

Augmented Reality by Hand Gesture Recognized Commands and Movements

Mehdi Valinejad, *Student, IEEE*

Abstract—The abstract goes here.

Index Terms—Computer Vision, HCI, Hand Gesture Recognition, Augmented Reality, ARUco Markers, Landmark, MediaPipe.

I. INTRODUCTION

THE Human-Computer Interaction (HCI) represents a nuanced and interdisciplinary domain that intricately investigates the symbiotic relationship between humans and computer systems, with a primary focus on optimizing the design and utilization of technological interfaces. At its core, HCI delves into the intricacies of how individuals interact with computing devices, aiming to refine and elevate these interactions to levels of heightened effectiveness, efficiency, and overall user satisfaction.

Hand gesture recognition plays a significant role in Human-Computer Interaction (HCI) by providing a natural and intuitive means for users to interact with computers and devices. The relationship between hand gesture recognition and HCI lies in the application of gesture-based interfaces to enhance the way users communicate with and control technology.

Augmented Reality (AR) is a technology that overlays digital information and virtual objects on the real-world environment, creating an augmented view for users. Unlike virtual reality, which immerses users in a completely computer-generated environment, augmented reality enhances the real-world environment by adding computer-generated perceptual information.

The relationship between Human-Computer Interaction (HCI) and Augmented Reality (AR) is profound, as AR technologies heavily rely on effective HCI principles to deliver immersive and user-friendly experiences. Some of their interconnection aspects can be noted as User Interface and User Experience Design, Motion and Gesture Interaction, Spatial Awareness, Usability in 3D space, Feedback and user guidance, Context-Aware Computing, Accessibility in AR.

In recent years, the integration of cutting-edge technologies has propelled Human-Computer Interaction (HCI) to new heights, revolutionizing the way users engage with digital content. One captivating synergy within this realm lies at the intersection of Hand Gesture Recognition (HGR) and Marker-Based Augmented Reality (AR). This amalgamation presents a promising frontier for creating intuitive and immersive interactive experiences, where users can seamlessly navigate and manipulate augmented environments through natural hand gestures.

Hand gesture recognition, a pivotal facet of HCI, enables users to communicate with computing systems using natural

and instinctive movements. Leveraging advanced computer vision algorithms, these systems interpret and respond to the intricacies of hand gestures, introducing an unparalleled level of intuitiveness to human-computer interactions. Concurrently, Marker-Based Augmented Reality employs identifiable markers, such as QR codes or images, as triggers to overlay digital content onto the physical world. This technology augments reality by blending virtual elements with the user's immediate environment.

The convergence of hand gesture recognition and marker-based AR holds great promise for advancing the field of HCI. By integrating gesture recognition into AR interfaces, users can interact with digital content in a manner that mirrors real-world actions, fostering a more intuitive and engaging experience. This marriage of technologies addresses the limitations of traditional input methods, offering a more natural, efficient, and hands-free mode of interaction within augmented environments.

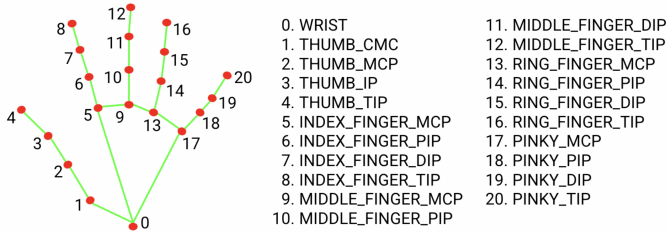
This paper explores the synergy between hand gesture recognition and marker-based AR, examining the implications for user interface design, interactive applications, and the overall user experience. We delve into the technical intricacies of combining these technologies, exploring how gesture-driven interactions can enhance the usability and accessibility of marker-based AR systems. Moreover, we investigate the potential impact on diverse application domains, from gaming and education to industrial training and collaborative workspaces.

As we navigate through the intricate landscape of this technological convergence, we seek to unravel the synergies and challenges that arise when marrying hand gesture recognition with marker-based AR. By doing so, we aim to contribute to the ongoing discourse in HCI, providing insights that inform the design and development of more natural and immersive interfaces for augmented reality environments. The journey into this fusion promises to unlock novel possibilities, shaping the future landscape of interactive technologies and user experiences.

II. PLANNED METHODS AND TOOLS

1. Landmark Detection

Landmark detection, also known as keypoint detection, refers to the process of identifying and locating specific points or features within an image or a visual scene. These points are often distinctive and can be used as references or landmarks for various computer vision tasks. Landmark detection is a crucial step in understanding and analyzing images, enabling applications in fields such as facial recognition, pose estimation, object tracking, and augmented reality [1].



The list of detected point of the hand by landmark hand detector.

Fig. 1. Landmark hand gesture detected joints

MediaPipe utilizes a specific model called the MediaPipe Hands for hand tracking, including hand landmark detection. The algorithm for hand landmark detection in MediaPipe Hands is based on a convolutional neural network (CNN) architecture.

The hand landmark model in MediaPipe can detect and track 21 3D landmarks on each hand, including key points such as the tips of the fingers, the base of the palm, and points along the hand contours. This enables precise tracking and recognition of hand gestures in real-time video or image streams.

2. ArUco markers

ArUco markers are a type of augmented reality marker designed for computer vision applications, particularly for tasks such as camera calibration and augmented reality (AR) systems. These markers are black-and-white square patterns that are easy to detect and identify in images or video frames. using markers we are able to detect critical points in an video stream frame, generally, markers are used for Visual Pattern Recognition, Unique identification, camera calibration and etc.

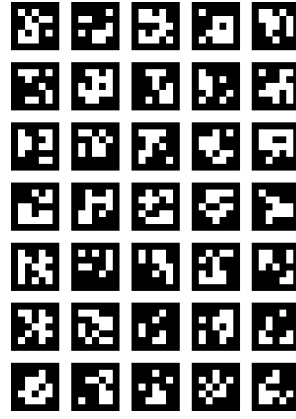
OpenCV (Open Source Computer Vision Library) provides a dedicated module for ArUco marker detection, and the implementation is based on the ArUco library. The ArUco module in OpenCV is specifically designed for easy integration of ArUco markers into computer vision applications.

3. Homography Finding

To estimate the transformation matrix (homography matrix) that relates corresponding points in two images of the same scene. OpenCV offers builtin homography detection methods that can be used for perspective point matching between real image and the way of residing in the the parent image visible area.

III. EXPECTED RESULTS

The expected result of landmark-based hand gesture detection combined with augmented reality (AR) can be a dynamic and interactive user experience where real-world hand movements are used to control and interact with virtual elements. Here's an overview of the potential outcomes:



Sample chart of ArUco Markers.

Fig. 2. ArUco Markers

Real-Time Hand Gesture Detection

Accurate and real-time detection of hand gestures using landmarks provides a natural and intuitive way for users to convey commands or interact with digital content.

Precise Tracking of Hand Poses

Landmark detection allows for precise tracking of the positions and orientations of key points on the hand, enabling accurate recognition of various hand poses and movements.

Virtual Object Manipulation

Users can interact with virtual objects overlaid on the real-world environment using their hand gestures. This could include grabbing, moving, rotating, or resizing virtual elements.

Gesture-Based Controls

Hand gestures can serve as a gesture-based control interface for AR applications. For example, a specific gesture might trigger an event, change the color or view, or initiate a specific action.

Immersive AR Experiences

Landmark-based hand gesture detection enhances the immersion of AR experiences by allowing users to engage with the virtual content in a more natural and hands-on manner.

Natural Interaction in AR Painting

In AR painting game, users can control characters or perform in-game actions using hand gestures detected by landmarks. This adds a layer of physicality and engagement to the painting experience.

Overall Expectation

The expected result is an interactive and immersive AR experience where users can use their hands and gestures to manipulate digital content, enhancing the naturalness and engagement of human-computer interactions in augmented reality.

IV. CONCLUSION

In this paper, we have explored the synergies of landmark detection, ArUco markers, and hand gesture recognition in the context of webcam video streams, leveraging the capabilities of the OpenCV library. The fusion of these technologies has paved the way for a multifaceted approach to enhancing Human-Computer Interaction (HCI), providing users with a more intuitive and immersive experience in real-time video environments.

The integration of landmark detection serves as the backbone of our approach, enabling precise tracking and recognition of key points on the human hand and face. This information is then seamlessly combined with the distinctive properties of ArUco markers, which act as visual anchors for accurate pose estimation and perspective transformations within the webcam video stream. Together, these components form a robust foundation for spatial awareness and object tracking in dynamic scenes.

Adding another layer of sophistication, hand gesture recognition brings a natural and gestural interface to our HCI paradigm. The ability to interpret and respond to hand movements in real time enhances the user's ability to interact with digital content in an instinctive and hands-on manner. Gestures such as Thumb Up/Down, Like/Unlike, Victory Sign can be translated into meaningful commands, expanding the range of possibilities for interactive applications within the webcam video stream.

Through our exploration, we have witnessed the potential applications of this fusion across diverse domains. In educational settings, users can manipulate 3D models or engage with virtual simulations through hand gestures and markers. In gaming environments, the combination of landmark detection and hand gesture recognition provides an immersive and interactive gaming experience. Moreover, the integration of ArUco markers facilitates accurate registration of virtual objects within the physical space captured by the webcam.

The significance of our approach extends to accessibility considerations, as gestural interfaces offer alternative interaction methods for individuals with physical disabilities. Furthermore, the real-time nature of our system enhances its applicability in various contexts, from augmented reality to video conferencing, where natural and expressive gestures can convey a richer form of communication.

Looking ahead, the fusion of landmark detection, ArUco markers, and hand gesture recognition in webcam video streams opens avenues for further research and development. Improvements in algorithmic efficiency, robustness, and the exploration of deep learning techniques may contribute to even more seamless and accurate interactions. Additionally, the integration of haptic feedback or voice commands could further enhance the multimodal nature of our HCI system.

In conclusion, our work signifies a step forward in the evolution of interactive experiences within webcam video streams. By combining landmark detection, ArUco markers, and hand gesture recognition with OpenCV, we have created a versatile and accessible example that empowers users to interact with digital content in a natural and engaging manner.

This fusion not only enriches HCI paradigms but also paves the way for innovative applications that bridge the gap between the virtual and physical worlds. As technology continues to evolve, so too will the opportunities to redefine how we engage with and perceive the digital landscape.

REFERENCES

- [1] F. Lin, B. Price, and T. Martinez, "Ego2hands: A dataset for egocentric two-hand segmentation and detection," 2021.

Mehdi Valinejad (S'12) received the B.S. degree in industrial engineering from the Azad University (South Tehran Branch) in 2012, and is currently working Master's. degree at the University of Bahcesehir at Istanbul.