MARKER-LESS AUGMENTED REALITY FOR SMART TEACHING

Koniki Venkata Balaji

MARKER-LESS AUGMENTED REALITY FOR SMART TEACHING

by
Koniki Venkata Balaji
Under the supervision
of
Dr. D. P. Dogra

Thesis Report submitted to the
Indian Institute of Technology Bhubaneswar
for partial fulfilment of the degree
of
Bachelors and Master of Technology (Dual Degree)

SCHOOL OF ELECTRICAL SCIENCES
INDIAN INSTITUTE OF TECHNOLOGY BHUBANESWAR
NOVEMBER 2021

ABSTRACT

In today's modern world, physical classrooms are getting transformed into virtual classrooms at a rapid pace. In such an environment, it is vital for both instructors and students to use the latest hand tracking and human-computer interaction technologies to explain and understand concepts. As there is a massive focus on developing the education sector, the use of modern technologies in a classroom setting is being given a lot of thought. Emerging technologies such as Augmented Reality(AR) and Virtual Reality(VR) have a great potential to enhance the teaching and learning process. Augmented reality is a technology that introduces artificial objects into the real-world scene for giving an interactive experience.Education can be made more interactive by using AR technology as it replaces memory-based learning with immersive experiences. This approach improves students' understanding of concepts and enhances their interest in learning. However, there is no reliable framework that augments a virtual object in the physical environment and gives an opportunity to interact with them with our bare hands. In this work, we propose a framework that supports augmented reality with marker-less human hand interaction features. It enables us to use specific hand gestures to perform operations such as zoom-in, zoom-out, rotation, etc., on the virtual objects.

Keywords: Hand Tracking, Augmented Reality, Virtual Reality, Human-Computer Interaction, Hand Gesture

Contents

List of Figures

Chapter 1

Introduction

Augmented reality (AR) and virtual reality (VR) are two technologies that have the word reality as a common term in their denominations, however, they differ in terms of the use of the real world or not. In fact, AR keeps the real physical world of a user and adds virtual content to his view in real-time [1]. While VR immerses a user within a synthetic world that imitates the real one, the user's sight, hearing and motion are stimulated to give the user the feeling of immersion in the virtual world [2]. Both of these technologies are revolutionary tools for the educational sector, they have many applications that promote learning and teaching processes. However, their use in education and especially in a classroom setting will impose additional requirements. classrooms and schools will be a good ground for AR/VR technologies. In this Thesis Work, we try to propose a learning framework based on AR technology that can be adopted in a smart classroom. The chapter is structured as follows: The first section lists the overview of the area. While the second section provides the main problem statement being solved in this thesis, the third section lists down the motivations and objectives of the study. Finally, the fourth section summarizes the proposed contributions.

1.1 Overview of the Area

Augmented reality (AR) allows to seamlessly insert virtual objects in an image sequence. A widely acknowledged definition of augmented reality is given by Azuma in the first survey dedicated to the subject [3]. *An AR system should combine real and virtual objects, be interactive in real-time, register real and virtual objects.* AR and VR are totally different in terms of their objectives. The former enriches the real

world of a user with a layer of digital information, the latter isolates a user from his real-world and immerses him in a virtual world. Moreover, they are distinct in the way they deliver the experience to the user.

AR can be used everywhere since users are not isolated from their real world. They can use mobile devices (tablets, smartphones) or smart glasses, which are comfortable tools. Users can carry out other activities while using AR glasses (walking or repairing equipment using smart glasses). AR is tied to the real world, and its components (people, real objects). It relies on its real elements to provide additional information to enrich the view of a user [4]. In AR, users still see the real world, including the surrounding objects and parts of their bodies (e.g. hands, feet) [5]. They do not feel any cybersickness while using AR contrary to VR [5]. By using AR, users can project information or instructions directly into real objects or real equipment [5]. Additionally, students can use AR in their school trips, in educational settings like museums and libraries, and even at home.

AR and VR have numerous differences; however, they have some common features. Both of them are technologies that allow users to have rich experiences using visual or other sensory simulations [6]. They require users' use of their senses and minds. Additionally, they afford the possibility of interaction with the content, which justifies users' engagement while using AR and VR technologies. No one of AR and VR technologies can replace the other. In fact, they are complementary technologies. Each one has its own role and responds to a specific need. On one side, AR augments the real world. It needs a real element, object, or scene as an input. It demands a real-time input to demonstrate an augmentation over it. Thus, the importance of AR technology lies in subjects and activities that aim at exploring characteristics, properties, or information about real facts or objects. For instance, the real human body is a point of interest for medicine students, fossils arouse curiosity for geology students, and historical artifacts draw attention for archeology students. On the other side, VR creates a new virtual world. Therefore, it is a good tool to experience phenomena that are impossible or difficult to experience in the real world. Students of mechanical engineering can design cars and experience them in VR. Learners can experience past historical events, go abroad, travel to space, and inspect the interior of the earth.

1.2 Problem Statement

The problem that is being solved here is to design a novel framework in a classroom setup by making use of Augmented Reality and hand tracking technologies. The proposed framework should detect a bare human hand in the incoming video feed and augment 3D virtual objects on the hand after camera pose estimation in every frame. It should also support hand gesture recognition to perform basic interactions with the augmented objects such as zoom-in, zoom-out, and rotation. The problem also demands the solution be real-time on commodity devices for its wide range of implementation.
The problem can be solved in the following four sub-tasks:

- Hand Features Extraction
- Camera Pose Estimation
- Virtual object Registration
- Hand Gesture Interaction

1.3 Motivation and Objectives of the Study

AR in education can provide multiple benefits not just in terms of a shift from traditional to digital but it is also said to have enhanced and effectively impacted understanding and interaction among the students and also have resulted in good results and in-depth understanding of the concept. Therefore it can be said that the combination of interactive education and engaging content can cause the students to remember what they have learned and gain fast skills and information in no time and without much chaos. AR can be applied in almost all the subjects that are taught to students of all age groups be it in schools or in colleges. They can range from reading to maths to languages to sciences to machines etc. All the above said features and the potential of AR technology to transform education are the motivations for this work.
The objectives of this work are:

1. To improve traditional teaching methods with the usage of emerging technologies such as AR

2. To propose a methodology that makes learning more immersive and interesting

3. To develop a learning framework that is low-cost and use fewer resources compared to commercial devices

1.4 Proposed Contributions

In an attempt to solve the problem, we came up with a solution that uses the MediaPipe Hand Tracking Model, OpenCV, and OpenGL. The framework doesn't require any special hardware or extra hardware resources to function. It is designed to perform satisfactorily on a commodity PC consisting of a decent webcam. Following are the main contributions of this work.

1. Implementation of camera pose estimation by making use of camera transformation from world coordinates to image coordinates. The intrinsic matrix of the camera is estimated with the help of a checkerboard.

2. Registration of 3D objects on bare hand detected by Mediapipe Hands model using OpenGL library.

3. Interaction with virtual objects such as zoom-in, zoom-out, and rotation using different hand gestures.

Chapter 2

Related work

AR has been employed in two main ways in the literature: marker-based augmented reality and marker-less augmented reality.When a device, such as a camera or a cell phone, detects a 2D image or a QR code, marker-based AR produces the intended result.Marker-less AR, on the other hand, is dependent on the device's specifications, such as GPS position.It's also known as location-based AR.Anatomy 4D is a type of MAR application in which a human body image serves as a marker.When the webcam detects the marker, it populates it with organs, bones, and nerves, allowing for a realistic depiction of human anatomy.While star walk in the form of marker-less AR application in which the user only aims the camera towards the sky to have a detailed view of stars, planets, and orbiting objects within a certain range [7]. Both of these forms are showing their dominance in many domains such as e-shopping, marketing, and education but the focus of our study is marker-less AR application.

In this digital era of dynamic technology companies need an innovative way of marketing in order to gain customer attention. Keeping in mind this critical demand of the market Bule et al. [8] used AR technique in order to grab customers' attention. They designed an AR system that detects the face of the customer through a webcam using a face detection algorithm. The system calculates the detected face position and places the comical slogan above the customer's head. The contents of the slogan are enriched with text and images. If no face is detected by the webcam it generates advertisements in the form of moving images. The application grabbed the attention of almost every visitor that passed by especially in groups.

Digital media these days is strongly influencing the shopping behavior of people hence for this purpose Khushal et al.[9] proposed a MAR application for buying furniture. Their proposed study mainly focuses on trying different furniture objects

for home without visiting the shops. The user places the marker at the position where he wants to place the object in the room. A webcam is used to capture the live feed of the room and detect the marker. The user selects the object from the database in order to try the best-fitted item in his room. In this way, the user can view the object in his room from different angles. Their proposed study addressed the issues of common people who generally find it time-consuming to visit furniture stores.

Le et al. [10] used MAR in combination with hand gesture recognition to address some of the basic critical issues in geometry. The aim of their study was to manipulate 3D shapes and figures of geometry with hand gestures. Webcam detects the visible marker and with the usage of toolkit virtual content is formed. Hand gestures then manipulate the position and motion of the virtual content to produce the desired results. Their study proved to be beneficial for the students having an interest in geometry and helped them to acquire good grades.

Similarly, Antonia et al. [11] in his studies showed how augmented reality technology is used to improve the learning activities of preschoolers. The main objective of their application was to show vertebrate animal classification which included the details of animal skins, their reproduction kinds, and temperature. They presented an animal park over an AR marker having a group of animals showing their gestures from different angles. They applied this approach to two groups of preschoolers that is control and the experimental group. The results of the experimental group showed that the MAR technique is effective in the learning activities of students at their initial stages.

The relationship between molecules in chemistry is an essential part to be understood by students in order to pass the course. For this purpose, Maier et al. [12] presented the concept of "Augmented Chemical Reactions". This concept implies the usage of the MAR technique which uses a physical cube surfaced with black and white patterns as a marker. The virtual contents were drawn at the top of the cube which forms an illusion showing a chemical reaction. Users can choose the molecule from the associated protein database. This study had a very positive effect on the students of chemistry.

Yilmaz et al. [13] presented the concept of "Educational Magic Toys (EMT)" which was developed using the MAR technique. The aim of the study was to boost children's imagination, skills, and activities. They used flashcards, match cards, and puzzles as markers to show virtual content. With the help of this MAR technique,

they taught the classification of fruits, animals, vegetables, vehicles, and colors to children having the age group 5-6. They conducted this experiment on a group of 33 children and 30 teachers. The experimental results showed that children showed great interest in the flashcards as 3D virtual objects appear on the card leaving a fascinating impact on them.

Augmented reality in education has its use cases as well, primarily for supporting and enhancing a variety of pedagogical approaches [14] and providing alternative learning environments [15]. Gaming is also a popular domain for AR technology as shown by the huge success of Pokemon Go [16].

Hand tracking is a significant component to creating a natural user interface and communication in AR/VR products and has been an active research area in the industry [17] [18] . Vision-based hand pose estimation has been studied for many years. Most of the previous works require specialized hardware like depth sensors[19][20][21] which are not lightweight to run real-time on commodity devices and thus are limited to platforms with powerful processors. In this report, we propose a teaching model based on a Real-time Hand Tracking solution designed by MediaPipe, Google [22]. We used the aforementioned model from MediaPipe to design a teaching framework that enables a user to augment virtual objects on hand and interact with them using hand gestures.

Chapter 3

Proposed Work

The framework design is explained in three sections. Section 3.1 describes how the MediaPipe solution extracts 21 hand key point coordinates in the bare human palm from an input image frame. Section 3.2 talks about the theory behind camera pose estimation and how it is done. Finally, Section 3.3 shows explains virtual objects are registered in real-world scenes, and gestures are used to interact with them. The overall framework is as shown below in Figure 3.1.

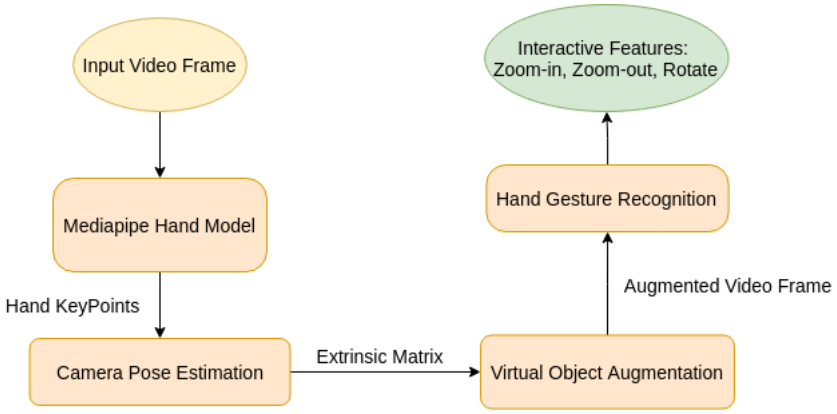


Figure 3.1: Framework Overview

3.1 Hand Landmark Extraction

The Hand Tracking pipeline consists of two models:

3.1.1 BlazePalm Detector

A palm detector was trained to estimate bounding boxes of rigid objects like palms and fists which is significantly simpler than detecting the entire hand in different orientations. Palms were modeled using square bounding boxes ignoring other aspect ratios. An encoder-decoder feature was used for a larger scene-context awareness even for small objects.

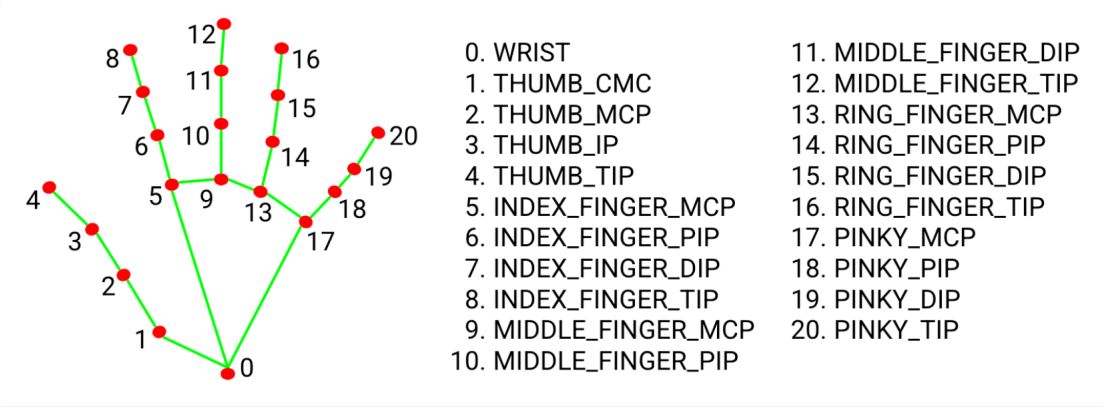


Figure 3.2: 21 Landmark Points detected by MediaPipe Hand Tracking Model

3.1.2 Hand Landmark Model

After running Palm detection over the whole image, the landmark model performs precise localization of 21 2.5D coordinates inside the detected regions via regression as shown in Figure 3.2. The model consistently learns the hand pose and is robust even to partially visible hands and self-occlusions. The model has the following outputs:

- 21 hand landmarks with x, y coordinates, and z relative depth
- A hand flag telling the probability of presence in the input image
- Binary classification of right or left-handedness

3.2 Camera Pose Estimation

This task is the most crucial for the performance of the entire application. This section deals with the problem of determining the position and orientation of the camera relative to the object (or vice-versa). We use the correspondences between 2D image pixels (and thus camera rays) and 3D object points (from the world) to compute the pose. 3D world coordinates are transformed into 2D image coordinates using the camera perspective transformation formula.

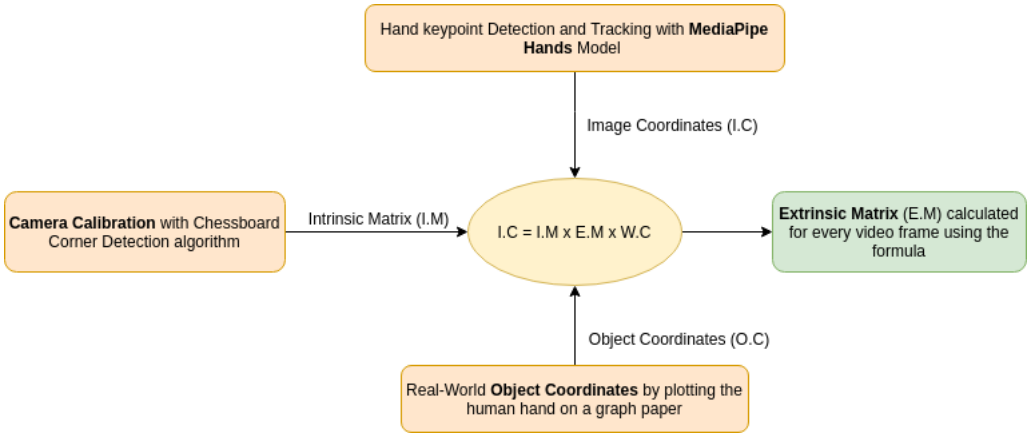


Figure 3.3: Camera Pose Estimation

3.2.1 Camera Calibration

Camera calibration involves estimating the parameters of a camera. This means gathering all the information (parameters or coefficients) about the camera required to determine an accurate relationship between a 3D point in the real world and its corresponding 2D projection (pixel) in the image captured by that particular camera. This involves obtaining two kinds of parameters:

1. **Intrinsic parameters:** Inherent features of the camera/lens system. E.g. focal length, optical center, and radial distortion coefficients of the lens.

2. **Extrinsic parameters:** This refers to the orientation (rotation and translation) of the camera with respect to some world coordinate system.

The intrinsic matrix of the camera which represents its inherent features is approximated using the camera calibrate function available in the OpenCV library.

Initially, nearly 10 checkerboard images are taken and corners are detected using the *findchessboardcorners* function in OpenCV. Then, the *calibratecamera* function takes image and object points of the checkerboard corners as input. The output consists of the Intrinsic matrix and distortion coefficient matrix which says how much distorted are the images.

3.2.2 Extrinsic Matrix Calculation

The main objective of this task is to calculate the extrinsic matrix of the camera for every video frame. It gives the estimated orientation of the camera w.r.t the virtual object. The equations that relate 3D point (X_w, Y_w, Z_w) in world coordinates to its projection (u, v) in the image coordinates are shown in Equations 3.1, Where $\mathbf{P}$ is a 3x4 Projection matrix consisting of two parts - intrinsic matrix($\mathbf{K}$) and extrinsic matrix($[\mathbf{R} \mid \mathbf{t}]$) that is a combination of 3x3 rotation matrix $\mathbf{R}$ and 3x1 translation $\mathbf{t}$ vector as given in Equation 3.2.

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \mathbf{P} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{3.1}$$

$$u = \frac{u'}{w'} \qquad v = \frac{v'}{w'}$$

$$\mathbf{P} = \overbrace{\mathbf{K}}^{\text{Intrinsic Matrix}} \times \overbrace{[\mathbf{R} \mid \mathbf{t}]}^{\text{Extrinsic Matrix}} \tag{3.2}$$

The intrinsic matrix $\mathbf{K}$ shown in Equation 3.3 is upper triangular consisting of x and y focal lengths (f_x, f_y), x and y coordinates of optical center in the image plane (c_x, c_y), and skew γ between the axis which is usually 0.

$$\mathbf{K} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

The camera's extrinsic matrix in Equation 3.4 describes the camera's location in the world, and what direction it's pointing. It has two components: a rotation matrix, $\mathbf{R}$, and a translation vector $\mathbf{t}$.

$$[R \mid t] = \left[\begin{array}{ccc|c} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \end{array}\right] \tag{3.4}$$

Six static hand key points among the detected are used as hand features to estimate the camera pose and virtual objects are augmented on the human hand using the perspective transformation. Their image-coordinates are provided by hand detection model output and object-coordinates are obtained by plotting the hand on a graph paper. Intrinsic parameters are estimated by camera calibrate function and extrinsic parameters are obtained by finding the unknown in the projection formula. The entire flow of camera pose estimation is described in the flowchart Figure 3.3.

3.3 Virtual Object Interaction

After obtaining all camera parameters, virtual objects are augmented on the required location using the same projection formula that was discussed earlier. As the camera transformation matrix projects the 3D world coordinates on to Image plane, the virtual object augments into the real scene environment. Initially, to see the stability of augmentation, simple figures such as 3D axis and cube were registered using only the OpenCV functions. After getting stable outputs, we started implementing the augmentation in the OpenGL environment which enables us to visualize complex 3D objects. The **OpenGL API** is implemented using the **Glumpy** package which offers wrapper functions that use raw OpenGL under the hood. This section explains the process in two parts, first virtual object augmentation and then interaction with hand gestures.

3.3.1 Object Augmentation

Six hand keypoints having indices as 0, 1, 5, 9, 13, and 17 are selected as image points to generate a coordinate plane on the palm. These points are preferred due to their relative static nature at all times irrespective of hand motion. In an OpenGL environment, rendering an object is always relative to the camera, and as such, the scene's vertices must also be defined relative to the camera's view. Model View Projection is a common series of matrix transformations that are applied to a vertex defined in model space, transforming it into clip space, i.e, Image plane, which can

then be rendered. The three matrices perform transformations from one space to another as discussed below.

1) Model Matrix:

A model matrix $\mathbf{M}$ is composed of an object's translation, rotation and scale transform $\mathbf{T}$, $\mathbf{R}$, and $\mathbf{S}$ respectively. Multiplying a vertex position v by model matrix transforms the vector into world space as shown in Equation 3.5.

$$\begin{aligned} \mathbf{M} &= \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{S} \\ v_{world} &= \mathbf{M} \cdot v_{model} \end{aligned} \tag{3.5}$$

2) View Matrix:

As all renderings are from some camera's perspective, all vertices must be defined relative to the camera. *Camera space* is the coordinate system defined as the camera at (0, 0, 0), facing down its -Z axis. The camera also has a model matrix defining its position in world space. The inverse of the camera's model matrix is the view matrix, and it transforms vertices from *world space* to *camera space* or *view space* as shown in Equation 3.6.

$$v_{camera} = \mathbf{V} \cdot \mathbf{M} \cdot v_{model} \tag{3.6}$$

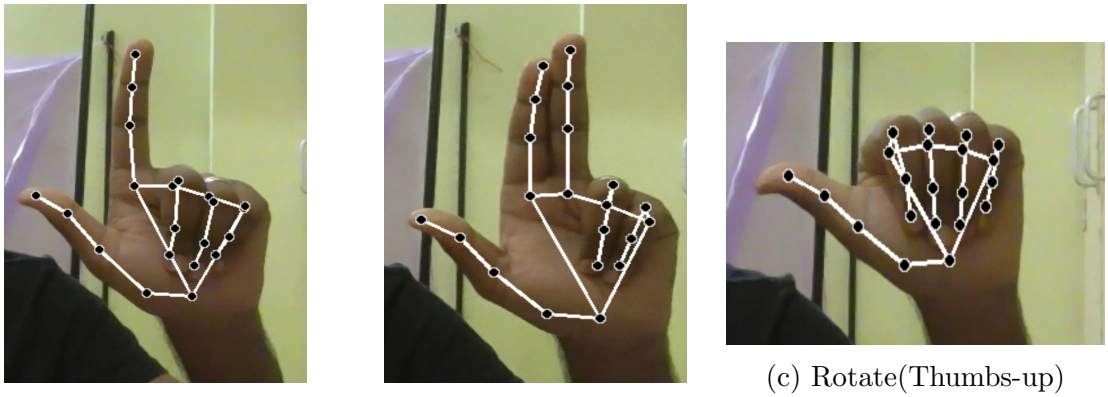
3) Projection Matrix: The projection matrix encodes how much of the scene is captured in a render by defining the extent of the camera's view. The two most common types of projection are *perspective* and *orthographic*. After applying a projection matrix, the scene's vertices are now in *clip space* as given in Equation 3.7.

$$v_{clip} = \mathbf{P} \cdot \mathbf{V} \cdot \mathbf{M} \cdot v_{model} \tag{3.7}$$

The model, view, and projection matrices transform vertices that start in *model space*, and then *world space*, *camera space*, and then *clip space*. The vertices are then transformed into *normalized device coordinates* via implicit perspective division. Finally, during rasterization, a viewport transform is applied to interpolated vertex positions, resulting in a window space position: an X and Y position of a texel in two dimensions, translating some point in 3D space relative to some viewer, into a specific pixel on a screen.

3.3.2 Gesture Interaction

Once the object is augmented in the scene, interactions with the object such as zooming, rotating are performed with the help of hand gesture detection. Three easy gestures are selected for three different tasks which can be detected by comparing relative positions of different hand keypoint coordinates detected by the Mediapipe model. To improve the efficiency of zooming, a queue of fixed size is maintained to store distances between the tips of the index and thumb fingers. Zoom-in and Zoom-out scaling are performed only when all the values in the queue are ascending and descending in order respectively. The three gestures are shown in Figure 3.4.



(a) Zoom-in(pinch-out) (b) Zoom-out(pinch-in) (c) Rotate(Thumbs-up)

Figure 3.4: Hand Gestures

The three gestures are performed as follows:

1. **Zoom-in:** The virtual object expands in size by moving the thumb and index finger away in the zoom-in gesture, i.e, making pinching out movement.

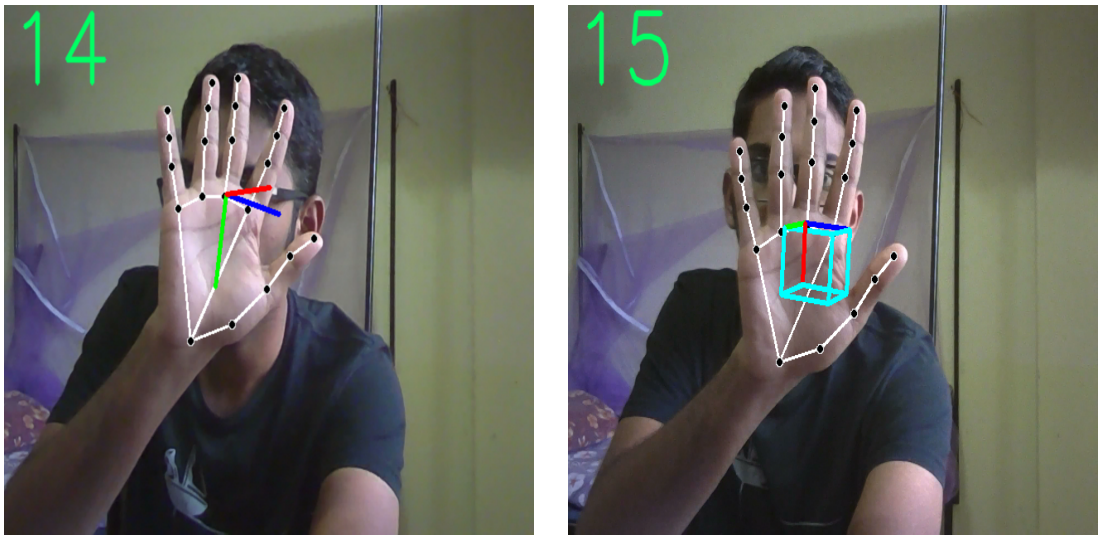
2. **Zoom-out:** the object compresses into smaller sizes by moving the thumb and the pair of index, middle fingers towards each other in the zoom-out gesture, i.e, making pinching in movement.

3. **Rotate:** Rotation of the object is obtained by folding all fingers except the thumb finger pointing outwards horizontally.

Chapter 4

Results

Initially, augmentation is carried out in an exclusively OpenCV environment, and the results are stable and satisfactory. The OpenCV environment makes use of projected 2D coordinates obtained from homogenous coordinates after camera transformation. These 2D coordinates are used to draw lines using OpenCV functions. 3D axis and cube are augmented on hand as shown in Fig 4.1.



(a) 3D Axis (b) 3D Cube

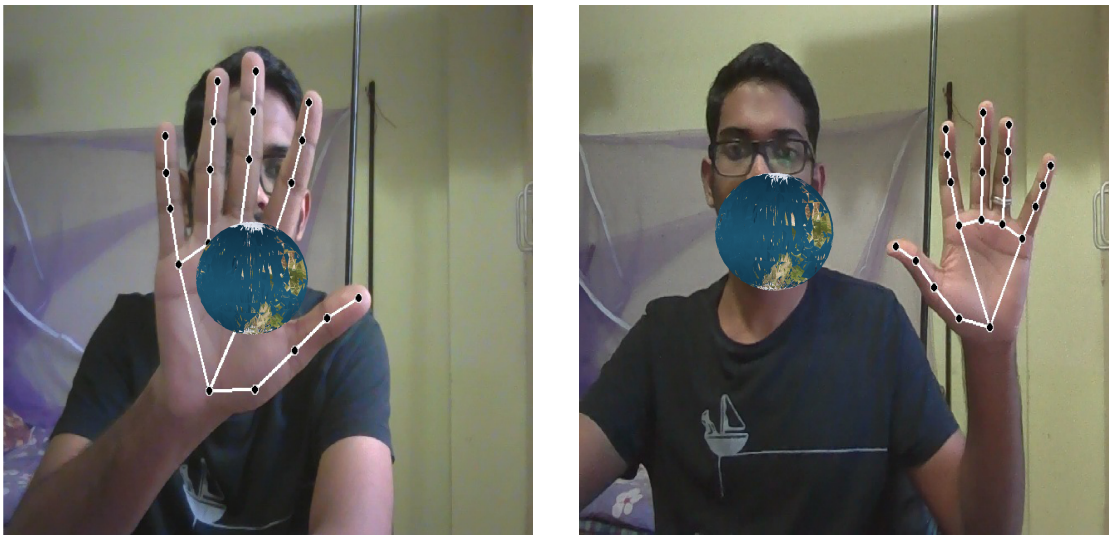
Figure 4.1: Augmentation using OpenCV

(a) Brain (b) Teapot (c) Cow

Figure 4.2: Augmentation of 3D Objects using OpenGL

3D objects of the human brain, teapot, and cow are augmented on the human hand as shown in Figure 4.2.The augmentation is stable and no lagging of frames has been observed in the live feed.

A more advanced 3D object of the globe has been augmented on the palm and then registered in the scene successfully. The results of the globe augmentation are shown in Figure 4.3.



(a) Globe Augmentation (b) Globe Registration

Figure 4.3: Augmentation of Globe object

Three actions zoom-in, zoom-out and rotate are performed on the globe by using corresponding hand gestures. In order to make the usage convenient and standard, left hand is used to register an object in the scene and right hand is used to make different gestures for interaction. The results of object interaction with the three gestures are shown in Figure 4.4.

(a) Zoom-in

(b) Zoom-out

(c) Rotate

Figure 4.4: Interaction of Globe object

Chapter 5

Future Work

Though the framework has established the proof of concept, there are many areas to improve upon to make the entire experience more engaging. Following are the few problems to work upon in the future:

1. The current focus is to make the rendering more accurate and stable. We have observed that the media pipe model is detecting a fake hand in the background occasionally. This makes the rendering of virtual objects inconsistent resulting in unstable output. To overcome this problem, we want to fix the virtual object in the scene and make the hand ineffective once the object has been augmented. Therefore, once the object is registered, the left-hand detection in the video frame doesn't have any effect on the object's position.

2. Another challenge is to implement the scaling effect when we move the left hand toward and away from the camera. Ideally, the object should become bigger when brought close to the camera and smaller when taken away from the it.

3. It is essential to explore the area of graphical objects to enrich the quality of virtual objects. In the future, we would like to introduce new virtual objects with better textures and new features such as building a scene by registering multiple objects simultaneously.

4. As this work's main objective is to improve the teaching methodology, we plan to compose lecture content with our framework by gathering a series of teaching modules such as explaining graphs, DNA structure, etc.

Chapter 6

Conclusion

In this report, we have verified the proof of concept for marker-less augmentation. The proposed framework takes the help of the Mediapipe Hands model for hand tracking and feature detection. It uses the theory behind camera perspective transformation to project 3D objects in the real-world scene onto the image plane. The camera calibration technique and camera pose detection estimate the camera's parameters. The model, view, and projection matrices transform vertices that start in model space, and then world space, camera space, and then clip space. The framework uses hand gestures to manipulate the virtual objects after augmentation by zooming and rotating operations. The Mediapipe Hands model is the best technique that can fulfill the marker-less feature of this framework. Since only static hand key-points are selected to create a coordinate plane on the palm, the augmentation is stable as long as the entire palm faces the camera. This framework can be adopted in a classroom setting for enriching the learning experience by developing good teaching content and adding new features.

Bibliography

[1] Tara J Brigham. Reality check: basics of augmented, virtual, and mixed reality. *Medical reference services quarterly*, 36(2):171–178, 2017.

[2] Christian Moro, Zane Štromberga, Athanasios Raikos, and Allan Stirling. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical sciences education*, 10(6):549–559, 2017.

[3] Ronald T Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.

[4] Xiao Li, Bo Xu, Yue Teng, Yi-tian Ren, and Zhu-min Hu. Comparative research of ar and vr technology based on user experience. In *2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings*, pages 1820–1827. IEEE, 2014.

[5] Jon Peddie. *Augmented reality: Where we will all live*. Springer, 2017.

[6] Greg Kipper and Joseph Rampolla. *Augmented Reality: an emerging technologies guide to AR*. Elsevier, 2012.

[7] Marybeth Green, Joy Hill Lea, and Cheryl Lisa McNair. Reality check: Augmented reality for school libraries. *Teacher Librarian*, 41(5):28, 2014.

[8] Jernej Bule and Peter Peer. Interactive augmented reality marketing system. 2013.

[9] Khushal Khairnar, Kamleshwar Khairnar, Sanketkumar Mane, and Rahul Chaudhari. Furniture layout application based on marker detection and using augmented reality. *International Journal of Engineering Science*, 4201, 2016.

[10] Hong-Quan Le and Jee-In Kim. An augmented reality application with hand gestures to support studying geometry. , pages 160–161, 2016.

[11] Antonia Cascales, Isabel Laguna, David Pérez-López, Pascual Perona, and Manuel Contero. An experience on natural sciences augmented reality contents for preschoolers. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 103–112. Springer, 2013.

[12] Patrick Maier and Gudrun Klinker. Augmented chemical reactions: An augmented reality tool to support chemistry teaching. In *2013 2nd Experiment@ International Conference (exp. at'13)*, pages 164–165. IEEE, 2013.

[13] Rabia M Yilmaz. Educational magic toys developed with augmented reality technology for early childhood education. *Computers in human behavior*, 54:240–248, 2016.

[14] Matt Bower, Cathie Howe, Nerida McCredie, Austin Robinson, and David Grover. Augmented reality in education–cases, places and potentials. *Educational Media International*, 51(1):1–15, 2014.

[15] Antigoni Parmaxi and Panayiotis Zaphiris. Developing a framework for social technologies in learning via design-based research. *Educational Media International*, 52(1):33–46, 2015.

[16] David Harborth and Sebastian Pape. Exploring the hype: Investigating technology acceptance factors of pokémon go. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 155–168. IEEE, 2017.

[17] Facebook. Oculus quest hand tracking. https://www.oculus.com.

[18] Snap Inc. Snapchat lens studio. https://lensstudio.snapchat.com.

[19] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *BmVC*, volume 1, page 3, 2011.

[20] Andrea Tagliasacchi, Matthias Schröder, Anastasia Tkach, Sofien Bouaziz, Mario Botsch, and Mark Pauly. Robust articulated-icp for real-time hand tracking. In

Computer Graphics Forum, volume 34, pages 101–114. Wiley Online Library, 2015.

[21] Chengde Wan, Thomas Probst, Luc Van Gool, and Angela Yao. Self-supervised 3d hand pose estimation through training by fitting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10853–10862, 2019.

[22] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020.