognition. Subsequently, the estimated fingertip positions and hand pose information are utilized for gesture recognition, interpreting the relative positions of fingertip landmarks and the hand's pose to identify specific gestures or movements. Machine learning models or rule-based algorithms are typically employed for this purpose, associating gestures with corresponding actions, such as selecting or panning. Finally, the recognized gestures and hand movements are translated into interactions with the digital canvas, facilitating actions such as drawing or zooming. Continuous processing of incoming data by the fingertip model ensures real-time feedback to the system, enabling seamless interaction with the digital canvas. In summary, the fingertip model within Air Pose Canvas is instrumental in accurately detecting and tracking fingertip positions, thus enabling natural and intuitive user experiences. The model's accuracy and responsiveness are paramount for the overall performance and usability of the system.

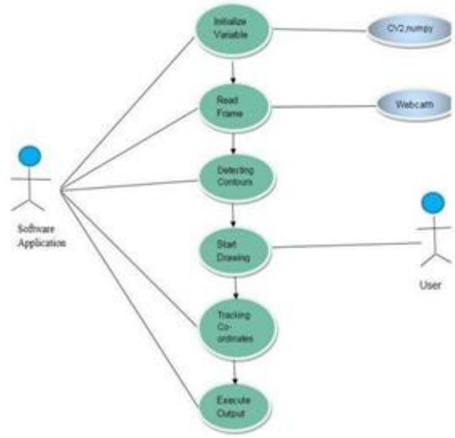

Figure 3 shows the Use Case Diagram.

Figure 2 depicts some common hand landmarks typically detected in hand landmark models, including fingertip positions, finger joints, palm centre, hand base, thumb connection point, and little finger connection point. These models are trained using deep learning techniques like CNNs and key point regression networks, producing landmark coordinates based on input images or video frames. The precision of landmark detection is critical for applications such as hand pose estimation for virtual reality, augmented reality, and gesture-based interfaces. The "User" represents individuals interacting with the Air Pose Canvas system in the use case diagram. The relationship between the "User" actor and the use cases is that the "User" actor is associated with all the use cases since users initiate and interact with each of these functionalities. The Air Pose Canvas with Hand Pose Estimation and Machine Learning is an advanced system that enables users to interact with digital content using hand gestures. Its functionality integrates computer vision, machine learning, and real-time hand pose estimation. The system relies on input from depth sensors or cameras, which capture the user's hand movements and gestures in a three-dimensional space.

ALGORITHM

Step 1. The sensors or cameras capture the user's hand movements and gestures in real-time.

Step 2. The data is processed by the Hand Pose Estimation Module, which estimates the 3D pose of the hands.

Step 3. The Gesture Recognition Module interprets the hand poses to identify specific gestures or movements.

Step 4. The Input Manager translates these recognized gestures into actions within the digital content environment.

Step 5. The Digital Content Renderer displays the digital canvas or environment, allowing users to interact with it using their hands.

Step 6. The User Interface component provides visual feedback and interaction options to the user.

Step 7. The Machine Learning Model Manager ensures that the machine learning models run smoothly and can be updated as needed.

Step 8. Optionally, networking capabilities can support collaborative interactions or data sharing.

Commonly used depth sensors include Microsoft's Kinect or structured light sensors. The depth data obtained from these sensors undergoes processing to create a point cloud or depth map representing the user's hand and its movements. This processing involves filtering, noise reduction, and background removal to isolate the hand from the environment. Subsequently, the processed depth data is utilized to detect and segment the user's hand from the background, typically employing hand detection algorithms to identify the hand's position and shape within the point cloud or depth map. At the system's core lies hand pose estimation, where the precise positions and orientations of the user's fingers and hand joints are determined in real-time. Deep learning models, such as convolutional neural networks (CNNs) or graph neural networks (GNNs), are trained to recognize and estimate hand poses accurately. The system interprets the estimated hand pose to recognize specific gestures and movements, employing gesture recognition algorithms to analyze the positions and trajectories of fingers and hand joints. Recognized gestures allow users to interact with the digital canvas, enabling actions like drawing, sculpting, manipulating 3D objects, or performing other tasks based on the recognized gestures. Real-time visual feedback is provided to users, displaying their hands and gestures on the screen or in a virtual environment, aiding their understanding of how the system interprets their hand movements. Users can customize their interaction experience by selecting different drawing tools, adjusting brush sizes or colours, and configuring other settings according to their preferences. They can utilize the system to create digital art, 3D models, educational content, or any other digital content requiring precise hand interaction. The system records and translates their gestures into digital strokes, shapes, or actions, allowing users to save or export their creations in various formats and share them on social media or with others. Continuous tracking of the user's hand movements and gestures ensures a real-time and responsive interaction experience, facilitating fluid and natural gestures. Security and privacy measures may be implemented depending on the application to protect user data and ensure secure interactions. The Air Pose Canvas with Hand Pose Estimation and Machine Learning amalgamates cutting-edge technologies, seamlessly bridging the physical and digital realms. It empowers users to express themselves creatively, interact with digital environments naturally, and open new avenues for creativity, accessibility, and human-computer interaction. Depending on the application, security and privacy measures may be implemented to protect user data and ensure secure interactions. The Air Pose Canvas with Hand Pose Estimation and Machine Learning is a fusion of cutting-edge technologies that seamlessly bridge the physical and digital worlds. It empowers users to express themselves, create content, and interact with digital environments using natural hand gestures, opening up new possibilities for creativity, accessibility, and human-computer interaction.

VII. CONCLUSION AND FUTURE SCOPE

In conclusion, the Air Pose Canvas with Hand Pose Estimation and Machine Learning is a testament to the boundless potential of merging cutting-edge technology with human creativity. It empowers users to interact with digital content naturally and artistically, making it a promising innovation with broad-reaching implications for numerous domains. As technology advances, we can anticipate an exciting future where the boundaries of digital expression are continually pushed, thanks to innovations like this. The Air Pose Canvas represents the future of human-computer interaction, where physical gestures and movements become powerful tools for digital expression. Its evolution may lead to even more sophisticated applications and creative possibilities.

REFERENCES

[1]. Sayli More, Prachi Mhatre, Shruti Pakhare, Surekha Khot (2022) "Air Canvas: Draw in Air" International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 02 | Feb 2022
[2]. Dr. Dayanand G. Savakar, Ayisha Daraveshi, Danesh Telsang (2022) "An Improved Method for Hand Gestures Recognition and Character Identification Using CNN Classifiers", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 11 Issue 09, September 2022
[3]. Abhik Swarnakar, Dr. Anuradha Kanade, Sonu Kumar Soni, Ashish Pal (2020) "Hand Gestures Recognition and Finger Count", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 ICSITS - 2020 Conference Proceedings
[4]. Akhil Dixit, Arush Agarwal (2019) "Hand Gesture Recognition System Using Computer Vision and Neural Networks", International Journal of Scientific & Engineering Research ISSN 2229-5518 Volume 10, Issue 6, June-2019
[5]. M. M. Hasan, & P. K. Mishra, (2012) "Improving Morphology Operation for 2D Hole Filling Algorithm", International Journal of Image Processing (IJIP), Vol. 6, No. 1, pp. 635-646.
[6]. T. R. Trigo & S. R. M. Pellegrino (2010) "An Analysis of Features for Hand-Gesture Classification", 17th International Conference on Systems, Signals and Image Processing (IWSSIP 2010), pp 412- 415.

[7]. M. M. Hasan & P. K. Mishra (2011) "HSV Brightness Factor Matching for Gesture Recognition System", International Journal of Image Processing (IJIP), Vol. 4, No. 5, pp 456-467. [4].

[8]. Luigi Lamberti & Francesco Camastra (2011) "Real-Time Hand Gesture Recognition Using a Color Glove", Springer 16th international conference on Image analysis and processing: Part I (ICIAP'11), pp 365-373.