ARFusion : Real Time Aruco Marker Augmentation.

Prateek Manoj Mhatre
Roll No: 35
Reg No: 220962158

Mustafa Murtaza Haji
Roll No: 22
Reg No: 220962076

Krish Didwania
Roll No: 39
Reg No: 220962244

***Abstract*—This study investigates the use of augmented reality (AR) in real-world settings by superimposing images over ArUco markers via a live camera stream using OpenCV and ArUco marker recognition. The technique ensures accurate and durable picture overlays, even in difficult situations like changing viewing angles or partial marker occlusions, by combining feature matching with homography-based warping. The AR overlays' alignment and stability are preserved by this dual-technique adaption. The system's usefulness extends across a number of applications, including dynamic object manipulation, AR-enhanced education, and interactive gaming, providing a fresh way to enrich visual experiences. The video overlay on a screen of the desired size is adjusted in this study using 6x6 fiducial markers based on marker spacing. Additionally, 4x4 markers make it easier to manipulate videos in many ways, such as up, down, and and edit functions, while 5x5 markers enhance the precision of object placement within AR environments.**

***Index Terms*—Augmented Reality (AR), OpenCV, ArUco Marker Detection, Feature Matching, Homography, Real-time Image Overlay**

I. Introduction

In order to improve a user's perception and engagement with their surroundings, computer-generated content is superimposed onto real-world situations in augmented reality (AR), which is a merging of the digital and physical worlds. A real-time direct or indirect view of a physical world improved by the addition of virtual computer-generated information is called augmented reality (AR). This integration is both interactive and registered in 3D, merging real and virtual objects seamlessly. According to the Reality-Virtuality Continuum introduced by Paul Milgram and Fumio Kishino, AR falls on a spectrum between the real environment and the purely virtual environment, positioning closer to the real world than to complete virtuality. This framework helps in understanding the varying degrees of augmented experiences, from slight enhancements of reality to fully immersive virtual environments.

With the rapid advancement of computer vision technologies, AR has found applications across various domains, including gaming, education, healthcare, and industrial design. This study leverages the capabilities of OpenCV and ArUco markers to create a robust AR system that enables real-time image and video overlays onto physical surfaces, thereby enriching the visual experience.

The use of ArUco markers, which are binary square fiducial markers that are simple to detect and individually identifiable in video streams, forms the basis of this investigation. These markers provide reference points for the positioning of digital content, acting as anchors in the real world. The goal of the project is to flawlessly superimpose a pre-made image or video in real-time onto an ArUco marker, with the overlay dynamically adjusting to the orientation and position of the marker. This is accomplished by combining feature matching methods with homography-based modifications, which guarantee that the digital content stays precisely aligned with the marker even in the face of difficult circumstances like changing lighting, angles, and partial occlusions.

Fig. 1. Milgram's Reality-Virtuality Continuum. The continuum spans from the real environment to a fully virtual environment, illustrating where augmented reality and augmented virtuality lie in relation to these extremes.

The incorporation of AR with video playback offers a revolutionary method of media consumption. AR makes it possible to create virtual screens that improve immersion and engagement by allowing videos to be projected onto surfaces in the viewer's surroundings. In addition to making watching more interesting, this technology gives us the freedom to turn any room into a dynamic viewing environment, completely changing the way and location of how we watch videos.

In domains where precise visualization is critical, the strategic placement of 3D objects with AR technology is essential. AR offers a way to put intricate 3D models in a real-world setting, giving businesses like interior design and architecture a realistic sneak peek at completed projects and interior layouts before actual work starts. By eliminating mistakes, cutting expenses, and maximizing both practical and aesthetic choices, this AR application greatly enhances project planning and client pleasure.

All things considered, the developments in AR made possible by the accuracy of ArUco markers not only make complicated AR applications more feasible, but they also broaden the use of AR technology and push the limits of how digital and physical realities may coexist to produce rich, engaging experiences. This study demonstrates the usefulness and potential of AR to revolutionize both professional and daily life.

II. Fiducial Markers in Augmented Reality

Because they offer trustworthy reference points that enable exact camera position estimation and accurate object localization, fiducial markers are essential in augmented reality (AR) and other computer vision applications. Computer vision algorithms identify these markers' unique, instantly identifiable patterns, which enable systems to ascertain the camera's orientation and position in relation to the marker. ArUco markers, specifically, are designed for quick, robust detection even in dynamic, real-time environments, making them ideal for AR applications such as overlaying virtual objects, video playback control, and interactive object placement. By serving as a consistent point of reference, these markers enhance the accuracy and responsiveness of AR experiences, helping to bridge the virtual and real worlds in various applications, from gaming to industrial design and education.

- **ArUco Markers**: Optimized for high-speed detection in real-time AR applications; widely used in robotics for quick camera calibration.
- **AprilTags**: Known for robust detection in varied lighting and motion conditions, often used in industrial AR setups and outdoor robotics.
- **QR Codes**: Primarily designed to store data such as URLs or text; commonly used in applications where data retrieval is prioritized over precise positioning.
- **ChArUco Boards**: A combination of ArUco markers and chessboard corners that enhances accuracy in corner detection, frequently used in camera calibration tasks.
- **RUNE-Tags**: Designed for high accuracy in occlusion-prone environments, ideal for applications requiring precise spatial responses under partial visibility.

We have used ArUco tags in this project to enable accurate camera position estimate, which makes it possible to insert virtual objects and videos in real-world settings with precision.

A. ArUco Markers

The design of ArUco markers, a popular kind of fiducial marker, is straightforward but efficient: a black square border with a special binary matrix in the middle. Computer vision algorithms can identify and decode the encoded data in this matrix, which is represented as a binary pattern. Once detected, the software can identify each unique marker and estimate its position and orientation (pose) relative to the camera, enabling precise spatial tracking.

The robust design of ArUco markers provides strong resilience against partial occlusions, such as when a part of the marker is obstructed by another object. Additionally, their high contrast and square shape allow for rapid and reliable detection, even in changing or dynamic environments where lighting and movement can vary significantly. This makes them especially suitable for applications that require real-time tracking, such as augmented reality (AR), robotics, and drone navigation, where accurate pose estimation and stability are essential for effective performance.

1) Detection and Decoding Process: Detection of ArUco markers is typically performed using the following steps:

1) **Identification of marker outlines using adaptive thresholding**: Adaptive thresholding is applied to distinguish high-contrast regions, allowing for the isolation of potential marker outlines in various lighting conditions. This process identifies the square shapes of ArUco markers by detecting contours and verifying the presence of a solid black border surrounding a binary matrix.
2) **Decoding the binary patterns within the detected markers**: Once markers are detected, the binary matrix within each marker is read and decoded. Each binary pattern is unique, allowing the system to determine the marker's ID and access the specific information it encodes, enabling accurate identification and further calculations, such as pose estimation.

To estimate the position of the detected ArUco marker relative to the camera, a mathematical model is used to transform the 3D world coordinates of the marker to 2D image coordinates as follows:

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where:

- s: The scaling factor that adjusts the coordinates based on the marker's distance from the camera.
- K: The camera's intrinsic parameters matrix, containing values such as focal lengths and optical center offsets, which reflect the camera's internal properties.
- $[R|t]$: A combination of the rotation matrix R and the translation vector t. The rotation matrix R defines the orientation of the marker relative to the camera, while the translation vector t specifies its position in the camera's coordinate frame.
- (X, Y, Z): The 3D coordinates of a point on the marker in real-world space.
- (x, y): The 2D coordinates of the corresponding point in the camera's image plane, where the marker appears.

This transformation model enables precise pose estimation, allowing augmented reality applications to accurately overlay virtual objects on the detected marker in real-world scenes.

2) Application in Video Playing and Object Placement: ArUco markers can be used to anchor digital content in physical areas during video playback. For example, by aligning a video with an ArUco marker, it can be superimposed on a particular spot or item in the actual world. This approach allows the video content to stay in place relative to the marker, even as the camera moves, providing a seamless augmented reality experience. Such functionality enhances interactivity by enabling users to view media in context with their surroundings, making it particularly useful for interactive installations or museum exhibits where content is tied to physical objects.

Fig. 2. Kinds of Fiducial Markers

For object placement, ArUco markers facilitate precise and real-time positioning of 3D models in an AR environment. When a marker is detected, the system can accurately overlay virtual objects at its location, aligning with the user's perspective. This feature is invaluable in applications like virtual furniture arrangement for interior design, where users can visualize how furniture will look in a room before purchasing. Additionally, in educational settings, ArUco markers can be used to create interactive learning tools, where students manipulate virtual objects, such as molecular models or historical artifacts, within the classroom. This makes learning more engaging and visually informative, as students can interact directly with augmented content in a hands-on manner.

3) Advancements and Future Prospects: Recent advancements in machine learning and computer vision have led to improvements in the detection algorithms for ArUco markers, enhancing their accuracy and reducing their susceptibility to environmental factors such as lighting conditions and occlusions. Ongoing research continues to expand their applicability in more complex AR systems, promising even more seamless integration of AR technologies into everyday life.

III. OpenCV Functions for Tag Detection and Augmented Reality

A. Overview

For creating applications requiring tag identification and augmented reality, OpenCV offers a strong collection of functions. From simple image processing to intricate transformations and AR overlays, these features make a wide range of activities easier.

B. Key Functions

1) cv::aruco::detectMarkers: The `detectMarkers` function is a part of the ArUco module in OpenCV, designed to identify fiducial markers within an image. It processes an input image to locate markers and returns their IDs and the coordinates of their corners.

- **Parameters:** This function takes the grayscale image, the dictionary specifying the type of markers to look for, and parameters that fine-tune the detection process.
- **Usage:** It is crucial for initializing the tracking of objects in AR applications, allowing the system to understand where virtual objects should appear in the real world.

2) cv::aruco::estimatePoseSingleMarkers: Once markers are detected, `estimatePoseSingleMarkers` calculates the 3D pose of each marker relative to the camera. This function is vital for AR, as it determines how objects should be rotated and scaled in 3D space to align correctly with the physical world.

- **Parameters:** It requires the corners of the marker, the size of the marker, camera calibration parameters (camera matrix and distortion coefficients), and outputs rotation and translation vectors.
- **Usage:** Essential for placing 3D models in an AR scene accurately.

3) cv::perspectiveTransform: This function applies a perspective transformation to a set of points. In the context of AR, `perspectiveTransform` is used to adjust the coordinates of a virtual object so that it fits the perspective of the viewing camera, based on the pose estimation.

- **Parameters:** It takes points and the transformation matrix derived from the real-world coordinates of the markers.
- **Usage:** It aligns the virtual and real-world coordinates, enabling realistic integration of virtual content into the live video feed.

4) cv::warpPerspective: `warpPerspective` warps an image using a given transformation matrix. This function is particularly useful in AR for rendering images or videos onto detected markers.

- **Parameters:** Requires the source image, output matrix size, and the transformation matrix.
- **Usage:** Used to overlay and adjust images or videos onto surfaces defined by markers, crucial for immersive AR experiences.

IV. Methodology

The first step in the study is to use OpenCV's `cv2.aruco` module to identify ArUco markers in the video feed. Accurate digital content overlay is made possible by ArUco markers, which offer a strong basis for identifying reference locations in the video. The system determines the homography matrix, a transformation matrix that is used to warp the intended picture or video material onto the marker's location within the video feed, after a marker has been recognized. Even when the camera perspective shifts, this transformation makes

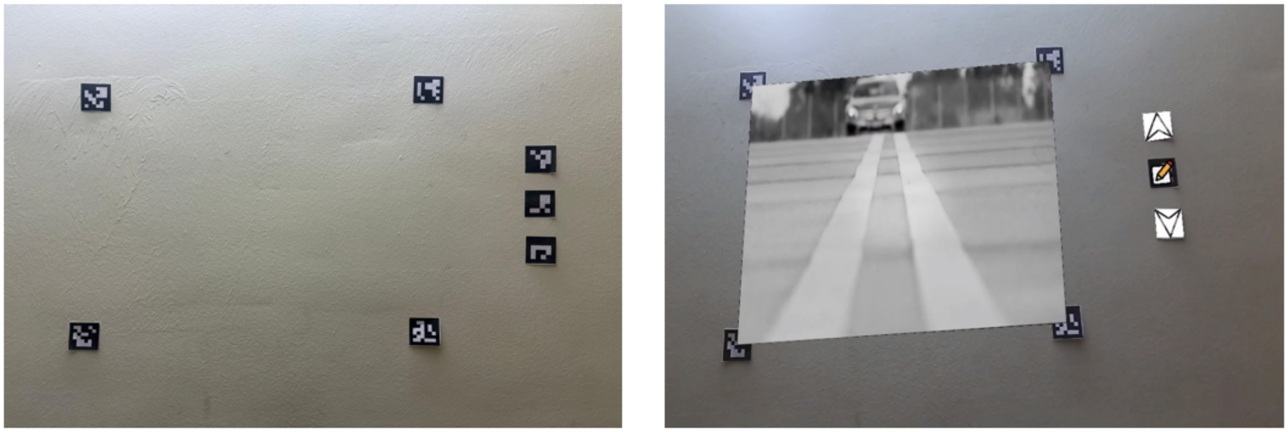

Fig. 3. The ArUco markers placed on a wall, configured for augmented reality tracking. (Right) A video dynamically overlaid onto the wall using the positions of the ArUco markers, demonstrating real-time video mapping in an augmented reality application.

sure that the overlaid content stays appropriately scaled and positioned in reference to the marker. To enable a wide range of applications, the project incorporates cutting-edge video and object interaction capabilities employing ArUco markers. Real-time effect application, interactive video navigation, and accurate virtual object presentation in an augmented reality environment are all made possible by these markers. The system's ability to track and interact with markers enhances user engagement by providing a seamless AR experience where digital content is contextually anchored to the physical environment.

- **Marker Detection**: The process begins with the detection of ArUco markers in the video feed. ArUco markers are detected using OpenCV's cv2.aruco module, which provides functions to detect these fiducial markers and estimate their pose. Upon detection, the corners of the marker are identified, which serve as the reference points for subsequent transformations. The accuracy of this detection phase is crucial, as any errors in identifying the marker's position and orientation could lead to misalignment of the overlay.
- **Homography Calculation**: Once the marker is detected, the system calculates the homography matrix, a transformation matrix that maps the four corners of the marker in the physical world to the corresponding corners in the overlay image. The homography matrix is essential for warping the overlay image to fit precisely onto the marker, even as the marker moves or rotates within the video feed. This transformation ensures that the overlay remains anchored to the marker's position, providing a seamless AR experience.
- **Feature Matching**: While homography provides a basic alignment of the overlay, additional refinement is often necessary to ensure that the overlay remains accurate under varying conditions, such as when the marker is viewed from an angle or is partially obscured. This is where feature matching comes into play. The system detects keypoints and descriptors in both the ArUco marker and the overlay image using algorithms like SIFT (Scale-Invariant Feature Transform) or ORB (Oriented FAST and Rotated BRIEF). By matching these features, the system can adjust the overlay dynamically, correcting any distortions that may occur due to changes in perspective.
- **Image Overlay** : The final stage involves warping the overlay image using the homography matrix and applying the feature-based corrections derived from the matching process. This ensures that the overlay remains stable and accurately aligned with the marker in real-time. The overlaid image or video is then displayed on the marker within the video feed, creating the illusion of digital content seamlessly integrated into the physical environment.

A. Video Manipulation with ArUco Markers

1) Multiple Marker Tracking for Video Playing: In this project, we utilize 6x6 ArUco markers to define the vertices of a virtual screen, onto which videos are dynamically displayed. These markers are tracked in real-time, allowing users to adjust the position and scale of the projected video interactively. As the markers move, the system continuously calculates the homography matrix to adjust the video frame to fit within the area defined by the markers. This transformation ensures that the video display is correctly aligned and scaled according to the markers' position, even as the user changes the markers' orientation. The real-time tracking of these markers is essential for providing a smooth and immersive viewing experience, allowing users to seamlessly interact with the video content. The video display instantly adapts to any changes in the marker's position or orientation, offering a versatile and dynamic method of video projection that enhances interactivity and engagement.

2) Control Markers for Video Navigation and Effects: In addition to the 6x6 markers used for video playing, we designate 4x4 ArUco markers to control specific video functions and visual effects. These markers are tracked in real-time to trigger the following actions:

- One marker allows users to skip to the next video in the playlist.
- Another marker takes the user back to the previous video.
- A third marker enables the cycling of various video effects, such as grayscale, canny edge detection, Gaussian blur, and negative image filters.

To interact with these controls, users simply move the designated markers into view of the camera. The system detects these markers and triggers the corresponding function, enabling seamless video navigation or visual effect application in real time. This real-time marker tracking ensures that the system responds immediately to user input, making the interaction intuitive and efficient. By leveraging the precise detection of control markers, users can interact with video content dynamically, adjusting playback and applying effects without the need for traditional interfaces like remote controls or buttons.

B. Augmented Reality Object Display

Additionally, our solution incorporates the use of 5x5 ArUco markers to enable the augmented reality display of 3D objects from STL or OBJ files. Multiple objects can be placed into the physical environment as seen through the camera thanks to ArUco markers, which are linked to specific objects. This setup provides users with the ability to place objects in a real-world context, offering an immersive experience where digital content appears seamlessly integrated with the surroundings.

The system's capacity to scale the items on display in real-time in response to human contact with the ArUco markers is one of its primary characteristics. Because the object's size is dynamically changeable, users can see the 3D model at various scales to fit their own requirements or tastes. Additionally, the quantity of ArUco markers found in the surroundings is strongly correlated with the quantity of things shown. A matching 3D object can be rendered for every marker found, enabling flexible and adaptable interactions in multi-object configurations.

At this early level of development, we have used OpenCV-supported basic 3D models from STL or OBJ files to generate an artificial spider. Although this method illustrates the fundamental functionality, the project's future development intends to include realistic, high-quality 3D items, such vases or other complex models, for improved realism and display. These improvements will increase the system's potential for applications in fields requiring high-fidelity 3D representation, such as product visualization, virtual prototyping, and interior design.

This feature is especially helpful in applications like educational tools, where students can manipulate and interact with virtual objects to better understand complex concepts, or in design scenarios, like architecture or interior design, where users can see how various models or objects fit within a physical space. The ability to scale objects and add more based on the number of markers provides a dynamic and customizable augmented reality experience.

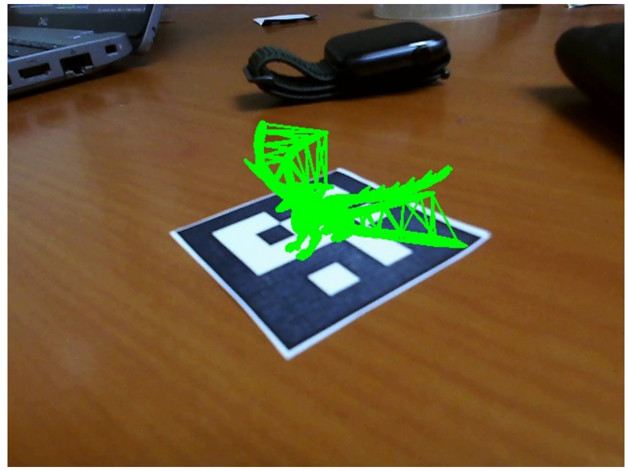

Fig. 4. A 3D display of a dragon from a STL file onto a 5x5 aruco tag

V. Salient Features

A. Real-Time Interaction and Feedback

- **Real-Time Processing:** The system operates in real-time, providing immediate visual feedback to user interactions. This feature is critical for applications where timing and responsiveness are crucial, such as in interactive gaming and live training simulations.
- **Immediate Response to Environmental Changes:** Our application dynamically adjusts to changes in the environment, such as variations in lighting and movement, ensuring consistent performance without lag.

B. Stability and Accuracy

- **Robust Overlay:** The combination of homography and feature matching ensures that the overlay remains stable even with changes in perspective, lighting, or partial occlusion. This robustness is essential for maintaining an immersive user experience in AR applications.
- **Enhanced Tracking Precision:** Employing advanced algorithms for marker detection, the system minimizes errors in tracking and overlay, which is particularly beneficial for precise tasks in educational and industrial AR applications.

C. Versatility and Scalability

- **Versatility:** The application can be adapted for various uses, from interactive gaming to AR-assisted learning and dynamic advertising. This adaptability allows it to be integrated into multiple industries including education, marketing, and entertainment.

- **Customizable Overlays:** Users can customize overlays and interactions, which enhances user engagement and allows for personalized experiences in applications like virtual showrooms and interactive learning environments.
- **Scalability:** The system is designed to scale seamlessly, supporting multiple markers and overlays simultaneously. This scalability facilitates the expansion of the application to larger environments and more complex scenarios without a loss in performance.

D. Innovative Applications

- **Interactive Gaming:** By integrating real-world objects into the gaming environment using AR markers, the system provides a unique gaming experience that blends physical and digital gameplay.
- **AR-assisted Learning:** The system enhances educational content by overlaying digital information onto physical objects and environments, making learning more interactive and engaging.
- **Dynamic Advertising:** Advertisers can create immersive and interactive campaigns that engage consumers directly by overlaying digital content onto physical spaces, such as shopping malls and public squares.

These features collectively enhance the functionality and applicability of the AR system, making it a versatile tool for both entertainment and professional applications.

VI. Conclusion

Using OpenCV and ArUco markers, this study effectively combines augmented reality (AR) with real-world settings to produce accurate and reliable real-time overlays of photos and videos. The incorporation of feature matching and homography-based transformations ensures that digital material faithfully adheres to the physical marks, managing challenges such as changing viewpoints and obstacles.

ArUco markers serve as a robust basis for these overlays, ensuring dynamic adaptability to variations in marker placement and orientation. The versatility of this approach is showcased in its broad range of applications, including interactive gaming, educational enhancements, and engaging advertising, all benefiting from the interactive, real-time capabilities of the system.

Additionally, the proposed framework can handle complex AR settings with many overlays and markers because it is versatile and scalable. This project highlights the revolutionary potential of AR in diversifying visual interactions across several sectors and not only demonstrates the successful deployment of AR with easily accessible tools, but it also opens the door for future advancements in AR technologies.

References

[1] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image studyion;Augmented Reality;Homography;ArUco Markers

[2] Bekhit, Ahmed. (2022). Augmented Reality Using OpenCV. 10.1007/978-1-4842-7462-0_6.

[3] Rogeau, Nicolas & Tiberghien, Victor & Latteur, Pierre & Weinand, Yves. (2020). Robotic Insertion of Timber Joints using Visual Detection of Fiducial Markers. 10.22260/ISARC2020/0068.

[4] Carmigniani, J., Furht, B., Anisetti, M. et al. Augmented reality technologies, systems and applications. Multimed Tools Appl 51, 341–377 (2011). https://doi.org/10.1007/s11042-010-0660-6

[5] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image projection;Augmented Reality;Homography;ArUco Markers

[6] Yap, H.J. and Taha, Z. and Eng, T.H. and Chew, J.Y. (2009) Design and development of a marker-based augmented reality system using OpenCV and OpenGL. In: Asia Pacific Industrial Engineering and Management Systems Conference, 14-16 Dec 2009, Kitakyushu, Japan.

[7] J. Zhang et al., "A perspective transformation method based on computer vision," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 765-768, doi: 10.1109/ICAICA50127.2020.9182641. keywords: Cameras;Face;Visualization;Automobiles;Blindness;Computer vision;Distance measurement;OpenCV;Computer vision;Blind spots;Affine transformation

[8] K. Shah, M. Pandey, S. Patki and R. Shankarmani, "A Virtual Trial Room using Pose Estimation and Homography," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 685-691, doi: 10.1109/ICICCS48265.2020.9120947. keywords: Industries;Instruments;Urban areas;Sociology;Pose estimation;Control systems;Mobile applications;Virtual trial room;OpenCV;object detection;TensorFlow lite;augmented reality;clothes

[9] Kalaitzakis, M., Cain, B., Carroll, S. et al. Fiducial Markers for Pose Estimation. J Intell Robot Syst 101, 71 (2021). https://doi.org/10.1007/s10846-020-01307-9

[10] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image projection;Augmented Reality;Homography;ArUco Markers

ARFusion : Real Time Aruco Marker Augmentation.

Prateek Manoj Mhatre
Roll No: 35
Reg No: 220962158

Mustafa Murtaza Haji
Roll No: 22
Reg No: 220962076

Krish Didwania
Roll No: 39
Reg No: 220962244

***Abstract*—This study investigates the use of augmented reality (AR) in real-world settings by superimposing images over ArUco markers via a live camera stream using OpenCV and ArUco marker recognition. The technique ensures accurate and durable picture overlays, even in difficult situations like changing viewing angles or partial marker occlusions, by combining feature matching with homography-based warping. The AR overlays' alignment and stability are preserved by this dual-technique adaption. The system's usefulness extends across a number of applications, including dynamic object manipulation, AR-enhanced education, and interactive gaming, providing a fresh way to enrich visual experiences. The video overlay on a screen of the desired size is adjusted in this study using 6x6 fiducial markers based on marker spacing. Additionally, 4x4 markers make it easier to manipulate videos in many ways, such as up, down, and and edit functions, while 5x5 markers enhance the precision of object placement within AR environments.**

***Index Terms*—Augmented Reality (AR), OpenCV, ArUco Marker Detection, Feature Matching, Homography, Real-time Image Overlay**

I. INTRODUCTION

In order to improve a user's perception and engagement with their surroundings, computer-generated content is superimposed onto real-world situations in augmented reality (AR), which is a merging of the digital and physical worlds. A real-time direct or indirect view of a physical world improved by the addition of virtual computer-generated information is called augmented reality (AR). This integration is both interactive and registered in 3D, merging real and virtual objects seamlessly. According to the Reality-Virtuality Continuum introduced by Paul Milgram and Fumio Kishino, AR falls on a spectrum between the real environment and the purely virtual environment, positioning closer to the real world than to complete virtuality. This framework helps in understanding the varying degrees of augmented experiences, from slight enhancements of reality to fully immersive virtual environments.

With the rapid advancement of computer vision technologies, AR has found applications across various domains, including gaming, education, healthcare, and industrial design. This study leverages the capabilities of OpenCV and ArUco markers to create a robust AR system that enables real-time image and video overlays onto physical surfaces, thereby enriching the visual experience.

The use of ArUco markers, which are binary square fiducial markers that are simple to detect and individually identifiable in video streams, forms the basis of this investigation. These markers provide reference points for the positioning of digital content, acting as anchors in the real world. The goal of the project is to flawlessly superimpose a pre-made image or video in real-time onto an ArUco marker, with the overlay dynamically adjusting to the orientation and position of the marker. This is accomplished by combining feature matching methods with homography-based modifications, which guarantee that the digital content stays precisely aligned with the marker even in the face of difficult circumstances like changing lighting, angles, and partial occlusions.

Fig. 1. Milgram's Reality-Virtuality Continuum. The continuum spans from the real environment to a fully virtual environment, illustrating where augmented reality and augmented virtuality lie in relation to these extremes.

The incorporation of AR with video playback offers a revolutionary method of media consumption. AR makes it possible to create virtual screens that improve immersion and engagement by allowing videos to be projected onto surfaces in the viewer's surroundings. In addition to making watching more interesting, this technology gives us the freedom to turn any room into a dynamic viewing environment, completely changing the way and location of how we watch videos.

In domains where precise visualization is critical, the strategic placement of 3D objects with AR technology is essential. AR offers a way to put intricate 3D models in a real-world setting, giving businesses like interior design and architecture a realistic sneak peek at completed projects and interior layouts before actual work starts. By eliminating mistakes, cutting expenses, and maximizing both practical and aesthetic choices, this AR application greatly enhances project planning and client pleasure.

All things considered, the developments in AR made possible by the accuracy of ArUco markers not only make complicated AR applications more feasible, but they also broaden the use of AR technology and push the limits of how digital and physical realities may coexist to produce rich, engaging experiences. This study demonstrates the usefulness and potential of AR to revolutionize both professional and daily life.

II. Fiducial Markers in Augmented Reality

Because they offer trustworthy reference points that enable exact camera position estimation and accurate object localization, fiducial markers are essential in augmented reality (AR) and other computer vision applications. Computer vision algorithms identify these markers' unique, instantly identifiable patterns, which enable systems to ascertain the camera's orientation and position in relation to the marker. ArUco markers, specifically, are designed for quick, robust detection even in dynamic, real-time environments, making them ideal for AR applications such as overlaying virtual objects, video playback control, and interactive object placement. By serving as a consistent point of reference, these markers enhance the accuracy and responsiveness of AR experiences, helping to bridge the virtual and real worlds in various applications, from gaming to industrial design and education.

- **ArUco Markers**: Optimized for high-speed detection in real-time AR applications; widely used in robotics for quick camera calibration.
- **AprilTags**: Known for robust detection in varied lighting and motion conditions, often used in industrial AR setups and outdoor robotics.
- **QR Codes**: Primarily designed to store data such as URLs or text; commonly used in applications where data retrieval is prioritized over precise positioning.
- **ChArUco Boards**: A combination of ArUco markers and chessboard corners that enhances accuracy in corner detection, frequently used in camera calibration tasks.
- **RUNE-Tags**: Designed for high accuracy in occlusion-prone environments, ideal for applications requiring precise spatial responses under partial visibility.

We have used ArUco tags in this project to enable accurate camera position estimate, which makes it possible to insert virtual objects and videos in real-world settings with precision.

A. ArUco Markers

The design of ArUco markers, a popular kind of fiducial marker, is straightforward but efficient: a black square border with a special binary matrix in the middle. Computer vision algorithms can identify and decode the encoded data in this matrix, which is represented as a binary pattern. Once detected, the software can identify each unique marker and estimate its position and orientation (pose) relative to the camera, enabling precise spatial tracking.

The robust design of ArUco markers provides strong resilience against partial occlusions, such as when a part of the marker is obstructed by another object. Additionally, their high contrast and square shape allow for rapid and reliable detection, even in changing or dynamic environments where lighting and movement can vary significantly. This makes them especially suitable for applications that require real-time tracking, such as augmented reality (AR), robotics, and drone navigation, where accurate pose estimation and stability are essential for effective performance.

1) Detection and Decoding Process: Detection of ArUco markers is typically performed using the following steps:

1) **Identification of marker outlines using adaptive thresholding**: Adaptive thresholding is applied to distinguish high-contrast regions, allowing for the isolation of potential marker outlines in various lighting conditions. This process identifies the square shapes of ArUco markers by detecting contours and verifying the presence of a solid black border surrounding a binary matrix.
2) **Decoding the binary patterns within the detected markers**: Once markers are detected, the binary matrix within each marker is read and decoded. Each binary pattern is unique, allowing the system to determine the marker's ID and access the specific information it encodes, enabling accurate identification and further calculations, such as pose estimation.

To estimate the position of the detected ArUco marker relative to the camera, a mathematical model is used to transform the 3D world coordinates of the marker to 2D image coordinates as follows:

$$s \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where:

- s: The scaling factor that adjusts the coordinates based on the marker's distance from the camera.
- K: The camera's intrinsic parameters matrix, containing values such as focal lengths and optical center offsets, which reflect the camera's internal properties.
- $[R|t]$: A combination of the rotation matrix R and the translation vector t. The rotation matrix R defines the orientation of the marker relative to the camera, while the translation vector t specifies its position in the camera's coordinate frame.
- (X, Y, Z): The 3D coordinates of a point on the marker in real-world space.
- (x, y): The 2D coordinates of the corresponding point in the camera's image plane, where the marker appears.

This transformation model enables precise pose estimation, allowing augmented reality applications to accurately overlay virtual objects on the detected marker in real-world scenes.

2) Application in Video Playing and Object Placement: ArUco markers can be used to anchor digital content in physical areas during video playback. For example, by aligning a video with an ArUco marker, it can be superimposed on a particular spot or item in the actual world. This approach allows the video content to stay in place relative to the marker, even as the camera moves, providing a seamless augmented reality experience. Such functionality enhances interactivity by enabling users to view media in context with their surroundings, making it particularly useful for interactive installations or museum exhibits where content is tied to physical objects.

(a)ArUco Tag (b)April Tag (c)QR Code (d)ChArUco Tag (e)RUNE Tag

Fig. 2. Kinds of Fiducial Markers

For object placement, ArUco markers facilitate precise and real-time positioning of 3D models in an AR environment. When a marker is detected, the system can accurately overlay virtual objects at its location, aligning with the user's perspective. This feature is invaluable in applications like virtual furniture arrangement for interior design, where users can visualize how furniture will look in a room before purchasing. Additionally, in educational settings, ArUco markers can be used to create interactive learning tools, where students manipulate virtual objects, such as molecular models or historical artifacts, within the classroom. This makes learning more engaging and visually informative, as students can interact directly with augmented content in a hands-on manner.

3) Advancements and Future Prospects: Recent advancements in machine learning and computer vision have led to improvements in the detection algorithms for ArUco markers, enhancing their accuracy and reducing their susceptibility to environmental factors such as lighting conditions and occlusions. Ongoing research continues to expand their applicability in more complex AR systems, promising even more seamless integration of AR technologies into everyday life.

III. OpenCV Functions for Tag Detection and Augmented Reality

A. Overview

For creating applications requiring tag identification and augmented reality, OpenCV offers a strong collection of functions. From simple image processing to intricate transformations and AR overlays, these features make a wide range of activities easier.

B. Key Functions

1) cv::aruco::detectMarkers: The `detectMarkers` function is a part of the ArUco module in OpenCV, designed to identify fiducial markers within an image. It processes an input image to locate markers and returns their IDs and the coordinates of their corners.

- **Parameters:** This function takes the grayscale image, the dictionary specifying the type of markers to look for, and parameters that fine-tune the detection process.
- **Usage:** It is crucial for initializing the tracking of objects in AR applications, allowing the system to understand where virtual objects should appear in the real world.

2) cv::aruco::estimatePoseSingleMarkers: Once markers are detected, `estimatePoseSingleMarkers` calculates the 3D pose of each marker relative to the camera. This function is vital for AR, as it determines how objects should be rotated and scaled in 3D space to align correctly with the physical world.

- **Parameters:** It requires the corners of the marker, the size of the marker, camera calibration parameters (camera matrix and distortion coefficients), and outputs rotation and translation vectors.
- **Usage:** Essential for placing 3D models in an AR scene accurately.

3) cv::perspectiveTransform: This function applies a perspective transformation to a set of points. In the context of AR, `perspectiveTransform` is used to adjust the coordinates of a virtual object so that it fits the perspective of the viewing camera, based on the pose estimation.

- **Parameters:** It takes points and the transformation matrix derived from the real-world coordinates of the markers.
- **Usage:** It aligns the virtual and real-world coordinates, enabling realistic integration of virtual content into the live video feed.

4) cv::warpPerspective: `warpPerspective` warps an image using a given transformation matrix. This function is particularly useful in AR for rendering images or videos onto detected markers.

- **Parameters:** Requires the source image, output matrix size, and the transformation matrix.
- **Usage:** Used to overlay and adjust images or videos onto surfaces defined by markers, crucial for immersive AR experiences.

IV. Methodology

The first step in the study is to use OpenCV's `cv2.aruco` module to identify ArUco markers in the video feed. Accurate digital content overlay is made possible by ArUco markers, which offer a strong basis for identifying reference locations in the video. The system determines the homography matrix, a transformation matrix that is used to warp the intended picture or video material onto the marker's location within the video feed, after a marker has been recognized. Even when the camera perspective shifts, this transformation makes

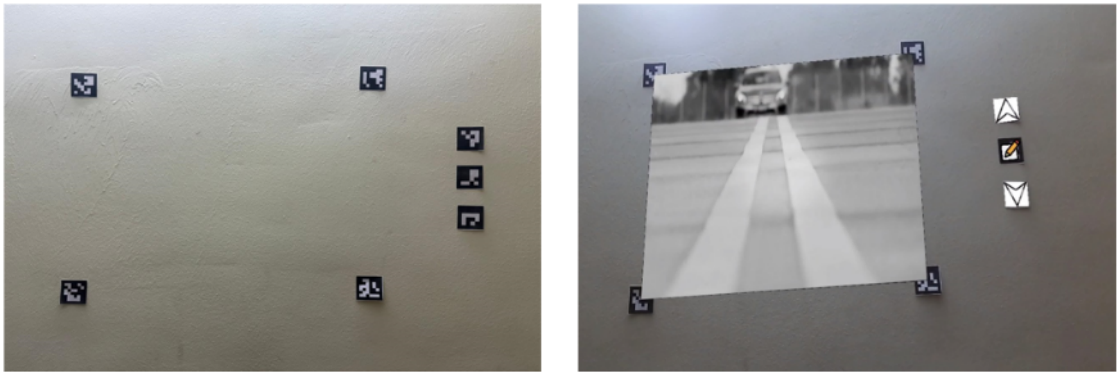

Fig. 3. The ArUco markers placed on a wall, configured for augmented reality tracking. (Right) A video dynamically overlaid onto the wall using the positions of the ArUco markers, demonstrating real-time video mapping in an augmented reality application.

sure that the overlaid content stays appropriately scaled and positioned in reference to the marker. To enable a wide range of applications, the project incorporates cutting-edge video and object interaction capabilities employing ArUco markers. Real-time effect application, interactive video navigation, and accurate virtual object presentation in an augmented reality environment are all made possible by these markers. The system's ability to track and interact with markers enhances user engagement by providing a seamless AR experience where digital content is contextually anchored to the physical environment.

- **Marker Detection**: The process begins with the detection of ArUco markers in the video feed. ArUco markers are detected using OpenCV's cv2.aruco module, which provides functions to detect these fiducial markers and estimate their pose. Upon detection, the corners of the marker are identified, which serve as the reference points for subsequent transformations. The accuracy of this detection phase is crucial, as any errors in identifying the marker's position and orientation could lead to misalignment of the overlay.
- **Homography Calculation**: Once the marker is detected, the system calculates the homography matrix, a transformation matrix that maps the four corners of the marker in the physical world to the corresponding corners in the overlay image. The homography matrix is essential for warping the overlay image to fit precisely onto the marker, even as the marker moves or rotates within the video feed. This transformation ensures that the overlay remains anchored to the marker's position, providing a seamless AR experience.
- **Feature Matching**: While homography provides a basic alignment of the overlay, additional refinement is often necessary to ensure that the overlay remains accurate under varying conditions, such as when the marker is viewed from an angle or is partially obscured. This is where feature matching comes into play. The system detects keypoints and descriptors in both the ArUco marker and the overlay image using algorithms like SIFT (Scale-Invariant Feature Transform) or ORB (Oriented FAST and Rotated BRIEF). By matching these features, the system can adjust the overlay dynamically, correcting any distortions that may occur due to changes in perspective.
- **Image Overlay** : The final stage involves warping the overlay image using the homography matrix and applying the feature-based corrections derived from the matching process. This ensures that the overlay remains stable and accurately aligned with the marker in real-time. The overlaid image or video is then displayed on the marker within the video feed, creating the illusion of digital content seamlessly integrated into the physical environment.

A. *Video Manipulation with ArUco Markers*

1) Multiple Marker Tracking for Video Playing: In this project, we utilize 6x6 ArUco markers to define the vertices of a virtual screen, onto which videos are dynamically displayed. These markers are tracked in real-time, allowing users to adjust the position and scale of the projected video interactively. As the markers move, the system continuously calculates the homography matrix to adjust the video frame to fit within the area defined by the markers. This transformation ensures that the video display is correctly aligned and scaled according to the markers' position, even as the user changes the markers' orientation. The real-time tracking of these markers is essential for providing a smooth and immersive viewing experience, allowing users to seamlessly interact with the video content. The video display instantly adapts to any changes in the marker's position or orientation, offering a versatile and dynamic method of video projection that enhances interactivity and engagement.

2) Control Markers for Video Navigation and Effects: In addition to the 6x6 markers used for video playing, we designate 4x4 ArUco markers to control specific video functions and visual effects. These markers are tracked in real-time to trigger the following actions:

- One marker allows users to skip to the next video in the playlist.
- Another marker takes the user back to the previous video.
- A third marker enables the cycling of various video effects, such as grayscale, canny edge detection, Gaussian blur, and negative image filters.

To interact with these controls, users simply move the designated markers into view of the camera. The system detects these markers and triggers the corresponding function, enabling seamless video navigation or visual effect application in real time. This real-time marker tracking ensures that the system responds immediately to user input, making the interaction intuitive and efficient. By leveraging the precise detection of control markers, users can interact with video content dynamically, adjusting playback and applying effects without the need for traditional interfaces like remote controls or buttons.

B. Augmented Reality Object Display

Additionally, our solution incorporates the use of 5x5 ArUco markers to enable the augmented reality display of 3D objects from STL or OBJ files. Multiple objects can be placed into the physical environment as seen through the camera thanks to ArUco markers, which are linked to specific objects. This setup provides users with the ability to place objects in a real-world context, offering an immersive experience where digital content appears seamlessly integrated with the surroundings.

The system's capacity to scale the items on display in real-time in response to human contact with the ArUco markers is one of its primary characteristics. Because the object's size is dynamically changeable, users can see the 3D model at various scales to fit their own requirements or tastes. Additionally, the quantity of ArUco markers found in the surroundings is strongly correlated with the quantity of things shown. A matching 3D object can be rendered for every marker found, enabling flexible and adaptable interactions in multi-object configurations.

At this early level of development, we have used OpenCV-supported basic 3D models from STL or OBJ files to generate an artificial spider. Although this method illustrates the fundamental functionality, the project's future development intends to include realistic, high-quality 3D items, such vases or other complex models, for improved realism and display. These improvements will increase the system's potential for applications in fields requiring high-fidelity 3D representation, such as product visualization, virtual prototyping, and interior design.

This feature is especially helpful in applications like educational tools, where students can manipulate and interact with virtual objects to better understand complex concepts, or in design scenarios, like architecture or interior design, where users can see how various models or objects fit within a physical space. The ability to scale objects and add more based on the number of markers provides a dynamic and customizable augmented reality experience.

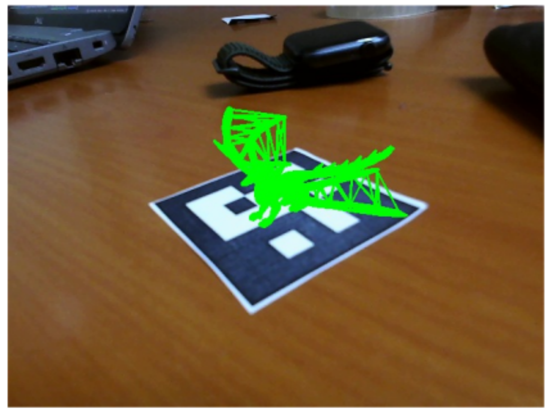

Fig. 4. A 3D display of a dragon from a STL file onto a 5x5 aruco tag

V. Salient Features

A. Real-Time Interaction and Feedback

- **Real-Time Processing:** The system operates in real-time, providing immediate visual feedback to user interactions. This feature is critical for applications where timing and responsiveness are crucial, such as in interactive gaming and live training simulations.
- **Immediate Response to Environmental Changes:** Our application dynamically adjusts to changes in the environment, such as variations in lighting and movement, ensuring consistent performance without lag.

B. Stability and Accuracy

- **Robust Overlay:** The combination of homography and feature matching ensures that the overlay remains stable even with changes in perspective, lighting, or partial occlusion. This robustness is essential for maintaining an immersive user experience in AR applications.
- **Enhanced Tracking Precision:** Employing advanced algorithms for marker detection, the system minimizes errors in tracking and overlay, which is particularly beneficial for precise tasks in educational and industrial AR applications.

C. Versatility and Scalability

- **Versatility:** The application can be adapted for various uses, from interactive gaming to AR-assisted learning and dynamic advertising. This adaptability allows it to be integrated into multiple industries including education, marketing, and entertainment.

- **Customizable Overlays:** Users can customize overlays and interactions, which enhances user engagement and allows for personalized experiences in applications like virtual showrooms and interactive learning environments.
- **Scalability:** The system is designed to scale seamlessly, supporting multiple markers and overlays simultaneously. This scalability facilitates the expansion of the application to larger environments and more complex scenarios without a loss in performance.

D. Innovative Applications

- **Interactive Gaming:** By integrating real-world objects into the gaming environment using AR markers, the system provides a unique gaming experience that blends physical and digital gameplay.
- **AR-assisted Learning:** The system enhances educational content by overlaying digital information onto physical objects and environments, making learning more interactive and engaging.
- **Dynamic Advertising:** Advertisers can create immersive and interactive campaigns that engage consumers directly by overlaying digital content onto physical spaces, such as shopping malls and public squares.

These features collectively enhance the functionality and applicability of the AR system, making it a versatile tool for both entertainment and professional applications.

VI. Conclusion

Using OpenCV and ArUco markers, this study effectively combines augmented reality (AR) with real-world settings to produce accurate and reliable real-time overlays of photos and videos. The incorporation of feature matching and homography-based transformations ensures that digital material faithfully adheres to the physical marks, managing challenges such as changing viewpoints and obstacles.

ArUco markers serve as a robust basis for these overlays, ensuring dynamic adaptability to variations in marker placement and orientation. The versatility of this approach is showcased in its broad range of applications, including interactive gaming, educational enhancements, and engaging advertising, all benefiting from the interactive, real-time capabilities of the system.

Additionally, the proposed framework can handle complex AR settings with many overlays and markers because it is versatile and scalable. This project highlights the revolutionary potential of AR in diversifying visual interactions across several sectors and not only demonstrates the successful deployment of AR with easily accessible tools, but it also opens the door for future advancements in AR technologies.

References

[1] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image studyion;Augmented Reality;Homography;ArUco Markers

[2] Bekhit, Ahmed. (2022). Augmented Reality Using OpenCV. 10.1007/978-1-4842-7462-0_6.

[3] Rogeau, Nicolas & Tiberghien, Victor & Latteur, Pierre & Weinand, Yves. (2020). Robotic Insertion of Timber Joints using Visual Detection of Fiducial Markers. 10.22260/ISARC2020/0068.

[4] Carmigniani, J., Furht, B., Anisetti, M. et al. Augmented reality technologies, systems and applications. Multimed Tools Appl 51, 341–377 (2011). https://doi.org/10.1007/s11042-010-0660-6

[5] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image projection;Augmented Reality;Homography;ArUco Markers

[6] Yap, H.J. and Taha, Z. and Eng, T.H. and Chew, J.Y. (2009) Design and development of a marker-based augmented reality system using OpenCV and OpenGL. In: Asia Pacific Industrial Engineering and Management Systems Conference, 14-16 Dec 2009, Kitakyushu, Japan.

[7] J. Zhang et al., "A perspective transformation method based on computer vision," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 765-768, doi: 10.1109/ICAICA50127.2020.9182641. keywords: Cameras;Face;Visualization;Automobiles;Blindness;Computer vision;Distance measurement;OpenCV;Computer vision;Blind spots;Affine transformation

[8] K. Shah, M. Pandey, S. Patki and R. Shankarmani, "A Virtual Trial Room using Pose Estimation and Homography," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 685-691, doi: 10.1109/ICICCS48265.2020.9120947. keywords: Industries;Instruments;Urban areas;Sociology;Pose estimation;Control systems;Mobile applications;Virtual trial room;OpenCV;object detection;TensorFlow lite;augmented reality;clothes

[9] Kalaitzakis, M., Cain, B., Carroll, S. et al. Fiducial Markers for Pose Estimation. J Intell Robot Syst 101, 71 (2021). https://doi.org/10.1007/s10846-020-01307-9

[10] D. S. Asanvitha Gundala, S. S. Alamuri, A. Firdaus and G. K. Kumar, "Implementing Augmented Reality Using OpenCV," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-4, doi: 10.1109/DELCON54057.2022.9753233. keywords: IEEE Sections;Conferences;Data visualization;Detection algorithms;Augmented reality;Image projection;Augmented Reality;Homography;ArUco Markers

CV_final rpt

ORIGINALITY REPORT

4%	2%	3%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	patents.google.com Internet Source	1%
2	Antonio Sartal, Diego Carou, J. Paulo Davim. "Enabling Technologies for the Successful Deployment of Industry 4.0", CRC Press, 2020 Publication	<1%
3	Walter Vanzella, Natalia Grion, Daniele Bertolini, Andrea Perissinotto, Marco Gigante, Davide Zoccolan. "A passive, camera-based head-tracking system for real-time, three-dimensional estimation of head position and orientation in rodents", Journal of Neurophysiology, 2019 Publication	<1%
4	hdl.handle.net Internet Source	<1%
5	link.springer.com Internet Source	<1%
6	fastercapital.com Internet Source	<1%

7 Y. Sato, M. Nakamoto, Y. Tamaki, T. Sasama, I. Sakita, Y. Nakajima, M. Monden, S. Tamura. "Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization", IEEE Transactions on Medical Imaging, 1998 <1%

Publication

---

8 A S Priambodo, F Arifin, A Nasuha, Muslikhin, A Winursito. "A Vision and GPS Based System for Autonomous Precision Vertical Landing of UAV Quadcopter", Journal of Physics: Conference Series, 2022 <1%

Publication

---

9 Osslan Osiris Vergara Villegas, Vianey Guadalupe Cruz Sánchez. "Augmented Reality - Fundamentals and Applications", CRC Press, 2024 <1%

Publication

---

10 rfos.fon.bg.ac.rs <1%

Internet Source

---

11 Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, Misa Ivkovic. "Augmented reality technologies, systems and applications", Multimedia Tools and Applications, 2010 <1%

Publication

12	Yuanpeng Liu, Jun Zhou, Yida Li, Yuqi Zhang, Yongqiang He, Jun Wang. "A high-accuracy pose measurement system for robotic automated assembly in large-scale space", Measurement, 2021 Publication	<1%

Exclude quotes	On	Exclude matches	< 3 words
Exclude bibliography	On		